International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-5, June 2019

Predictive Analysis in Intelligent Healthcare Framework Using Big Data Applications

T. Papitha Christobel, A. Sasi Kumar

Abstract: ***Recently, logical research on healthcare services request to expand an intelligent choice to offer sound life office with ahead of time disorder discovery to the individual. In the utmost recent time, healthcare services ventures are creating masses of unstructured or semi-structured certainties which want to be investigated and treated continuously. In this paper, we have planned a healthcare services framework to address sufferer's natural, and enthusiastic condition and additionally the previous wellbeing records with genetical records. The data formed by methods for the patient and the healing centers are accumulated in high-performance computer server, and the logical history, notwithstanding genetical data, is gathered from the cloud synchronization. We developed a probabilistic dimensions securing plan to investigate the insights and take after MapReduce algorithm in High Performance Computing (HPC) to make shape database. The contraption holds an actualities distribution center which gives a two-way collaboration among HPC and cloud for intuitive quantities hoarding. In this exploration, we show an expectation algorithm that is completed in cloud server to expect a patient's issue. We apply Artificial Neural Network, Random Forest, SVM, C5.0 and Naive Bayes for expectation examination and demonstrate the side by methods for feature appraisal on the ones algorithms.***

Index Terms: ***Big Data, Cloud Computing, Artificial Neural Network, Random Forest, SVM, C5.0, Naive Bayes***

I. INTRODUCTION

In the recent era, food is not that much healthy provided to the people. Due to the reason of food habits and pollution, people are facing big health issues. To provide a solution, researchers are focusing on developing a health care system. The human interaction made with the system to offer good assistance with medicine based on their patient's physical condition. This is one of the greatest growing learning areas that attract numerous specialists and researchers most recent time. These days, clinical treatment wishes an interactive and sensible machine that could manage a major natural dataset with human-pc interchange to investigate the greatest profitable measurements and give a higher cure identified with wellbeing. At show, it's no longer immense to expect or design a shrewd social insurance contraption that can have association with human in a matter of seconds and an intense way. To offer some extraordinary offices identified with medical services, the National Health Reform units a couple of wants and targets which screen the wellbeing progress and to find the essential changes. The concentrating spot of immense extent of wellbeing administrations look into today is the specific field of the human services. This may be drawing nearer to the prevalent wellness contributions scientists [1].

Manuscript published on 30 June 2019.
* Correspondence Author (s)
T. Papitha Christobel*, PhD Research Scholar, Department of Information Technology, School of Computing Sciences, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Pallavaram, Chennai, India.
A.Sasi Kumar, Professor, Department of Information Technology, School of Computing Sciences, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Pallavaram, Chennai, India.

A. HCI in Data Collection

Biosensor based IoT devices joined with smart wireless technology and data mining strategy techniques better human healthcare framework to investigate data with intelligent patient observing technique. This new intelligent healthcare framework will most likely gather numerous patients' data (for example legitimate and enthusiastic data of patients) in an intuitive way utilizing biosensors based IoT devices and sensor-less devices inside a brief period. Analysts think about that, medical data gathering is one of the huge segment in a healthcare [2]. Data gathering is massively important to analyze a patient's illness and investigation of other significant reports with future ailment expectation. Along these lines, human computer collaboration is basic to gather data from patient to guarantee more noteworthy number of data with less exertion and less time which should be dissected.

B. Healthcare with Big Data

Intelligent healthcare frameworks produce monstrous volume of continuous medical data and those data accompany unstructured or semi-structured format. The current intelligent healthcare frameworks utilize electronic wellbeing records to store those data. American Hospital Association demonstrated that employments of Electronic Health Records turned out to be twofold from 2009 to 2011 [3]. As detailed by the intelligent healthcare data analysis, 150 Exabyte's of medical data are delivered by USA healthcare in 2011 [4]. In 2014, this sum is come to zettabytes [5].

C. Properties of Medical Big Data

Intelligent healthcare framework must be consulted with this immense volume of restorative data with big data investigation to find the concealed examples and scan for unrevealed connection with the patient's past medical data. In medical big data nalytics, the significant test is to manage the characteristics of huge data which is characterized by 5Vs: Volume, Velocity, Variety, Value, and Veracity.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

In this new intelligent healthcare framework, patient's data which are gathered from different sources portray the volume of data. The landing rate of data speaks to the speed of data.
The social insurance data, for example, Electrocardiography (ECG), Electromyography (EMG), clinical reports, specialist's note originate from numerous sources and those accompany structure, semi-structured and unstructured arrangement that depict the assortment. Those data should be investigated to discover the significant data which is considered as the estimation of data. There are numerous unsure number of state in data which we called hidden data are represent to the veracity.

D. Structure of Big Data

Some advanced technology is utilized to examine big data, for example, tensors, cloud computing and some clever system which can deal with missing data from huge measure of medical data [6]. Be that as it may, handling of huge data prompts challenges as there are numerous semi-structuredand unstructured data. IBM medical services scientists found that, 80% of big data in human services are unstructured and they additionally guarantee that to recover productive data about patients, those unstructured or semi-structureddata should be examined to make those data in structuredorganization [7]. Another blog named MapR Converged Data Platform tells that in the present human services condition, 75% or a greater amount of the data by certain assessments is unstructured data [8]. Then again, a portion of the scientists demonstrates that computerized data are developing quickly. They think about the measure of the structuredand unstructured data of the most recent decade. As indicated by International Data Corporation (IDC) look into, among all the all out gradual advanced data practically 90% data are unstructured[9]. Some of the time big data known as grimy data and before huge data investigation, we should structure those medical data for better examination result. Unstructured and semi-structureddataset can hold significant data about data and those unstructured or semi-structured dataset causes off base data about the dataset [10].
In this paper, we contribute the accompanying can be outlined.

• Interactive data gathering plan to get data with less exertion utilizing a portion of the IoT devices and sensors.
• Significant data examination of unstructured and semi structured data to make an structureddatabase and store those data productively.
• A two-route association with utilizing good data distribution center with HPC and cloud synchronization which can store the medical data for what's to come.
• This framework can manage manifestations of the diesease, passionate data of the patient just as the chronicled medical data and genetical data about the patients.
• An expectation model which can anticipate both cost estimation and sickness prediction in a quicker manner.

II. LITERATURE SURVEY

Healthcare framework is improving consistently for the last half-decade for the constant research of the analysts. These days, Healthcare framework is ending up further developed than past. It has loads of Internet of Things (IoT) devices and Electronic Health Records (EHR) innovation to records the majority of the patient's medical data effectively. Be that as it may, scientists, as yet investigate about the medical framework with the innovations to build up some intelligent human healthcare framework. BDAEH is one of the as of late proposed healthcare system, which takes the patient's coherent and enthusiastic data as the data and after that can investigate those data to anticipate the patient's disease [14]. Yet, BDAEH can't manage the past medical data of the patient and furthermore can't break down the patient's genetical data.
AIWAC is a proposed human healthcare system that can accumulate data utilizing input data devices, look at the received data, anticipate the diesease and can direct the interaction [15]. ReTiHA is another framework which can persistently screen the patient and can accept those data as the info data and after that that framework can tip wellbeing encouraged to the patients [16]. Some analyst's exploration about the medical services with the critical data examination and give a algorithm to expectation model with 98% exactness and structure a probabilistic data accumulation with the assistance of various IoT devices and sensors [17].

There is likewise having a few systems to gather medical data (for Cerebral Palsy) intelligently from the patient utilizing Electrical Medical Records (EMR) [18]. A few scientists depict the huge data which is the current to the human services and the future open doors about the medical services [19], [20]. They likewise demonstrated that the exploration about huge data dissecting in social insurance is expanding quicker. There is likewise some examination about the assessment of big data in medical services data utilizing data mining approach [21]. A portion of the frameworks were proposed which can gather data from EHR and after that utilization AI algorithm that can ready to examine those data [22].

Healthcare choice emotionally supportive network is recommended that can break down the medical data utilizing the assistance of neural system and other data mining algorithm, for example, Naive Bayes and C4.5 [23]. A portion of the specialists utilize the diverse AI algorithm, for example, arbitrary timberland, neural nets and supported trees. They additionally actualized progressively conceivable models, for example, naives Bayes, logistic regression, and single decision trees for a contextual analysis of the emergency clinics [24].

In light of those contextual investigations they demonstrated which algorithm is predictive to foresee the illness of the patient. There is likewise having a few looks into about huge data in medical research [25], [26].

III. PROPOSED METHOD

A. Introduction

Our proposed intelligent healthcare framework engineering has concentrated on three stages which were appeared in Figure .1 the data gathering with the extraordinary collaboration of smart biosensor devices, break down that monstrous volume of data of individual patients just as emergency clinics for future illness prediction and discover the example of comparable patients. These three stages are execute in three distinctive layer called primary layer, operating layer and application layer. In this paper, we will examine the design of those three layer and figure the engineering of the cloud based server for the framework.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

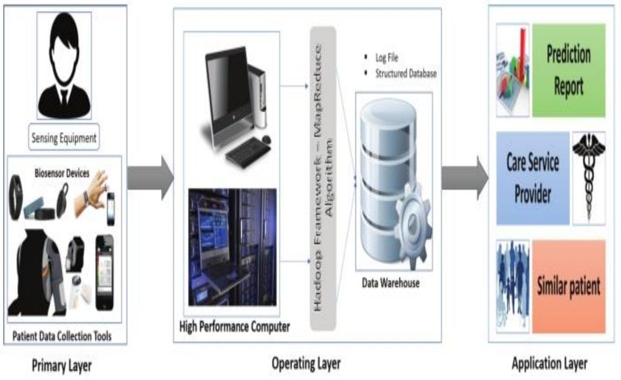

Figure 1. Three phases of intelligent healthcare Framework: Primary layer, Operating layer, and Application Layer

B. System Architecture

The framework will figure the biomedical data, and passionate state of the specific patient through smart biosensor devices and other data, for example, patient's medical history will be gathered from hospital server. Those data will be sent to the High-Performance Computer server for essential investigation. In our framework, the data gathering stage in HPC is called primary layer of data investigate conspire where every unstructured data are being assembled. The following layer is operating layer where some programming model, for example, Hadoop MapReduce system will be relevant for making an structured database with some log file and metadata. Those metadata and database will make a decent client collaboration in data mangement. In the wake of making the log record and database, those will be coordinated and sent to another elite server which we called an data stockroom. The log record is utilized as metadata for giving quick seeking capacity and examination. In application layer, data stockroom sends the database to the server farm for data stockpiling which is situated in the cloud. Cloud gives the office of quick processing to analyze a sickness and foresee the future malady and empowers a lot more offices.

C. Hadoop MapReduce Framework

In data gathering stage, HPC gathers healthcare data in a consistent time space. The preparing of those broad data is a most testing undertaking for showing signs of improvement yield. Hadoop gives HDFS to store those data in an structureddatabase. MapReduce algorithm empowers to make the structured distributed database from raw unstructured and semi-structured data. MapReduce has two essential class: the reducer class and the mapper class. Reducer class marge indistinguishable keyed data from per the algorithm guidance and give the structured database which store in HDFS. Mapper class take the data record as data and put a key with a comparing an incentive to the data record. At that point Map capacity mix those record, sort them and send to reducer class. This MapReduce algorithm design is appeared in Figure 2.

For applying the MapReduce system, let us consider HPC gives θ number of dynamic server appear in a set S = {S1,S2......Sθ}.

Algorithm: MapReduce Framework , Input : Unstructured or semi-structured data collection. Output: Create a structure data index.

Mapper function to create the pair of key and qualities

1. Input the data collection and take it in InData.
2. for each line segment and each row: do
3. Data=Strip and Split the line
4. If number of section in Data! = length Data.head: then
5. Skip that line from Data
6. Else
7. Store those data creating pair of key and qualities.
8. Go to the new line in Data.
9. end if
10. end for

Decrease function to create the structure data

11. Set oldKey = Null.
12. for each line section and each line: do
13. Data=Strip and split the line.
14. if number of segment in Data!=length Data.head:then
15. Skip that column from Data.
16. else on the off chance that any incentive from Data ==0, at that point
17. Skip that column from Data.
18. else
19. Store every one of the data in database as structure data.
20. end if
21. end for

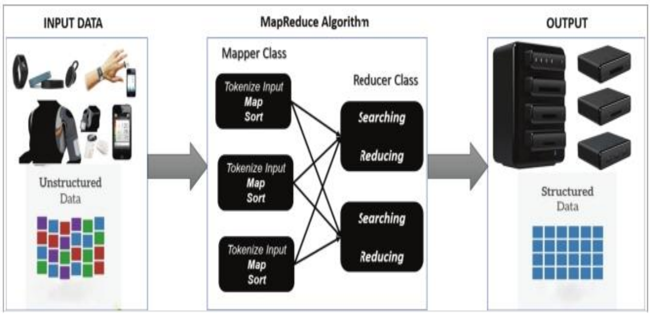

Figure 2. MapReduce Framework

IV. EXPERIMENTAL RESULT

Managing with big data and assess the result of the framework demonstrates the proficiency of framework. In this paper, we will examine the assessment of the prediction algorithm and furthermore demonstrates the examination between two datasets with applying different AI algorithm.

Table 1. Evaluation criteria for Diabetics prediction before implementing Genetic-Algorithm

	ANN	Random Forest	SVM	C5.0	Naive Bayes
Accuracy	76.47%	74.67%	75.98%	70.74%	72.92%
Kappa	49.33%	43.20%	45.71%	37.48%	41.42%
Sensitivity	80.85%	77.01%	77.43%	77.46%	78.23%
Specificity	68.75%	69.12%	72.30%	59.77%	63.41%
Precision	80.01%	85.51%	87.58%	75.86%	79.31%
Recall	80.84%	77.01%	77.43%	77.46%	78.23%
F1	81.41%	81.04%	82.20%	76.65%	78.86%
Misclassification Error	23.52%	25.32%	24.02%	29.25%	27.08%

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

Table 2. Evaluation criteria for Diabetics prediction after implementing Genetic algorithm

	ANN	Random Forest	SVM	C5.0	Naive Bayes
Accuracy	76.92%	73.80%	74.67%	69.43%	71.17%
Kappa	50.68%	39.66%	42.60%	35.48%	37.32%
Sensitivity	81.88%	74.85%	76.36%	77.37%	76.51%
Specificity	68.67%	70.68%	70.31%	57.60%	61.25%
Precision	81.29%	88.27%	86.89%	73.10%	78.62%
Recall	81.88%	74.85%	76.36%	77.37%	76.51%
F1	81.58%	81.01%	81.29%	75.17%	77.55%
Misclassification Error	23.08%	26.20%	25.32%	30.57%	28.82%

A. Performance Evaluation for Diabetics Prediction

In diabetic dataset arrangement, we anticipate the paired result where 0 signifies the patient who does not have diabetics, and 1 demonstrates the diabetes tolerant. In the wake of applying the prediction algorithm, we get various outcomes for hereditary and non-hereditary algorithm approach for various classifiers. An examination is appeared Table 1 and Table 2, Evaluation Criteria for Diabetics patient data which predict the evaluation assessment of all classifier for diabetic dataset when applying the hereditary algorithm. Table 3 and Table 4 speaks to the framework time which is taken to fabricate the model.

Table 3. Model Construction time for Diabetics Prediction table before genetic algorithm incorporated.

	ANN	Random Forest	SVM	C5.0	Naive Bayes
CPU Time	9.26	0.45	0.03	0.35	0.23
System Time	0.00	0.03	0.00	0.02	0.01
Elapsed Time	9.51	0.74	0.04	0.58	0.05

Table 4. Model Construction time for Diabetics prediction after genetic algorithm incorporated.

	ANN	Random Forest	SVM	C5.0	Naive Bayes
CPU Time	5.31	0.45	0.23	0.68	0.03
System Time	0.00	0.00	0.00	0.08	0.00
Elapsed Time	5.65	0.24	0.03	0.45	0.05

We get the most astounding exactness in the counterfeit neural system at the diabetic dataset, yet this model sets aside higher effort to build up the model. Subsequent to actualizing the hereditary algorithm, we can limit the model development time, yet it marginally diminishes the exactness of ANN where other model's precision is expanded. The precision and misclassification blunder graph in Figure 3 and Figure 4 the presentation assessment of diabetics patient appeared in Figure 5 and Figure 6.

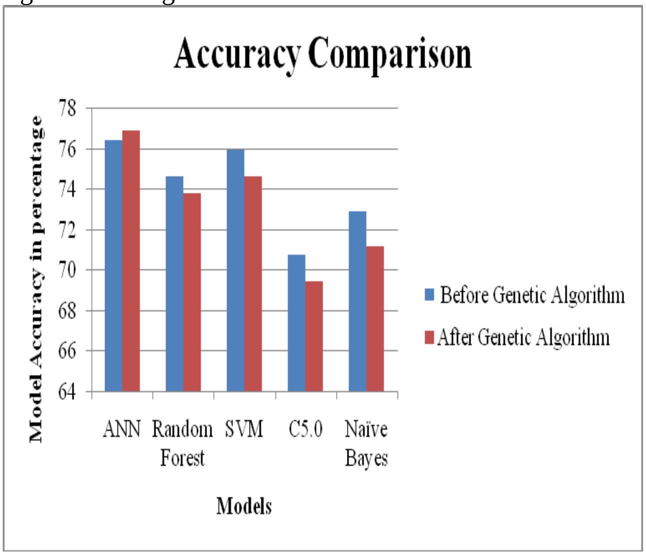

Figure 3. Model Accuracy for diabetic's prediction

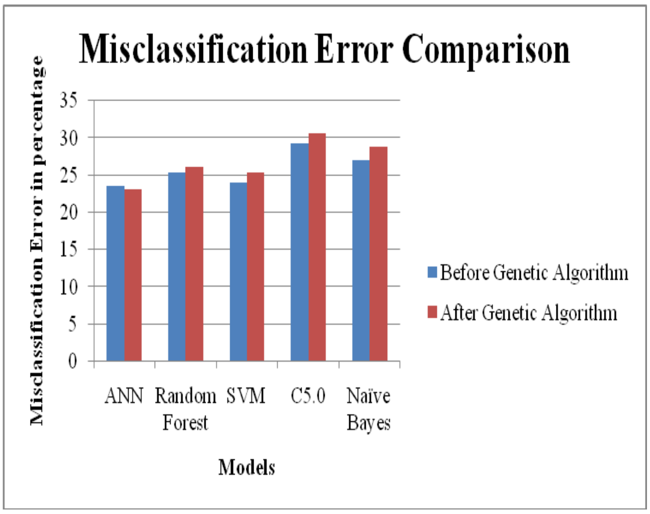

Figure 4. Model Misclassification Error (MCE) for diabetic's prediction

Published By:
Blue Eyes Intelligence Engineering & Sciences Publication (BEIESP)

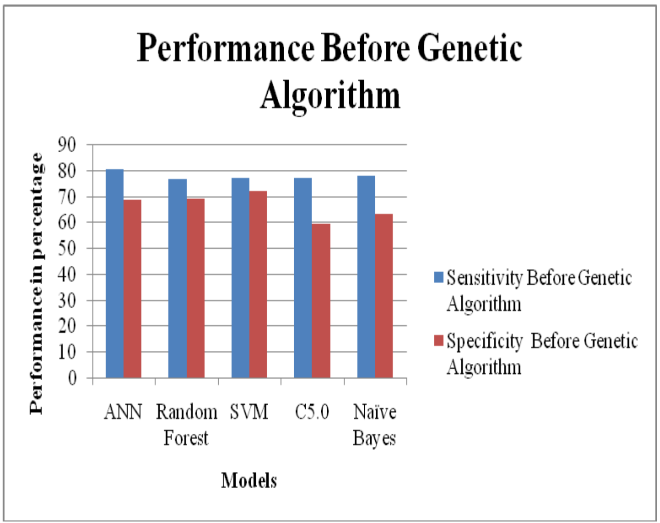

Figure 5. Performance evaluation for diabetics prediction before genetic algorithm

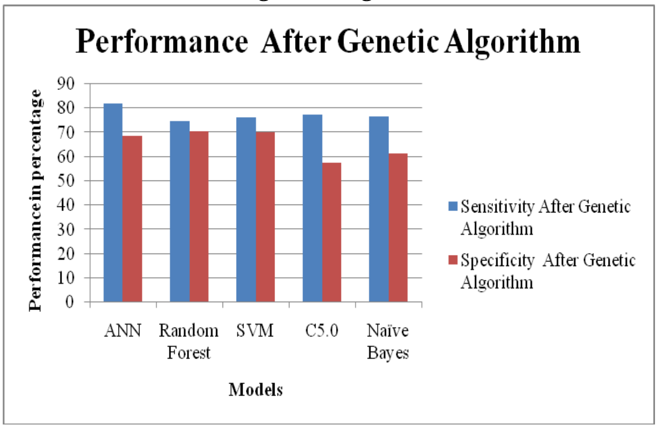

Figure 6. Performance evaluation for diabetics prediction after genetic algorithm

which can manage the large volume of patient's data along with the full outline of patient's medical conditions. Other than data accumulation, we have planned a probabilistic data obtaining plan which will examine the big measure of unstructured data and those plans are productive for a loud situation. A data warehouse is introduced to store data and empowers different capacities which can make two-way association with HPC and cloud server. In this paper, we have implemented some prediction model algorithm on existing dataset and demonstrated the exhibition of those models. We have demonstrated a side by side correlation of certain data mining techniques such as Artificial Neural Network, Random Forest, SVM, C5.0 and Naive Bayes classifier on healthcare data.

V. FUTURE WORK

Our future work is to build up an investigative model for wellbeing data representation utilizing Augmented Reality and Virtual Reality. Signal Procession and picture examination can uncover another measurement in the medical services industry which will permit a lot more highlights in an intelligent healthcare framework. Subsequently, managing sign and image data is another part of future endeavors.

REFERENCES

1. M. Zhao, F. Adib, and D. Katabi, “Emotion recognition using wireless signals”, Mobicom’16, October 2016.
2. M. Chen, Y. Ma, J. Song, C. F. Lai, and B. Hu, “Smart clothing: Connecting human with clouds and big data for sustainable health monitoring”, Mobile Netw. Appl., vol. 21, no. 5, 2016, pp. 825–845.
3. T. B. Murdoch and A. S. Detsky, “The inevitable application of big data to healthcare”, JAMA. 309 (13):, April 2013, p. 1352.
4. W. Raghupathi and V. Raghupathi, “Big data analytics in healthcare: promise and potential”, Health Inf. Sci. Syst., vol. 2, no. 1, 2014, pp. 1–10.
5. “The digital universe of opportunities: Rich data and the increasing value of the internet of things.” November. 2017. Available: https://www.emc.com/collateral/analyst-reports/idc-digital-universe-2014pdf
6. E. M. Mirkes, T. J. Coats, J. Levesley, and A. N. Gorban, “Handling missing data in large healthcare dataset: A case study of unknown trauma outcomes”, Computers in Biology and Medicine. 75: June 2016, pp. 203–216.
7. Ibm big data in a minute: Transforming unstructured data into better healthcare outcomes, August 26, 2014. Available: http://www.ibmbigdatahub.com/video/ibm-big-data-minute-transforming-unstructured-data-better-healthcare-outcomes accessed: 18. September. 2017.
8. C.McDonald, “5 big data trends in healthcare for 2017”, https://mapr.com/blog/5-big-data-trends-healthcare-2017/ accessed: 08. October. 2017, Feb. 13 2017, pp. 514–522.
9. D. L. Rizzatti,“Digital data storage is undergoing mind-boggling growth”, in http://www.eetimes.com/author.asp?section id=36&doc id=1330462 accessed: 18. September. 2017, Sept 9 2016.
10. J. O’Donoghue and J. Herbert, “Data management within mhealth environments: Patient sensors, mobile devices, and databases”, Journal of Data and Data Quality. 4 (1): 5:1-5:20, October 2012.
11. R. A. Hernandez, J. Dil, B. Fisher, and T. M. Green, “Visual analytics and human computer interaction”, interaction, January + February 2011.
12. O. Linda A. Winters-Miner, “Seven ways predictive analytics can improve healthcare,” 2014. [Online]. Available: https://www.elsevier.com/connect/seven-ways-predictive-analytics-can-improve accessed: 17 September 2017.
13. Y. Lee, H. Bang, and D. J. Kim, “How to establish clinical prediction models”, Endocrinology and Metabolism 31.1, 2016, pp. 38–44.
14. 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, 2016, pp. 6901–6909.
15. M. Chen, Y. Zhang, Y. Li, M. M. Hassan, and A. Alamri, “Aiwac: Affective in- teraction through wearable computing and cloud technology”, IEEE Wireless Commun. vol. 22, no.1, February 2015, pp. 20–27.
16. K. Dolui, S. Mukherjee, S. K. Datta, and V. Rajamani, “Retiha: Real time health advice and action using smart devices” , Proc. Int. Conf. Control Instrum. Com- mun. Comput. Technol. (ICCICCT), July 2014, pp. 979–984.
17. P. K. Sahoo, S. K. Mohapatra, and S. L. Wu, “Analyzing healthcare big data with prediction for future health condition”, IEEE Access, vol. 4, 2016, pp. 9786–9799.
18. I. Raharjo, T. G. Burns, J. Venugopalan, and M. D. Wang, “Development of user- friendly and interactive data collection system for cerebral palsy”, 2016 IEEE- EMBS International Conference on Biomedical and Health Informatics (BHI), Las Vegas, NV, 2016, pp. 406–409.
19. J. Andreu-Perez, C. C. Y. Poon, R. D. Merrifield, S. T. C. Wong, and G. Z. Yang, “Big data for health”, IEEE Journal of Biomedical and Health Informatics, vol. 19, no. 4, July 2015, pp. 1193–1208.
20. A. Das, T. Adhikary, M. Razzaque, M. Alrubaian, M. M. Hassan, Z. Uddin, and S. Biao, “Big media healthcare data processing in cloud: a collaborative resource management perspective”, Journal:Cluster Computing, March 2013, p. 160.
21. B. Boukenze, H. Mousannif, and A. Haqiq, “Predictive analytics in healthcare system using data mining techniques” , CS & IT-CSCP, 2016, pp. 1–9.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

22. K. Ng, A. G. S. R. Steinhubl, W. F. Stewart, B. Malin, and J. Sun, "Paramo: A parallel predictive modeling platform for healthcare analytic research using electronic health records", J Biomed Inform. 48:, April 2014, pp. 160–170.
23. M. Dey and S. S. Rautaray, "Study and analysis of data mining algorithms for healthcare decision support system", IJCSIT,Vol. 5 (1), 2014, pp. 470–477.
24. R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm, and N. Elhadad, "Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmis- sion", Proceeding KDD '15 Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,Sydney, Australia, 2015, pp. 1721–1730.
25. S. Ram, W. Zhang, M. Williams, and Y. Pengetnze, "Predicting asthma-related emergency department visits using big data", IEEE Journal of Biomedical and Health Informatics, vol. 19, no. 4, July 2015, pp. 1216–1223.
26. H. Wu, W. Pan, X. Xiong, and S. Xu, "Human activity recognition based on the combined svm & hmm" , 2014 IEEE International Conference on Data and Automation (ICIA), Hailar, 2014, pp. 219–224.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)