ISSN No. 0976-5697

Volume 8, No. 5, May – June 2017

International Journal of Advanced Research in Computer Science

REVIEW ARTICE

Available Online at www.ijarcs.info

# Comparative Study of Techniques used in Healthcare Monitoring using Mining Mechanism

Rekha Rani
Computer Engineering & Technology
Guru Nanak Dev University, Amritsar, India

Kamaljit Kaur
Computer Engineering & Technology
Guru Nanak Dev University, Amritsar, India

***Abstract:*** Health predictions are critical area focused upon by technology. The technology plays great role in predicting diseases accurately and at good rate. This work is focused on analyzing techniques which are used for prediction purposes. The parameters for disease detection are fetched by the use of sensors and are maintained in the form of datasets. Internet of Things(IoT) is used primarily to collect data from the user. The techniques such as K nearest neighbour, Euclidean distance and ARIMA are considered for evaluation. Relative advantages and disadvantages are also highlighted in this work. Accuracy of data could be at stake since sensor may malfunction during collection process.

***Keywords :*** Health Prediction, IoT, K Nearest Neighbour, Euclidean distance, ARIMA

## I. INTRODUCTION

Technology is enhancing and is widely used in monitoring health. Health related issues occur more frequently due to late discovery of disease. Technology however helps in predicting diseases at the early stages and hence preventing disease. [1]proposes Internet of Things to create android application for health care. [2]Internet of things (IoT) is widely used for collecting information regarding different parameters from the users and then techniques are applied to predict the disease. Techniques used for prediction purposes considered in this paper include [3]KNN, [4]Random Forest, [5]Euclidean distance and [6]ARIMA. The prediction generated varies depending upon the accuracy of data presented. Accuracy of data presented depends upon the sensors. Sensors if malfunction may produce inaccurate data. Fault tolerance hence is critical in such situations. This paper explore the applications of IoT for collection of data, techniques used to process data presented by sensors and then fault tolerance capabilities possessed by different techniques used to analyses data presented by sensors. Information generated from sensors is stored within dataset as more and more information in terms of attributes are collected, size of dataset increases. The collected information is then analyzed for accuracy and future prediction. The demonstration associated with analysis process is listed as under.

Rest of the paper is organized as :section 2 describes application of IoT in health care environment, section 3 describes parameter collection mechanism, section 4 describes approaches used to determine prediction accuracy for data collected through IoT, section 5 describes literature survey and comparison table and last section describes conclusion and future scope

## II. IOT IN HEALTH CARE WITH APPLICATIONS

[7], [8]IoT (Internet of Things) provide pervasive framework used to predict health status of persons. Variety of application domains that is associated with IoT including health care. Social and economic aspects are considered through IoT and hence are efficient for prediction purpose. In order to use IoT, state of the art network architecture is required. It is a vital component of IoT in health care. Backbone of IoT is supported through network architecture. It facilitates transmission and reception of health data. The network architecture is critical causing easy access platform for attributes collection. Effective network architecture causes social and economic benefits and also causes predictions which are accurate in nature. IoT network architecture along with issues tackled are given as under:

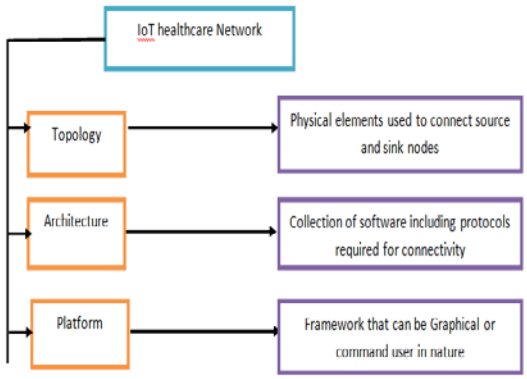


**Figure 1: Network Architecture for IoT Health care system**

[7], [8]Topology specifies physical equipments arrangement in IoT network. Topologies provide seamless health care environment achieved through distinct placement of

elements in network. Sensors used in the environment are of distinct type hence formed grid is heterogeneous in nature. Grid consisting of different types of sensors collects vital signs and sensor data such as temperature, blood pressure, electrocardiograms and oxygen levels forming topology associated with IoT network. Figure 2 describes how grid is used for collection of enormous vital signs and convert heterogeneous grids into hybrid computing grids. The architecture of computing grid is as shown below.

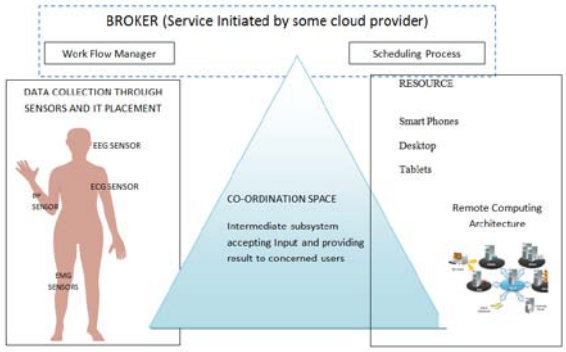


**Figure 2: IoT based ubiquitous Healthcare solution**

[1][10] Applications of IoT in health care are of prime importance. Services are provided through the support of IoT and are consumed by consumer using application support environment provided by protocol suite supported by IoT. Allocation areas are interactively shown as under:

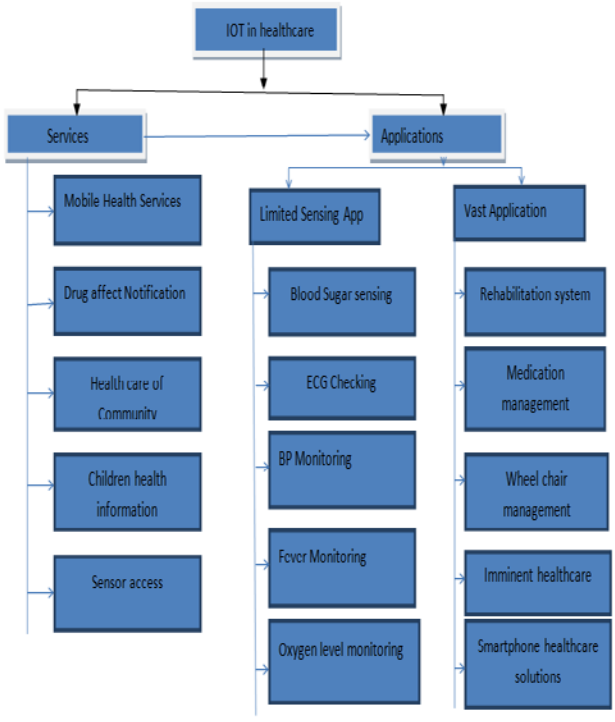


**Figure 3: IoT in health care environment**

Health care environment can involve single condition or clustered environment. Single condition indicates disease detection on the basis of single parameter list. The clustered condition indicates disease detection on the basis multiple parameters. To detect the disease number of techniques is devised. These techniques are prediction accuracy is listed through the following examples.

Example 1: Consider sensors attached to human body and collecting information about person heart rate(X), blood pressure(Y) and Temperature (Z). Collected information forms a dataset of following structure in the form of time series

**Table 1: Dataset of persons records with ID, Time, X,Y and Z.**

| ID | Time | X | Y | Z |
|---|---|---|---|---|
| 1 | 9:00 AM | 120 | 150 | 100 |
| 1 | 9:15 AM | 123 | 144 | 99 |
| 1 | 9:30 AM | 122 | 146 | 101 |
| 1 | 9:45 AM | 124 | 145 | 100 |
| 1 | 10:00 AM | 123 | 146 | 99 |
| 2 | 9:00 AM | 121 | 150 | 102 |
| 2 | 9:15 AM | 125 | 141 | 101 |
| 2 | 9:30 AM | 126 | 142 | 105 |
| 2 | 9:45 AM | 121 | 149 | 99 |
| 2 | 10:00 AM | 120 | 156 | 98 |

## III. PARAMETER FETCHING MECHANISM

The evaluation of Internet of Things greatly facilitates the diagnosis process of patients. [11]Suggests monitoring of records associated with patients is becoming possible with the utilization of IoT. In order to accomplish this task, [12][13]small IP based wireless sensor (Proximity sensors) is attached with the patient body. [14]Proposes sensor based sensing application that helps in monitoring the psychological parameter like heart rate and blood pressure remotely and frequently. The record so obtained can be stored over the cloud so that patient record can be retrieved as and when required. The proposed work studies the applications of the IoT in the field of health care along with management policies used to enhance security of records stored within cloud.

### A. COLLECTION OF PARAMETERS THROUGH IoT

The parameters collection is integral part of health care. Collection of parameters is accomplished through sensors and organized in the form of tabular structure. [15], [16]as more and more data is collected Big Data is formed, it is organized to form dataset. The process of collecting parameters includes the sensors implanted on different parts of the body. The sensor produces information that are stored in the memory as person performs distinct activities. Overall organization of internet of things in parameters collection is organized as follows:

**Table 2: Parameter Collection Settings.**

| PARAMETER COLLECTION "PLACEMENT OF SENSORS" ALONG WITH ITS TYPE | | |
|---|---|---|
| Parameters | Description | Utilization Example |
| Human Body | Devices attached inside or outside human body | Devices used to maintain well being of humans. Applications include disease management, increased productivity etc. |
| Home based environment | Homes and Building where people live in | Sensors used in security systems |
| Business Store | Places where customers engage in transactions | Stores, Banks, maul etc involving large number of people. |
| Offices | Place where intellectuals interact with each other for business | Management of energy and security enhancement services in buildings |
| Organization like factories, industries etc. | Mostly used in production | Places where repetitive work is done like in hospitals, inventory systems. |
| Sites where actual work is done | User specific customer environment | Oil Mining and construction environment |
| Cars and other moving vehicles | System which work inside moving vehicles | Vehicles including cars, jeeps etc used to monitor consumption of fuel. |
| Urban Environment | Cities | Smart Cities |
| Miscellaneous | Between Urban and rural area | Including rail tracks , roads etc. used to detect blockage if any |

Collection of parameters collected through the above listed source form dataset. For detection of disease related to Activities, dataset from UCI website can be drawn. Parameters collected for disease detection is obtained by placing sensor over or inside the body. Next section describes techniques used to predict diseases through data obtained through sensors.

## IV. TECHNIQUES USED TO ANALYSE DATA COLLECTED THROUGH SENSORS

### A. IMPLICATION OF KNN

The dataset can be analyzed by the use of K-nearest neighbor technique. Let the value of K=2 and implication is assigned on X, Y and Z. The test point of x is considered to be 120. Test point y is considered to be 143.The test point of z is 99. Then KNN generates

**Table 3: KNN deviation results.**

| Id | $X= \sum \frac{X-120}{N}$ | $Y=\sum(Y-143)/N$ | $Z=\sum(Z-99)/N$ |
|---|---|---|---|
| 1 | 2.16 | 2.66 | 0.66 |
| 2 | 2.16 | 4.83 | 2 |

The decision boundary is created for detecting and forecasting purpose in case of KNN. The decision boundary is established by the use of rules. These rules are in the form of IF-THEN form. The threshold if assumed to be 2.5 for X, 3 for Y and 1.5 for Z then X is falsified, person 2 is detected with disease corresponding to Y and Z.

### B. IMPLICATION OF EUCLIDEAN DISTANCE

Euclidean distance is used to calculate the distance of noted dataset vales from the test point. The threshold value is compared with obtained value to determine anomalies. The test point of x is considered to be 120. Test point y is considered to be 143.

**Table 4: Euclidean Distance Results**

| Id | $X= \sqrt{\frac{(X-X_i)^2}{Total_{observation}}}$ | $Y= \sqrt{\frac{(Y-Y_i)^2}{Total_{observation}}}$ | $Z= \sqrt{\frac{(Z-Z_i)^2}{Total_{observation}}}$ |
|---|---|---|---|
| 1 | 2.7 | 3.5 | 0.93 |
| 2 | 2.7 | 5.53 | 3.12 |

The threshold if assumed to be 2.5 for X, 3 for Y and 1.5 for Z then Person 1 is detected with Z, person 2 is detected with disease corresponding to X, Y and Z. The hybridization of KNN + Euclidean distance can be used in the future research for optimization.

### C. IMPLICATION OF ARIMA

[17][18], [19] For accurate prediction of disease detection the Auto regressive moving average model is used. By using mathematical model the modification in time series are to be done in ARIMA. This model is based on adjustment of observed values. The goal is to obtain the differences of observed value and value obtained from the model close to zero. This model can predict accurately difference between the stationery and non stationery series. Working of this model is described through the following diagram :

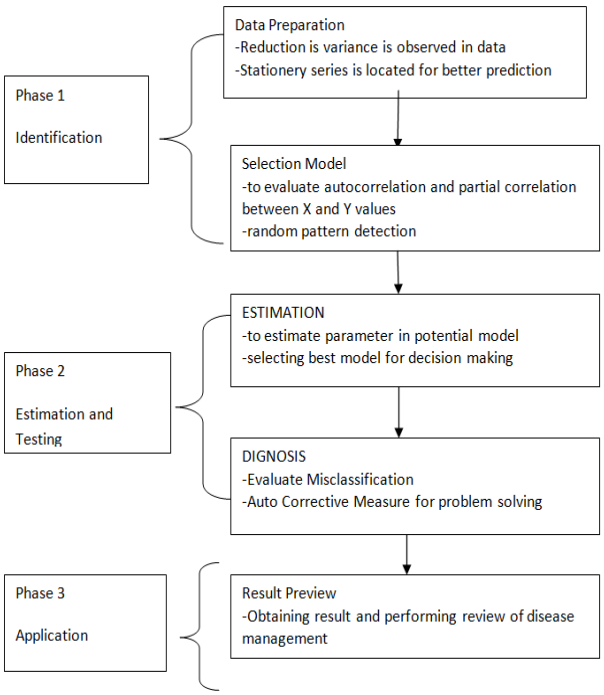


**Figure 4: Working of ARIMA model for Health prediction**

ARIMA model has multiple phases associated with it. This model can be merged with Euclidean distance and KNN for better performance in health prediction.

## V. LITERATURE SURVEY

In [20], gather the related data that show the significance of data mining in human services. As the measure of gathered health information is expanding altogether consistently, it is trusted that a solid examination instrument that is equipped for taking care of and dissecting huge wellbeing information is fundamental. The [21] proposes developing requests on administration quality and cost regulation test the medicinal services associations around the globe. They are especially essential concerning regular sickness or restorative conditions. One technique that can be utilized to address these issues is the use of human services data frameworks for choice support and information administration.[22]suggests affiliation rules mining process empowers the end clients to break down, comprehend, and utilize the separated learning in a canny framework or to bolster the basic leadership forms. To discover significant affiliation rules from a substantial number of excess guidelines, this paper proposes a more profound mining process, multi-mode and high esteem affiliation rules mining. In [23]this paper, Data Mining is presented and in addition huge information in the system of Healthcare. Besides, the Data Mining for aggregated information is examined. Particularly , their complexities of the different ranges wellbeing and therapeutic research. At long last, machine learning calculations have been utilized as a part of request to study handling Healthcare information.[24]suggests development in the human services information has roused us in exploring vigorous and adaptable models for information mining. For Classification issues Information Gain (IG) based Decision Tree is one of the well-known decisions. Nonetheless, contingent on the way of the dataset, IG based Decision Tree may not generally perform well as it lean towards the characteristic with more number of particular esteems as the part quality.[25]Proposes an examination toolbox in view of open-source modules that encourage the investigation of wellbeing correlated datasets. We outline our system by giving a definite investigation of doctor and clinic appraisals information. Our method ought to demonstrate profitable to programming engineers, enormous information designers, healing center executives, strategy creators and patients.[26]suggests linkage between human services administration and distributed computing methods has drawn much consideration recently. Up to the present, most works concentrate on IT framework movement and the administration of disseminated social insurance information as opposed to exploiting data covered up in the information. In this paper, we propose to investigate social insurance information by means of cloud-based human services information mining administrations. Specifically, we propose a cloud-based medicinal services information digging structure for social insurance information mining administration improvement. Under such system, we additionally build up a cloud-based medicinal services information mining administration to foresee patients future length of remain in clinic.[27]Proposes Human services information is viewed as exceptionally critical to specialists in this field. Such data must be distributed with strategies that keep the personality of patients shrouded particularly when managing touchy data. Distributing such data makes it more powerless against aggressors. Accordingly, numerous procedures were proposed to safeguard the protection of medicinal services information. In this paper, we illustrated a review for the models and systems that are utilized for distributing information about patients.[28]discuss huge volumes of medicinal information are in effect persistently created, yet their esteem is extremely undermined by our powerlessness to make an interpretation of them into learning and, at last, activities. Information mining methods permit the extraction of already obscure fascinating examples from expansive datasets, yet their many-sided quality constrains their useful dispersion. Information driven examination is a multi-step prepare, in which social insurance experts define investigation objectives and survey extricated learning, while PC researchers handle the non-trifling errand of driving the digger framework examination movement.[29]suggest keen on building a portable based BMI forecast framework utilizing Kinect and information mining methods with the goal that everyone can without much of a stretch screen their BMI wherever by taking a depiction of their face. The basis behind this is the instinct that there is a connection between the state of one's face and

one's BMI esteems, which individuals frequently follow up on while seeing a companion has either put on or shed pounds.

### A. COMPARISON OF LITERATURE REVIEW

**Table 5: Comparison of literature**

| AUTHORS | PARAMETERS | DATA GATHERING MECHANISMS | MERITS AND DMERITS |
|---|---|---|---|
| [20], | Accuracy | Survey uses Electronic health record | It is a survey paper describing advantages of many distinct techniques hence optimal approach can be selected for research. No new technique is suggested through this literature |
| [21] | ERROR RATE ACCURACY | ELECTRONIC HEALTH RECORD | It present a case study to be implemented in practical applications The specific environment is considered in which accuracy is an issue and can be further enhanced. |
| [22] | Accuracy | Rule Based Approach | Fast mechanism of data gathering is accomplished through proposed mechanism Accuracy however can further be improved. |
| [23] | Accuracy | Clinical Decision Support System | Data gathering is with accuracy but missing value are not handled |
| [24] | CORRELATION ACCURACY | Correlation based Ratio Analysis | Correlation is used hence specific information about health care is gathered efficiently Missing Value is not considered |
| [25] | Prediction Accuracy | Open Dataset | Graphical interface is used hence visualization is better Missing value however is not considered. |
| [26] | Future Prediction | Heterogeneous health care data | Future prediction can be efficiently done Missing value are not handled |
| [27] | Accuracy Privacy | Raw medical dataset | Privacy is efficiently managed Missing value is not handled |

Findings from existing literature is listed as under

1. Most of classifiers focuses on accuracy of classification without considering missing values
2. In case missing values are introduced accuracy decreases substantially.
3. Most probable values can be used to replace missing values to enhance accuracy which is missing in exiting literature.

## VI. CONCLUSION AND FUTURE WORK

Health prediction is critical for the early detection of disease. Large number of models is devised for this purpose. This paper described most proficient models used for prediction purpose. ARIMA model is multi objective model used to predict diseases. This model can be used to enhance accuracy in case missing values are detected within the dataset. Missing values can be handled using most probable value derived using Euclidean distance along with KNN in ARIMA.

Accuracy can be enhanced through the hybridization of KNN and Euclidean distance in ARIMA.

## REFERENCES

[1] J. Mohammed, C.-H. Lung, A. Ocneanu, A. Thakral, C. Jones, and A. Adler, “Internet of Things: Remote Patient Monitoring Using Web Services and Cloud Computing,” in 2014 IEEE International Conference on Internet of Things(iThings), and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom), 2014, pp. 256–263.

[2] D. V. Dimitrov, “Medical internet of things and big data in healthcare,” Healthc. Inform. Res., vol. 22, no. 3, pp. 156–163, 2016.

[3] V. S. H. Rao and M. N. Kumar, “A new intelligence-based approach for computer-aided diagnosis of Dengue fever.,” IEEE Trans. Inf. Technol. Biomed., vol. 16, no. 1, pp. 112–8, Jan. 2012.

[4] J. Abawajy, S. Member, M. Chowdhury, and A. Kelarev, “Hybrid Consensus Pruning of Ensemble Classifiers for Big Data Malware Detection,” vol. 3, no. 2, pp. 1–11, 2015.

[5] D. Sinwar and R. Kaushik, “Study of Euclidean and Manhattan Distance Metrics using Simple K-Means Clustering,” Int. J. Res. Appl. Sci. Eng. Technol., vol. 2, no. 5, pp. 270–274, 2014.

[6] D. Banerjee, “Forecasting of Indian stock market using time-series ARIMA model,” in 2014 2nd International Conference on Business and Information Management (ICBIM), 2014, pp. 131–135.

[7] M. Maksimovic, V. Vujovic, and B. Perisic, “A custom Internet of Things healthcare system,” in 2015 10th

Iberian Conference on Information Systems and Technologies (CISTI), 2015, pp. 1–6.

[8] P.-C. J. Chung, "Impacts of IoT and Wearable Devices on Healthcare," in Proceedings of the 12th International Conference on Advances in Mobile Computing and Multimedia - MoMM '14, 2014, pp. 2–2.

[9] L. M. R. Tarouco et al., "Internet of Things in healthcare: Interoperatibility and security issues," in 2012 IEEE International Conference on Communications (ICC), 2012, pp. 6121–6125.

[10] V. M. Rohokale, N. R. Prasad, and R. Prasad, "A cooperative Internet of Things (IoT) for rural healthcare monitoring and control," in 2011 2nd International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology (Wireless VITAE), 2011, pp. 1–6.

[11] "How the Internet of Things Is Revolutionizing Healthcare - white paper - IOTREVHEALCARWP.pdf." .

[12] S. Abdelwahab, B. Hamdaoui, M. Guizani, and T. Znati, "R EPLISOM : Disciplined Tiny Memory Replication for Massive IoT Devices in LTE Edge Cloud," vol. 4662, no. c, 2015.

[13] W. Xu et al., "A context detection approach using GPS module and emerging sensors in smartphone platform," in 2014 Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS), 2014, pp. 156–163.

[14] H.-W. Guo, Y.-S. Huang, C.-H. Lin, J.-C. Chien, K. Haraikawa, and J.-S. Shieh, "Heart Rate Variability Signal Features for Emotion Recognition by Using Principal Component Analysis and Support Vectors Machine," in 2016 IEEE 16th International Conference on Bioinformatics and Bioengineering (BIBE), 2016, pp. 274–277.

[15] D. Li, H. W. Park, E. Batbaatar, Y. Piao, and K. H. Ryu, "Design of Health Care System for Disease Detection and Prediction on Hadoop Using DM Techniques," pp. 124–129.

[16] G. Vaishali and V. Kalaivani, "Big Data Analysis for Heart Disease Detection System Using Map Reduce Technique," vol. V.

[17] D. V Jose and G. Sadashivappa, "a N Ovel E Nergy E Fficient R Outing a Lgorithm for W Ireless S Ensor N Etworks," Int. J. Wirel. Mob. Networks, vol. 6, no. 6, pp. 15–25, 2014.

[18] Y. Pan, M. Zhang, Z. Chen, M. Zhou, and Z. Zhang, "An ARIMA based model for forecasting the patient number of epidemic disease," 2016 13th Int. Conf. Serv. Syst. Serv. Manag. ICSSSM 2016, pp. 31–34, 2016.

[19] I. A. Permanasari, A.E. Hidayah, I.Bustoni, "SARIMA (Seasonal ARIMA) implementation on time series to forecast the number of Malaria incidence," Inf. Technol. Electr. Eng. (ICITEE),2013 Int. Conf. ., no. 2, pp. 2–6, 2013.

[20] M. H. Tekieh and B. Raahemi, "Importance of Data Mining in Healthcare," Proc. 2015 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Min. 2015 - ASONAM '15, pp. 1057–1062, 2015.

[21] A. Shahin, W. Moudani, F. Chakik, and M. Khalil, "Data Mining in Healthcare Information Systems: Case Studies in Northern Lebanon," 2014 Third Int. Conf. e-Technologies Networks Dev., pp. 151–155, 2014.

[22] L. Yang, Z. Li, and G. Luo, "MH-ARM: A Multi-Mode and High-Value Association Rule Mining Technique for Healthcare Data Analysis," 2016 Int. Conf. Comput. Sci. Comput. Intell., no. 71432002, pp. 122–127, 2016.

[23] Z. A. Mdaghri, M. El Yadari, A. Benyoussef, and A. El Kenz, "Study and analysis of Data Mining for Healthcare," Ieee, pp. 77–82, 2016.

[24] S. Roy, S. Mondal, A. Ekbal, and M. S. Desarkar, "CRDT: Correlation Ratio Based Decision Tree Model for Healthcare Data Mining," 2016 IEEE 16th Int. Conf. Bioinforma. Bioeng., pp. 36–43, 2016.

[25] A. R. Rao and D. Clarke, "A fully integrated open-source toolkit for mining healthcare big-data: Architecture and applications," Proc. - 2016 IEEE Int. Conf. Healthc. Informatics, ICHI 2016, pp. 255–261, 2016.

[26] P. Zhang, S. Hu, J. He, Y. Zhang, G. Huang, and J. Zhang, "Building cloud-based healthcare data mining services," Proc. - 2016 IEEE Int. Conf. Serv. Comput. SCC 2016, pp. 459–466, 2016.

[27] L. A. Abuwardih, W. Shatnawi, and A. Aleroud, "Privacy preserving data mining on published data in healthcare: A survey," 2016 7th Int. Conf. Comput. Sci. Inf. Technol., pp. 1–6, 2016.

[28] T. Cerquitelli, E. Baralis, L. Morra, and S. Chiusano, "Data mining for better healthcare: A path towards automated data analysis?," 2016 IEEE 32nd Int. Conf. Data Eng. Work. ICDEW 2016, pp. 60–63, 2016.

[29] C. H. Tai and D. T. Lin, "A Framework for Healthcare Everywhere: BMI Prediction Using Kinect and Data Mining Techniques on Mobiles," Proc. - IEEE Int. Conf. Mob. Data Manag., vol. 2, pp. 126–129, 2015.

 