A CLOUD-BASED DATA MINING TOOL FOR TRACKING HEALTHCARE IN RURAL AREA

[1]Vijesh Kumar Patel, [2]Alok Kumar, [3]Shahab Saquib

[1]Assistant Professor, [2]Assistant Professor, [3]Assistant Professor
[1]Department of Computer Science and Engineering,
[1]Bakhtiyarpur College of Engineering, Patna-800001, India

Abstract—The advancement of information and communication technology (ICT) makes healthcare more accessible and increases the efficacy of healthcare services provided in rural areas. A paradigm for tracking rural residents' healthcare was suggested in this research. In order to monitor healthcare, particularly for rural areas, the suggested framework is made to make use of data mining tools and a private cloud. This framework's principal goal is to raise healthcare standards and provide rural communities with access to medical facilities for general practice at the very least.

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

I. INTRODUCTION

Everyone's health is a major concern. A country cannot prosper if its citizens are ill. India's rural areas lack adequate healthcare facilities. The majority of Indians reside in rural areas. With respect to the country's population, India has a little over twenty-three thousand Primary Healthcare Centers (PHCs), which is insufficient.

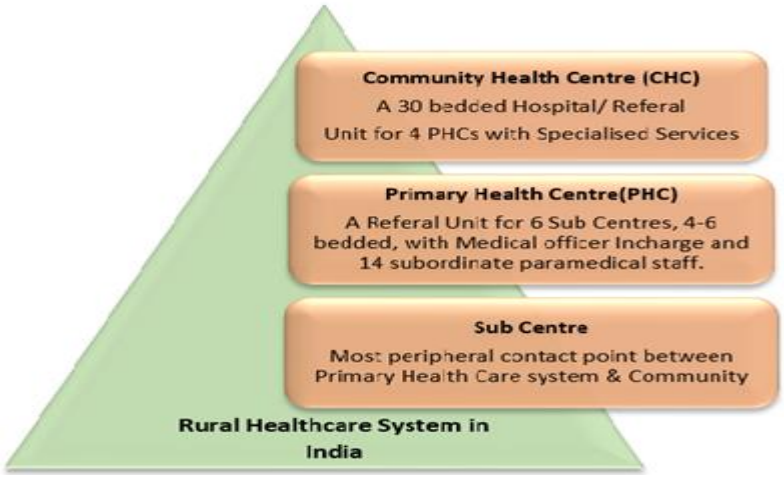

Figure 1 Rural Healthcare System in India [1]

India's rural healthcare system is depicted in Figure 1. A three-tiered approach is used to build the medical system in rural areas. First tiers are Sub-Centers (SC), PHCs are second tiers, and Community Health Centers are third tiers (CHC). The state government founded PHC, which serves as a point of contact between CHC and SC. Male health workers and ANMs (Auxiliary Nurse Midwives) run SC. However, the proportion of these health facilities to the population is insufficient [1]. Primary care is the type of healthcare that is not readily available in villages,

although it does exist for frequent and basic ailments as well as those for which prevention is more successful [2]. Approximately 700 million people reside in the nation's rural areas. In India, there are about 6 lakh villages. The majority of the locals have extremely low incomes. Because of their low income, even they lack access to basic healthcare facilities. Lack of access to clean drinking water, which is a major source of many ailments like Diarrhoea, Leukaemia, Fluorosis, etc., is another big issue in rural areas. Low physician-to-patient ratios in rural locations drive up healthcare costs and lead to inefficiencies in the system. The efficiency of basic healthcare might be improved and these facilities could be made more affordable with the use of ICT [2]. Individuals in rural areas lack access to basic and high-quality healthcare since there are insufficient PHCs and few doctors. India's rural and urban areas have different access to healthcare services [2]. In India, Figure 2 depicts the supply for healthcare services. Primary care, which offers medical treatment for common and minor illnesses as well as those for which protection is more impactful, is in highest demand. Secondary care, which requires ongoing medical attention and brief hospital stays, and tertiary care, which necessitates care from specialists or specialised clinics, are next in line. Figure 2 illustrates how healthcare availability varies in India between urban and rural areas.

Figure 2 Rural and Urban Health Divides

Primary healthcare facilities are nonexistent in rural areas, although corporate health facilities are readily available in urban areas. The ratio of hospital beds to population and the number of doctors per 1,000 people is lower in rural than in metropolitan areas, resulting in unequal access to hospital services. Villagers travel great distances to receive medical care in urban hospitals in the event of illness. In comparison to persons living in cities, they overspend on healthcare and hence incur more costs for the same condition. They cannot afford the healthcare services offered by urban hospitals because of their poverty. Urban and rural healthcare facilities differ greatly from one another. The maternal mortality ratio (MMR) in rural India is very high. The MMR calculates the number of maternal fatalities per 100,000 live births as the reason for a woman's death. In addition to having fewer institutional deliveries than in urban regions, rural communities lack a framework for tracking the health of expectant mothers. MMR is high in these places because of a lack of knowledge and accessibility issues with healthcare services [6]. Figure 3 displays the trends in MMR in India across several years.

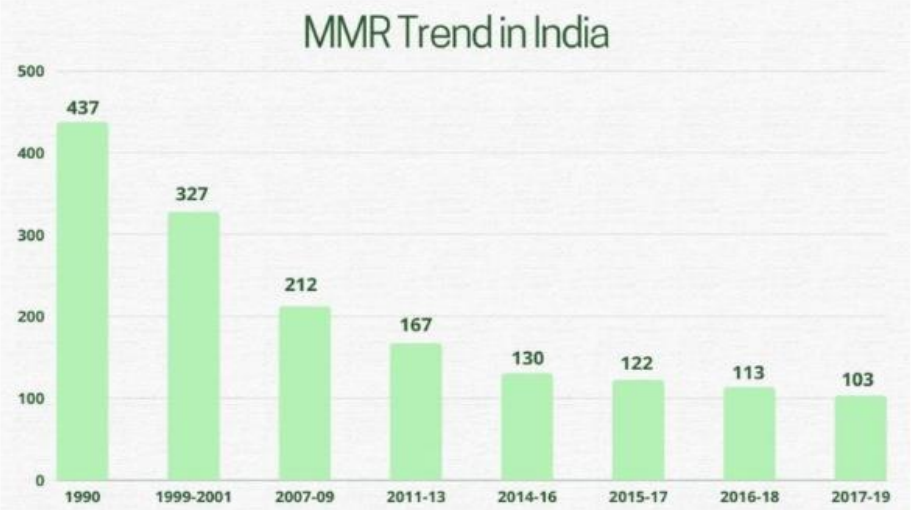

Figure 3 Maternal Mortality Ratio(1999-2019) in India [6]

India needs to increase its MMR by 97 by 2020 in order to meet the MDG (Millennium Development Goal), which will take a lot of work. Even though the government has launched a number of programmes, such JANANI SURAKSHA YOJANA, etc., to improve the MMR, these programmes still need to be managed well.

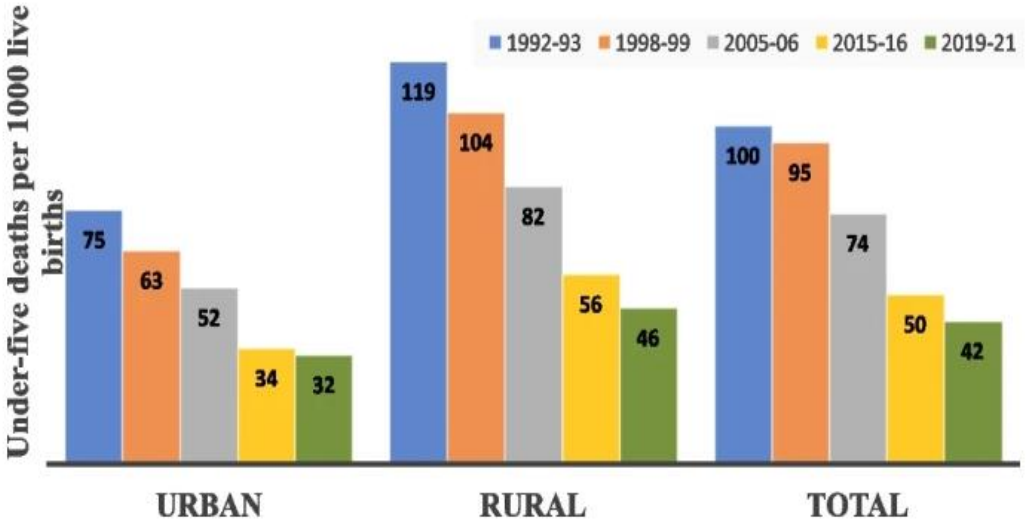

Figure 4 India's trends for the infant mortality ratio in its rural and urban areas [6]

The number of newborn fatalities (less than one year old) per 1000 children is known as the infant mortality rate, and it is likewise high in rural India. IMR in India is 42, as Figure 4 illustrates. However, there is a significant disparity between IMR in urban and rural regions.

II. CHALLENGES

While providing healthcare services to remote locations, many healthcare organizations encounter certain difficulties. These issues are mostly brought on by a lack of qualified labor in rural areas as well as inadequate rural connectivity. As seen in figure 5, we address the issues in this research from two different angles: the organizational perspective and the perspective of rural residents. The difficulties faced by rural residents include their inability to access healthcare facilities because of distance restrictions and their occasionally inability to pay for such treatments due to poor income. Lack of knowledge and awareness is another significant issue; many people are unaware of the different healthcare programmes that PHCs offer to them at no cost. Healthcare organizations deal with a number of issues as well. Their biggest obstacles are the social mores of the rural populace and a staffing shortage. In rural areas, self-medication is common. In order to overcome these obstacles, it is necessary to hire qualified personnel, raise rural residents' understanding of issues like diet and health, and offer them reasonably priced healthcare.

Figure 5 Challenges in rural areas

This research builds a framework for early disease identification utilizing data gathering and cloud computing techniques so that patients can receive timely and efficient treatment. Certain illnesses, like diabetes, TB, the flu, malaria, and others, are simple to identify and have standardized preventative measures. In rural areas, these diseases account for the majority of mortality. Therefore, our primary goal is to offer healthcare services, if not for all ailments, for those for which we can easily accurately diagnose using conventional techniques and whereas there are some common symptoms.

III. DATA MINING

A technique for extracting useful information from the massive amount of data to use in decision-making is called data mining. It started to exist in the mid-1990s [8]. Data mining is an effective technique that helps extract meaningful information and hitherto undiscovered patterns from large datasets. It assists the data owner in analyzing and identifying unexpected relationships between their data, which is beneficial for decision-making [9]. It is a crucial phase in the process of discovering new information. Fayyad et al. state that the knowledge discovery process consists of five stages: data collection is the first stage, pre-processing is the next, data migration is the third, data mining is done on the performed, and results evaluation is the last stage, illustrated in figure 6

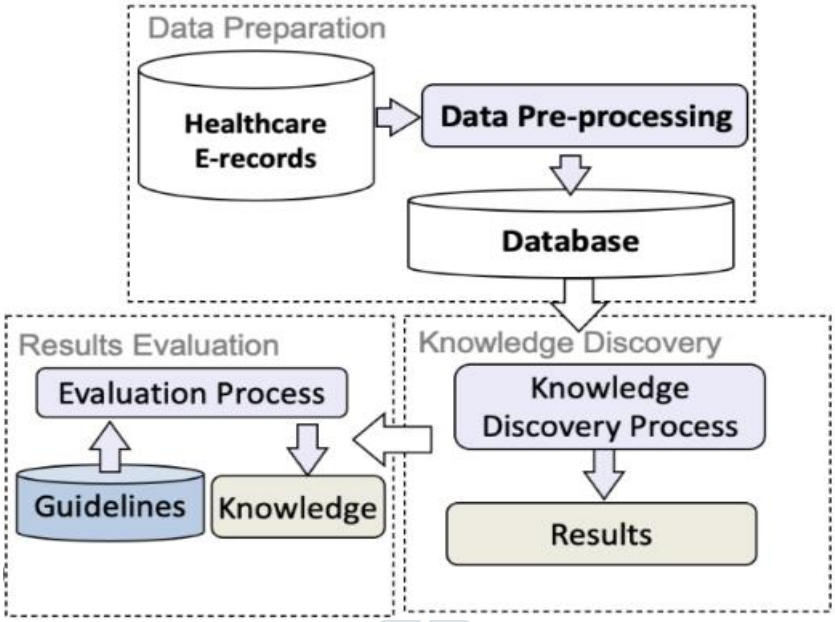

Figure 6 Procedure for Knowledge Discovery in Healthcare

The healthcare industry can benefit greatly from data mining techniques including association, clustering, and classification [12–25]. All of these data mining methods are significant and useful for diagnosing medical conditions, according to Tomar and Agarwal's analysis [12]. Among the different data mining approaches are association, clustering and classification. Data points are sorted into the appropriate classifications throughout the categorization process. That method is under supervision. Utilizing classification techniques, the classifier builds a model to forecast the class of data items. There are two categories of data points: the training phase and the test dataset. Training sets are used to train classifier models, and testing sets are used to verify the models' correctness [12]. Early illness prediction uses classification techniques as Artificial Neural Network, Least Square Twin Support Vector Machine, Ensemble Approaches, Decision tree, Support Vector Machine, and Weighted Support Vector Regression, among others [12–25]. Methods for clustering data points into comparable classes are employed. When compared to data points belonging to a different cluster, those belonging to almost the same cluster have a higher degree of similarity. The primary distinction between clustering and classifier techniques is that the former is unsupervised, while the latter is supervised [12]. Another crucial method for identifying connections or associations between data is association [9, 12].

IV. CLOUD COMPUTING

A sophisticated form of data technology called cloud computing enables the on-demand sharing of resources like devices, software, and hardware [35]. Figure 7 displays an outline of the cloud computing environment.

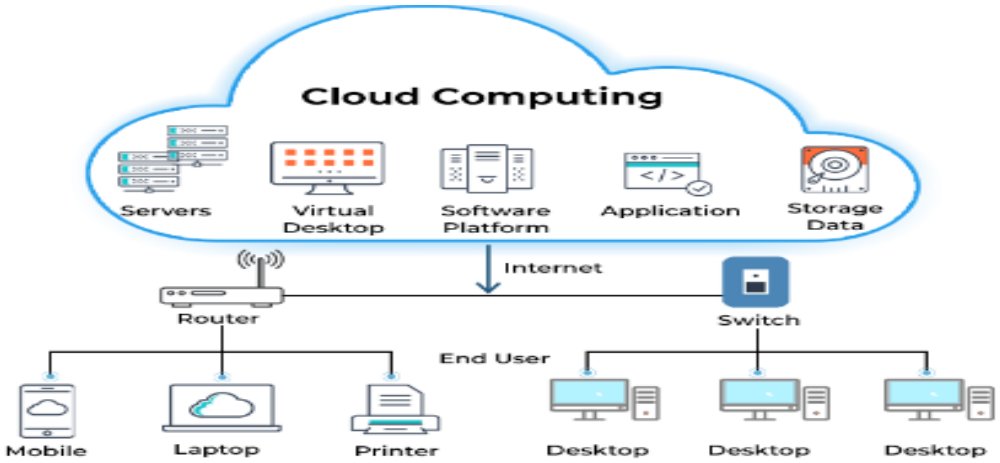

Figure 7 Architecture of Cloud Computing [36]

Three primary categories of cloud services exist, including [35]:

- Service-based infrastructure (SaaS): A basic cloud model for healthcare is offered, along with extra features like file storage, virtual machines, firewalls, and so forth.
- Platform-as-a-service (PaaS): This concept gives consumers access to the necessary online resources and services, such as a platform, on demand so they may develop applications. Additional software is not necessary to be installed in order to construct applications using the cloud. Services for planning, creating, testing, delivering, and maintaining a service are included in PaaS.
- User applications are provided as a service using the software-as-a-service (SaaS) concept. The program is available to users at all times and from any location. Among other things, it has web accounting and video conferencing.

A cutting-edge method that gives customers distributed access is cloud computing. Its flexible nature makes it a popular choice in the medical field [37–38]. The use of cloud software frameworks in healthcare has the following features:

1. A cloud-based infrastructure facilitates quicker access for both patients and doctors to medical facilities.
2. Since hardware segment is not a concern for physicians or healthcare organizations when building up the structure for patient healthcare monitoring, cloud computing offers a less expensive option.
3. Additionally, it gives patients and clinicians anytime, everywhere access to medical records.
4. Additionally, it enables synchronized access to the data in real-time and several concurrent users to obtain health services or data.

V. PROPOSED FRAMEWORK

Figure 8 shows the suggested structure. The information will be gathered from various Village Health center, clinics labs, medical professionals, hospitals, and mobile health workers. Each of these data sources is related to the paradigm that is being suggested. These sources' gathered data will be kept in one central location a primary health care server. With data mining, an integrated intelligent system is built. The early diagnosis of illnesses is aided by this decision support system. It also offers health advice to rural residents about how to maintain their health and prevent various diseases, as well as dietary and pharmaceutical recommendations based on disease

criteria. Additionally, it produces clinical alarms for patients in need of prompt attention from qualified professionals. The following services are offered by the cloud infrastructure:

- IaaS: Uses a central health server to provide computing infrastructure as a public utility.
- PaaS: Perched on the Infrastructure-as-a-Service (IaaS) stack, this platform offers middleware and interface functionalities to enable the simulation of outcomes through Pre-processing, Data Warehousing, Information Retrieval, and more.
- AaaS: provides application results across the World wide web in the medical cloud infrastructure, based on many medical datasets. A number of sources, including clinical alarms, dietary recommendations, early disease diagnosis, and health suggestions, have contributed to the existing dataset and its beneficial outcomes in the medical field.

Figure 8 is based on a hospital scenario that is divided into three layers, with the customer and health insurer interacting with each other:

- The proactive decisions and data mining methods receive patient and disease information.
- The vital patient data is provided via the hospital server and the state data mainframe.
- Additionally, findings and outputs are shared with the PHC or another healthcare facility.

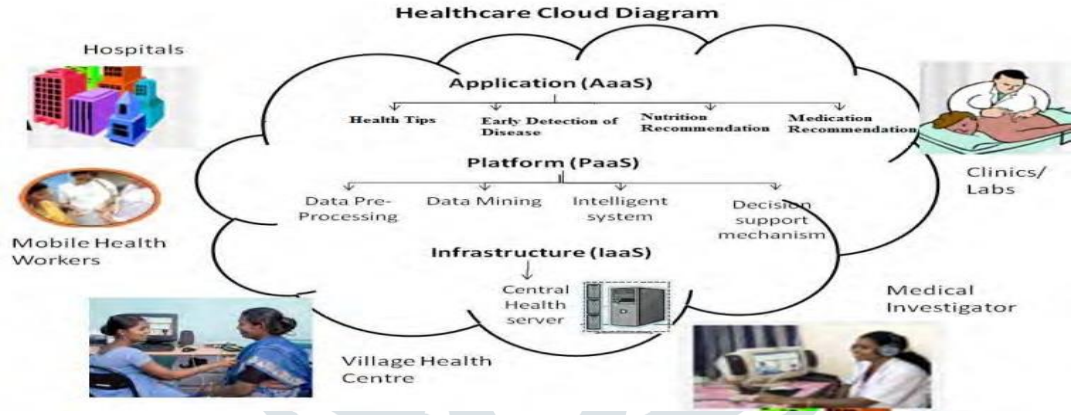

Figure 8 The suggested framework

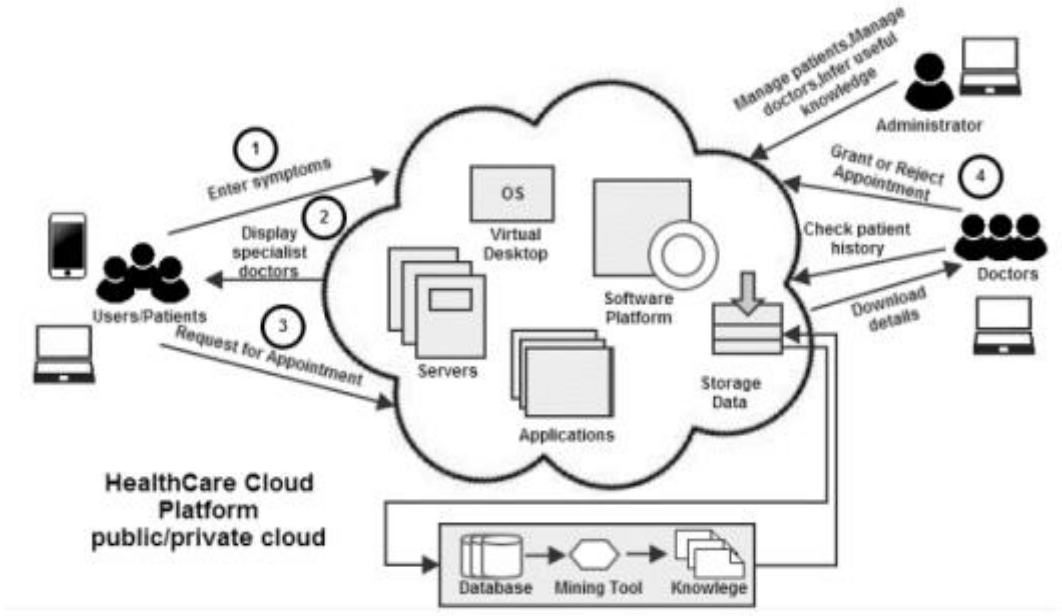

Figure 9 Diagram showing the connectivity of the healthcare cloud

The healthcare cloud's networking diagram is shown in Figure 9. Medical professionals can access the centralized health server, which is further linked to the data mining model, as depicted in the picture. A decision tree that assists in making decisions about patients' health and draws on the expertise of medical experts is built following an analysis of the data housed in the central health server. When a patient visits the PHC, this system evaluates their symptoms, makes a diagnosis, and recommends diet or medication to them, or it directs them to a qualified specialist based on their condition. This framework is primarily divided into three phases. We categories it into three groups: mining, pre-processing and data collection, which are further explained below:

Data Gathering: The purpose of this proposal's first phase is to collect and transfer patients' physical signs, such as those of sick or old individuals, pregnant women, and youngsters. The initial stage involves gathering patient data from different hospitals, PHCs, and health community centers. This stage collects a variety of vital physical patient data, such as blood sugar levels, heart rate, temperature, oxygen saturation, blood pressure, and ECG. The communication module uses the underlying Internet to communicate data while supporting 3G or wireless devices. The communicating module sends patient data that has been gathered by data collecting equipment to a central server that houses patient health data. PHC stores the patient's Electronic Health Record (EHR) on cloud storage, enabling users to access it from any location at any time. This system facilitates cloud-based information sharing amongst various medical organizations.

Computer Technology: Health information systems with varying specifications that were purchased from various vendors are used by all the medical facilities in the area. This will enable healthcare facilities to continue using their current information systems and participate in local health collaborations, mostly focused on illness management.

Information Mining: With data mining, an integrated intelligent system is built. Medical professionals or physicians can examine data gathered from various sources with the use of this decision support system. The construction of a decision-making model requires the analysis of the data that has been saved in the central health server using data mining techniques. In addition to offering advice on how to build this decision support system, medical professionals can also modify it based on their clinical experience treating patients.

Medical Services: The following healthcare services are offered by this framework:

- Digital Patient Information.
- Early illness identification.
- emergency referral framework.
- advice to the patient regarding diet and medication.
- health awareness advice for the patient.
- The remaining vaccination.
- focuses on creating markers to track disease and illness in children.
- The Integrated Child Development Scheme gives kids nutritional supplements based on weekly growth tracking.

This framework offers various advantages when appropriately applied, including:

- Time and money-saving. Avoid sending the patient on needless trips.
- raises the standard of medical care.
- In place of an untrained medical practitioner, the patient can receive knowledgeable expert counsel.
- This framework distinguishes between individuals who are seriously ill and in need of prompt attention from a qualified medical professional and those who are in the early stages and can be diagnosed with certain known symptoms.
- cooperation between healthcare professionals.
- Training non-medical experts is simple.

V. CONCLUSION

In India, the majority of people reside in rural areas. They lack access to basic medical care, and the quality of treatment in urban and rural locations varies greatly. Another significant issue in rural areas is the shortage of qualified medical professionals. This framework's major goal is to assist the less experienced non-medical specialists and other PHC workers in giving medical services to the villagers. There are several illnesses that have defined treatment protocols. The primary goal of this approach is to offer treatment, if not for all patients, for those who are afflicted with diseases for which standardized solutions exist, or for those who are recognized early and have identifiable symptoms. This framework also helps to distinguish between patients who are critically ill and in need of rapid specialist treatment, and those who are in the early stages and may be handled with ease. This framework's primary goal is to outperform both urban and rural medical systems. You may obtain information anywhere, at any time, thanks to cloud computing. Healthcare experts exchange their experiences and are linked to this framework, allowing them to treat patient's specific diseases and imparting knowledge that aids the non-medical professionals, or ANMs, at PHC, cure the patients.

REFERENCES

[1] http://www.nrhmassam.in/pdf/guideline2/bulletin/Rural%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] 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.

[4] 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.

[5] 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.

[6] http://censusindia.gov.in/vital_statistics/SRS_Bulletin s/MMR_release_070711.pdf.

[7] http://nhsrcindia.org/pdf_files/resources_thematic/Nutrition/NHSRC_Contribution/Discussion_Paper/Heal

th_and_Nutrition_in_Rural_India.pdf.

[8] Hand, D., Mannila, H., Smyth, P., Principles of data mining. MIT, 2001.

[9] Han, J., Kamber, M., Data mining: concepts and techniques. 2nd ed. The Morgan Kaufmann Series, 2006.

[10] 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.

[11] Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P., From data mining to knowledge discovery in databases. Commun. ACM 39 (11):24–26, 1996.

[12] 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.

[13] 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.

[14] 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).

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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).

[21] Fern'ndez G, DeLa Torre-díez I, Rodrigues JJ, "Analysis of the cloud computing paradigm on Mobile Health Records Systems" , IMIS 2012, IEEE, Sixth International Conference on 2012.

[22] Pino C, Di Salvo R, editors. A survey of cloud computing. architecture and applications in health. International Conference. on Computer Science and Electronics Engineering; 2013.

[23] 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.

[24] 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.

[25] 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.

[26] Agarwal, Sonali, and Divya Tomar. "A Feature Selection Based Model for Software Defect Prediction." International Journal of Advanced Science & Technology, 65, (2014).

[27] 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).

[28] 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.

[29] 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).

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] 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.

[35] Srinivasa Rao V. Nageswara Rao N K., E Kusuma K., Cloud Computing: An Overview , J ournal o f Theoretical a n d Applied Information Tec h nology , 2005.

[36] http://www.techiestate.com/top-2-cloud-computing- models/

[37] 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)

[38] 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).