Influence of Data Mining Techniques in Healthcare Research

P.Chandrakala[1],Dr.A.Sumithra[2],Dr.A.Saranya[3], .Dr.R.Bagavathi Lakshmi[4]
Assistant Professor,G.T.N Arts College, Dindigul.
Associate Professor,SNS College of Technology,Coimbatore
Lecturer, Madurai Kamaraj University,Madurai
Assistant Professor, V.V.VanniaPerumal College for Women, Virudhunagar

Abstract
Healthcare research is essential to modern society because increasing of new diseases makes the people despondent. In the 21st century, all the hospital becomes computerized and keeps all the medical records and transactions in the repository. One of the core research domains like image processing give significant advancement in healthcare and even data mining also contributed as well. Data mining is the research area that extracts knowledge from the data repository. The mining algorithms like clustering and classification can be applied in many areas like brain tumor detection, cancer detection, heart dieses, hospital management, cell mutation, etc. This study paper explores such algorithms, architectures, and problems that are developed by various researchers. This paper focus on the results belongs to a brain tumor and advanced mining techniques on medical records.
Keywords: Healthcare, Brain Tumor Detection, Data Mining

I. Introduction

The ultimate objective of any research is for the betterment of daily life and sometimes it's could be future development or supportive one. The core researches like image processing in healthcare [1], green computing [2, 3], IoT [4], etc. are directly influenced in human life or environment. All the data are collected in various fields and then accumulated at a appropriate places.Computational theories and different types of tools to assist humans in extractingtheir useful information (knowledge) from the huge volumes of data. The process of data mining methods for pattern discovery and extraction [1]. Data mining is the process of analyzing the raw data and extracts it. The process is defined as the discovery of unknown information from large volumes of data [5]. Among the data mining techniques developed in recent years, the data mining methods are including generalization process, characterization, classification process, clustering method, association rule, evolution, pattern matching, data visualization, and meta-rule guided mining. [2].

The paper further organized into the following sections. Section II explains the role of data mining in healthcare industry. In this section outline the mining algorithms with the application on healthcare and it highlights the contribution of clustering and classifications in that industry. This section also discussed contribution of mining algorithm on brain tumor and patient states and other disease. It discusses how to extract knowledge from the social media andcloud based applications. The cloud based mining algorithms is vital and its give the detailed output data. Section III gives the conclusion of this paper.

II. Data mining in healthcare System

Healthcare focus elaborated processes of the diagnosis method, treatment processand hindrance of illness, injury has undertaken and different physical and mental impairments in humans [6]. The health care industry in top mostcountries is evolving at a fast. The health care industry will be place with huge data and then generate big amounts of information with electronic medical records, administrative reports and finding the different benchmarking [7]. This health care knowledge are however being under-utilized. Data mining is a process toget new and valuable data from these massive amount of information. Data mining in health care is predicting various diseases for diagnosis for the doctors in creating clinical decision making. The discussion on the various strategies areutilized in the health care industry is mentioned below:

The clustering could be a major task of finite set classes or clusters[8]. The datasets employed in this study were collected from patient's clinical method and laboratory results with k-mean,k-median,x-means. The analysis for every of the algorithms is conducted exploitation the Davies-Bouldin Index. The k-means obtained the most effective results whereas x-means obtained a good result while the k-mediods obtained the worst results. From the results researchers provides a various styles of patients having a higher possibility of probability.

There are 5 brain tumordetection approaches which includes clustering, classification, genetic algorithm, neural network, regiongrowing and thresholding approaches. In M.S. Atkinsand B.T. Mackiewich [10] made experiments for automatic image segmentation using a threshold.JaskiratKaur et al [11]It fetches with image segmentation and feature extraction for recognizing images to be carried out. H. D.Cheng et al [12]Mainly focus on thresholding concept method based on histogram and then objects are scaterred.

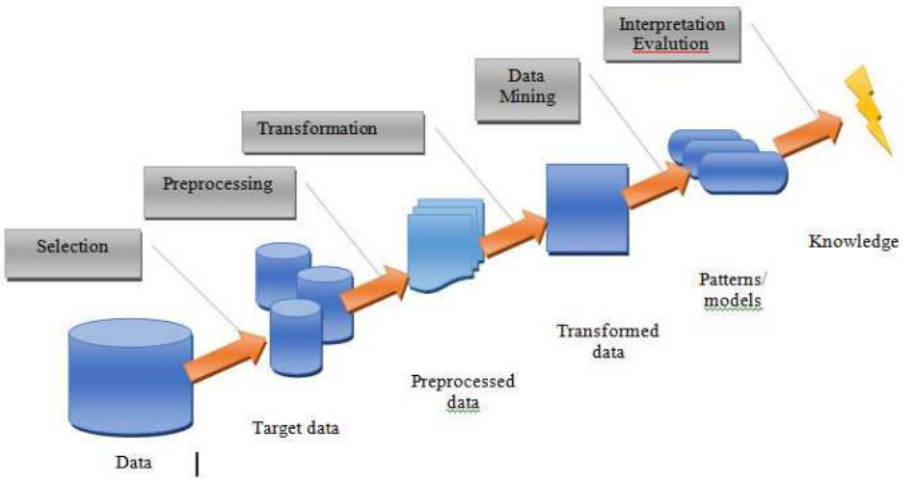

Fig 1 – Knowledge discovery of healthcare data

Duggal et al. [13] The comparison of numerous classification models which are predicted in hospital admission rate for some diabetic patients. It found the random forest (RF) formula [14] was the optimum classifier for some task. Strack et al. [15]The impact of glycatedhaemoprotein (HbA1c) measure on admission rate of some diabetic patients which are used to get the variable logistical regression on a dataset of patient diabetic from the records of a hundred thirty North American country hospitals. They all over that getting a measure of HbA1c for patients with diabetes could be a helpful predictor of admission rates.

Hachesu et al. [17] the data is collected with patients having coronary artery disease (CAD) which are used to have decision trees, support vector machines (SVM) and neural networks algorithms to predict the different LOS for these viscus patients. In their study, the result of SVMhas the best acceptable LOS. Mainly targeting LOS prediction for patients with diseases, there are studies it attempt to predict LOS for the patients who are hospitalized normally while not focused a particular sickness. Samples of this study embody Liu et al. [18] and Azari et al. [19].

B.V.Kiranmayee et al [20] proposes a unique design to notice brain tumor by using classifying the 2 medical image. This methodology works below 2 phases that's training and testing phases. Throughout training part, the scan pictures are load and designed the data set. Within the testing part the original image may be compared and test whether or not tumor found or not. The dataset may be designed through collected pictures from the web.

AymanAlahmar et al [21] projected a replacement methodology to predict long staying diabetics patients during a hospital. This formula used to classify the long keep patients as a result of they're prone to the hospital and also the algorithm used machine learning to predict. The dataset were collected from University of California because they maintain all diabetics' patients for very long time (table 1).

Table 1.Statisticson original dataset with selected data

Race	**Caucasian**	**African American**	**Hispanic**	**Asian**	**Others & Unknown**
	76008 (74.76%)	19211(18.87%)	2038(2.01%)	642 (0.64%)	3778(3.72%)
Time in Hospital	**Median**	**Average**	**Minimum**	**Maximum**	
	5 days	4.5 days	1.5 day	15 days	
Person Discharge Disposition	**Discharge to Home**		**Person Expired or Discharged to another healthcare facility, etc.**		
	60235 (59.18%)		41533 (40.82%)		

SopanGanpatSutar [22] proposed an interesting topic by applying data mining on social media. The drugs and other medical opinions are collected and discover knowledge from the social media users chat, post and other data. The mining algorithms are applied on the social media data and the information can visualize the actual states in the market. The architecture is shown in fig 2.

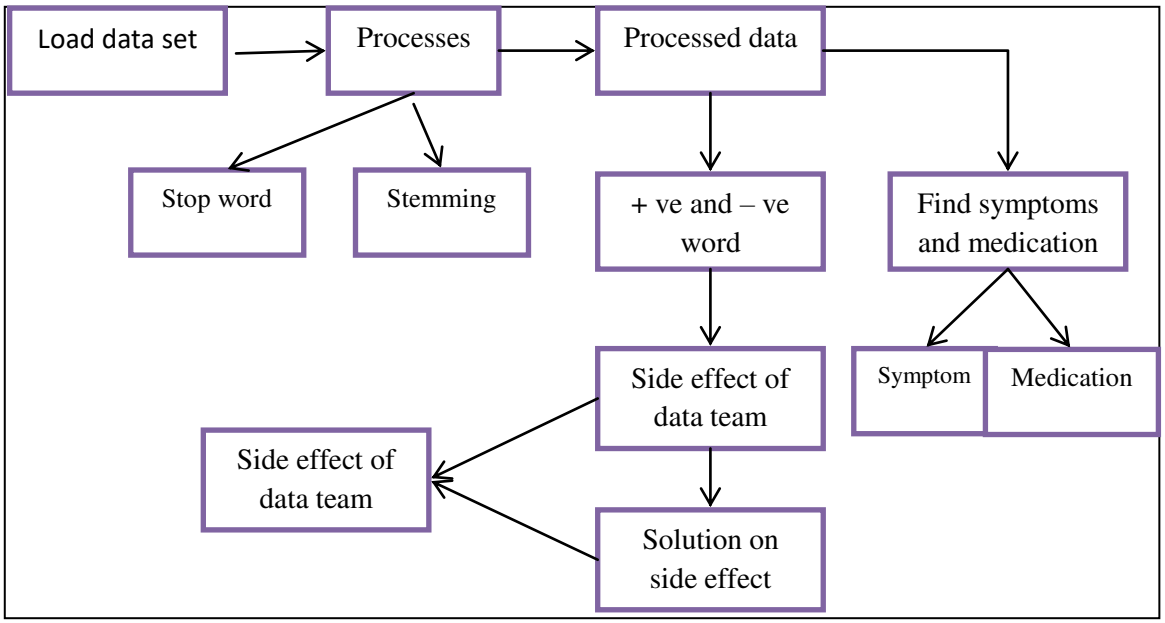

Fig 2. Mining social media data

Peng Zhang et al [24] propose to explore health care data via cloud-based healthcare data mining services; specifically focused on service development and other activities. Beneath such framework has to be develop a cloud-based health care data mining service to predict patient's length in hospital. During this projected framework,

1) population-level in healthcare data are scattered with local data sources are combined, that gives plentiful data for the information mining ;

2) Different computational approaches infrastructure and resources are often delivered by different cloud computing platforms in an exceedingly reliable, scalable, that satisfies the machine parts and financial aspects demand for building services in health care data mining;

3) The Method of service development is totally modularized, that makes the individual service development, to be very fast ,update, and maintenance;

El-Houssainy A. Rady et al [25] proposes a comparison result of chronic kidney disease. In this research the authors use Probabilistic Neural Networks(PNN), Multilayer Perceptron (MLP), Support Vector Machine (SVM) and Radial Basis Function (RBF) techniques to classify and predict the disease. The algorithms are applied in real-time dataset and the PNN will be the efficient algorithm to predict the disease.

Table 2 Overall classification accuracy rate in% and analysis execution time for each algorithms.

Algorithms	**Overall Accuracy Rate**	**Total Time for execution**
PNN	96.71%	0:00:13
SVM	60.71%	0:00:42
RBF	87.1%	2:29.7
MLP	51.56%	00:03.9

The Probabilistic Neural Network technique it gives a good classification results when compared with all techniques in classifying CKDstages (Figs. 3-7)

Fig 3.CKD Stage 1 Classification accuracy Rate in % for all algorithms.

Fig 4.CKD Stage 2 Classification accuracy Rate in % for all algorithms

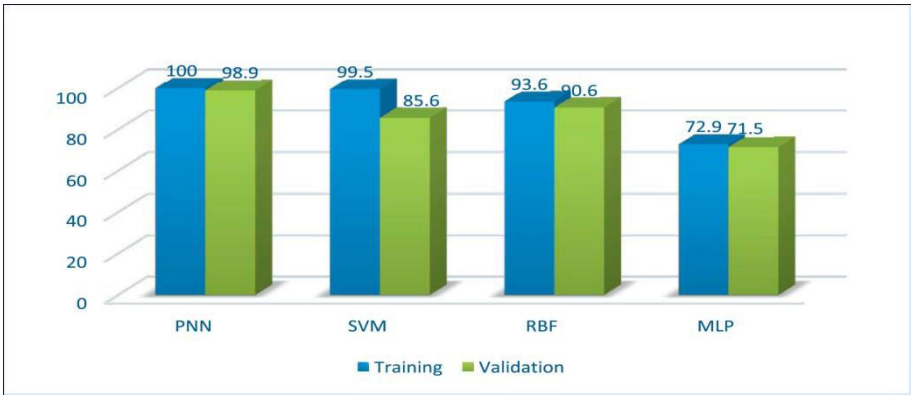

Fig 5.CKD Stage 3 Classification accuracy Rate in % for all algorithms.

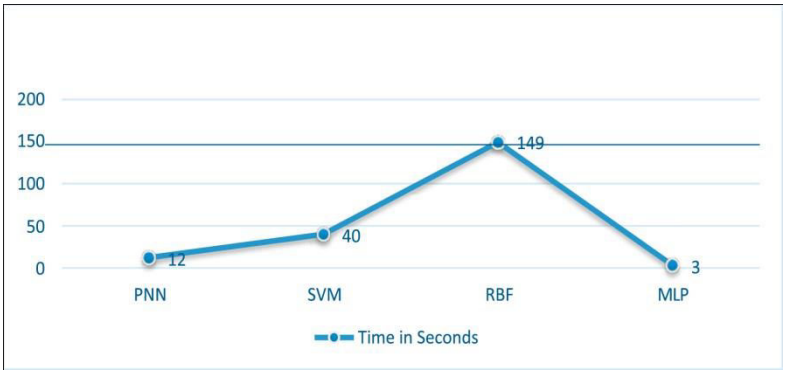

Fig 6.Overall Classification accuracy Rate in % for all used algorithms

Fig 7.Total Analysis taken for Execution time in seconds for all algorithms.

The research [26] aims to foundthe improvement of the accuracy of predicting cardiovasculardisease. Thisprediction models are developed by using different combination of prediction features, and sevenclassification techniques: k-NN, Decision Tree, Naive Bayes, Logistic Regression (LR), SupportVector Machine (SVM), Neural Network and Vote (a hybrid technique with Naïve Bayes andLogistic Regression).Final Experiment results shows the heart disease occurs in prediction model was developedusing the identified significant features and the best-performing data mining technique (i.e. Vote)achieves an accuracy of 87.4% in heart disease prediction (fig 8 –fig 10)

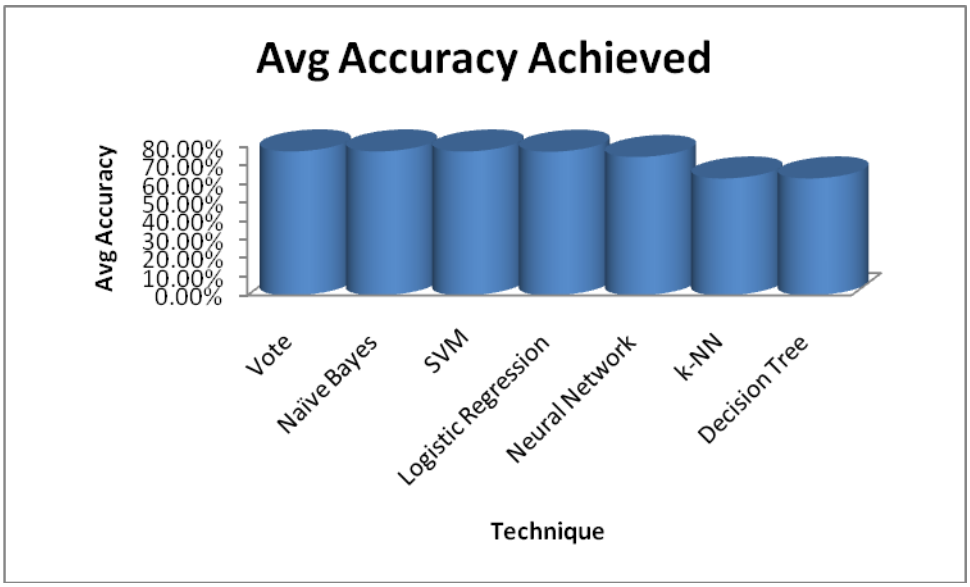

Fig 8. Average accuracy achieved by mining techniques

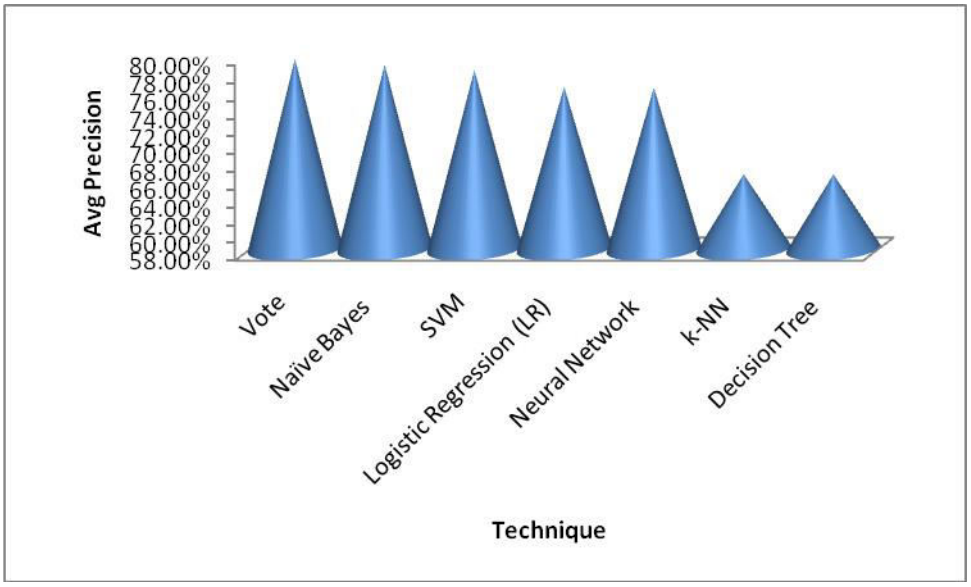

Fig 9.Average Precision Rate

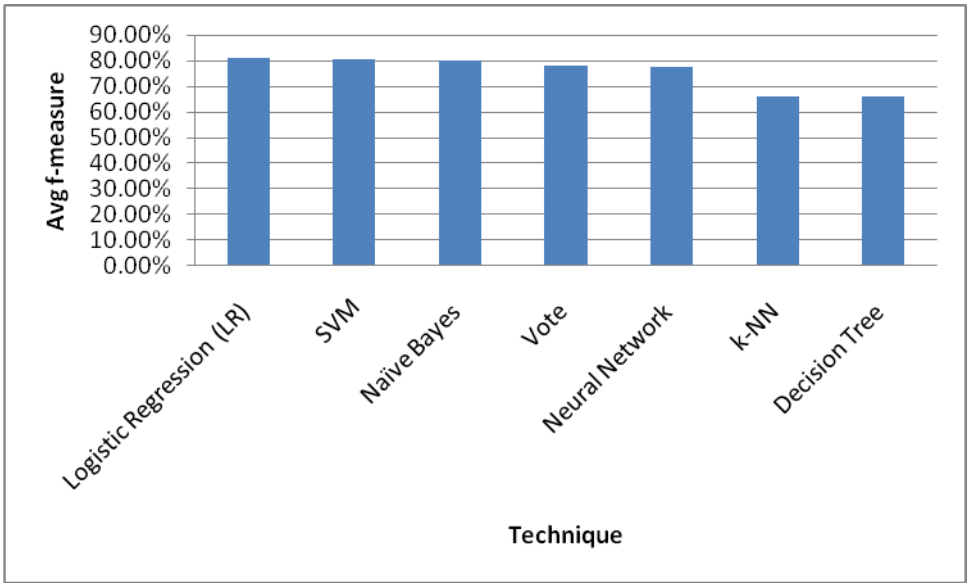

Fig 10.Average F-measure Frequency

III. Conclusion

Data mining is the key technique that applies in a variety of fields. Convergence on image processing, cloud-based, software engineering is significant but applying mining techniques in the healthcare industry is divine because it saves human lives. There is an open ground to apply mining algorithms in all areas of medical records. The records are uncountable and unimaginable so that it required the mining concept. This paper is an elaborate study of the result and the mining techniques that vary depend upon the dataset, size of dataset and application. The paper finds the some relavant characteristics in healthcare data sets are rich imbalanced data sets, so the major and the minor which are run by the classifier are not balanced and the result are prediction erroneous. At the end, the study highlights the result which already published and tested in real-time data and the suggestions have been made.

REFERENCES

[1] Tayade MC et al. "Role of image processing technology in healthcare sector: Review",International J. of Healthcare and Biomedical Research, Volume: 2, Issue: 3, April 2014, Pages 8-11.

[2] S.Pandikumar, M.Sumathi, "Energy Efficient Algorithm for High Speed PacketData Transfer on Smartphone Environment", International Journal of Engineering and Advanced Technology, Volume-8 Issue-6, August 2019.

[3]S.Pandikumar et al. "Principles and Holistic Design of Green Web Portal", International Journal of Computer Applications (0975 – 8887)Volume 65– No.9, March 2013.

[4] S.Pandikumar and RajappanVetrivel. "Internet of Things Based Architecture of Weband Smart Home Interface Using GSM." (2014), pp. 1721-1727.

[5] Milovic, B. (2011). Usage of DataMining in Making Business Decision.YU Info 2012 & ICIST 2012, (pp. 153-157).

[6] J.-J. Yang, J. Li, J. Mulder, Y. Wang, S. Chen, H. Wu, Q. Wang, and H. Pan, "Emerging information technologies forenhanced healthcare," Comput. Ind., vol. 69, pp. 3–11, 2015.

[7] N. Wickramasinghe, S. K. Sharma, and J. N. D. Gupta, "Knowledge Management in Healthcare," vol. 63, pp. 5–18, 2005

[8] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," AI Mag., pp.37–54, 1996.

[9] R. Veloso, F. Portela, M. F. Santos, Á. Silva, F. Rua, A. Abelha, and J. Machado, "A Clustering Approach for PredictingReadmissions in Intensive Medicine," Procedia Technol., vol. 16, pp. 1307–1316, 2014.

[10] M.S. Atkins and B.T. Mackiewich, 1 J.C. Bezdek. FullyAutomatic segmentation of the brain in MRI. IEEE T.Med.Imag., 17: 98–109.

[11] JaskiratKaur, Sunil Agrawal and RenuVig.; Comparative Analysis of Thresholding and Edge Detection Segmentation Techniques. International Journal of ComputerApplications 39(15):29-34, February 2012.

[12] H. D. Cheng, Y. H. Chen, and X. H. Jiang, Thresholding using two dimensional histogram and fuzzy entropyprinciple, وIEEE Trans. Image Processing, vol. 9, pp. 732-735, 2000.

[13] R. D. Duggal, S. Shukla, S. Chandra, B. Shukla, and S. K. Khatri,"Predictive Risk Modelling for Early Hospital Readmission ofPatients with Diabetes in India," International Journal of Diabetes inDeveloping Countries, 36(4), 2016, pp. 519-528.

[14] Dietrich, D., B. Heller, and B. Yang. "Data Science & Big DataAnalytics: Discovering, Analyzing, Visualizing and PresentingData,", John Wiley & Sons, 2015.

[15] B. Strack, J. P. DeShazo, C. Gennings, J. L. Olmo, S. Ventura, K. J.Cios, and J. N. Clore, "Impact of HbA1c Measurement on HospitalReadmission Rates: Analysis of 70,000 Clinical Database PatientRecords," BioMed Research International, vol. 2014, 2014.

[16] B. J. Lee, B. Ku, J. Nam, D. D. Pham, and J. Y. Kim, "Prediction ofFasting Plasma Glucose Status Using Anthropometric Measures forDiagnosing Type 2 Diabetes," IEEE Journal of Biomedical andHealth Informatics, 18(2), 2014, pp. 555-561.

[17] P. R. Hachesu, M. Ahmadi, S. Alizadeh, and F. Sadoughi, "Use ofData Mining Techniques to Determine and Predict Length of Stay ofCardiac Patients," Healthcare Informatics Research, 19(2), 2013, pp. 121–129.

[18] P. Liu, L. Lei, J. Yin, W. Zhang, W. Naijun, E. El-Darzi, "HealthcareData Mining: Predicting Inpatient Length of Stay," 3rd InternationalIEEE Conference on Intelligent Systems, 2006.
[19] A. Azari, V.P. Janeja, A. Mohseni, "Predicting Hospital Length ofStay (PHLOS): A Multi-Tiered Data Mining Approach," Data miningWorkshops (ICDMW), 2012, pp. 17-24.
[20]B.V.Kiranmayee et al, A Novel Data Mining Approach for BrainTumour Detection, International Journal of Computer Applications 39(18):129-140, Jan 2016.
[21] AymanAlahmar et al, Application of Data Mining Techniques to Predict the Length of Stay ofHospitalized Patients with Diabetes, 4th International Conference on Big Data Innovations and Applications pp. 38-44, 2018.
[22]S. G. Sutar, "Intelligent data mining technique of social media for improving health care," *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, 2017, pp. 1356-1360.
[23] F. Movahedi, Y. Zhang, R. Padman and J. F. Antaki, "Mining Temporal Patterns from Sequential Healthcare Data," *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, New York, NY, 2018, pp. 461-462.
[24] P. Zhang, S. Hu, J. He, Y. Zhang, G. Huang and J. Zhang, "Building Cloud-Based Healthcare Data Mining Services," *2016 IEEE International Conference on Services Computing (SCC)*, San Francisco, CA, 2016, pp. 459-466.
[25] El-Houssainy A. Rady et al,Prediction of kidney disease stages using data mining algorithms, Informatics in Medicine Unlocked, April 2019.
[26] Mohammad Shafenoor Amin et al, Identification of significant features and data mining techniques inpredicting heart disease, Journal of Telematics and Informatics, 2019.