Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 10, October 2021: 605-613

Performance analysis of various Map Reduce (MR) Models Using Internet of Things (IoT) Based Prediction of Cardiovascular Disease

[1]N.Rajesh Pandian, [2]D.Shanthi

[1]Assistant Professor, Department of Computer Science and Engineering,
PSNA College of Engineering and Technology, Kothandaraman Nagar, Dindigul, Tamil Nadu 624622
Email: nrajeshpandian@gmail.com
[2]Professor & Head, Department of Computer Science and Engineering,
PSNA College of Engineering and Technology, Kothandaraman Nagar, Dindigul, Tamil Nadu 624622
Email: dshan71@gmail.com

Abstract

In the modern era, Internet of Things (IoT) occupied the advanced features in the information and communication fields. The IoT platform is very useful for the human's healthcare monitoring in the various areas such as patient information management, real-time monitoring and healthcare management. According to the report, the mortality rate of (Cardiovascular Disease) CVD is more than 17.9 million globally for a year. Therefore the IoT data for CVD are very massive and huge in the size which is hard to access and extract information from it. The advanced Technique of Map Reduce (MR) model is used for a large set of data for the health predictions especially for the CVD. In this work, the survey of IoT based MR model is presented to observe the performance based accuracy and time consumptions. The basic of MR model is also discussed and then the overall performance of CVD Prediction is tabulated. The finest performance is determined from a survey in terms of accuracy, time consumption and error rate.

Keywords: **IoT, Map Reduce, Cardiovascular Disease, Accuracy**

I. INTRODUCTION

Heart diseases (HD) or a cardiovascular disease (CVD) is occurred in recent generations because of their unhealthy lifestyle and also some addictions like alcohol, smoking and enormous intake of fats. This lifestyle may also lead to various diseases like high blood pressure, hypertension, strokes and diabetics. According to the World Health Organization (WHO) [1], it is reported that the 17.9 millions of death that are caused globally for every year due to the CVD. The official statement of WHO confirmed that the good lifestyle and the earlier CVD prediction cause a maximum chance to reduce a CVD mortality rate. For the higher data analysis, the robust technology is required with a many features so that the system oriented techniques are processed.

The electronic system that is used to record all the data and collect the health oriented data for mining. Some of the tests that are carried for the CVD predictions are Electro cardiogram (ECG) and

Echocardiography (ECHO). These tests are supported for determining the hidden facts about the heart health and provide a decision making effectively. To reduce the CVD moratlity rate, the huge dataset for research is required. To obtain a huge dataset, the internet feature is required to collect all the sensitive data in a short time.

In recent time, internet is the major application that plays a vital role in the entire human's life. This internet is used in all the fields such as social media, health care, promotions, e-mails, educations and so on. In the medical care field, the application named as Internet of Thighs (IoT) is presented to reduce the medical error. An IoT is the technique that gathered information based on sensors which is attached with systems and devices. The statistic reported that the more than 50 billion of devices with sensors are accessed by internet.

The IoT is used to contribute in many applications which are connected a sensor data or networking between it namely disaster control [6], environmental control [5], smart system of cities [8] and health care [7]. The IoT is used to collect a data and provides a data in a huge quantity from a billion of Internet linked devices [10]. These IoT features are also used in the scholars societies [9] to provide an effective solutions and the predictions based on its data. In the medical field, the IoT data is collected from the various devices and to form a big dataset for the easy diagnosis.

This IoT based big data is applied for a Data mining which is used for easy and accurate predictions. The Data mining is an approach that is used to extract the knowledge for recognizing a different patterns and possible information from the big dataset using several algorithms. For accessing the IoT big data, the MapReduce (MR) method is used to process it. The MR algorithm is generally a parallel programming approach that can be handled. This method can be used to minimize the prediction issues of CVD from distributed and parallel programming interms of performance, error tolerance, balancing of load. Therefore the Mapreduce model is used to achieve a high scalability and reliability [11].

In this work, the review of Iot based CVD prediction using a MR model is presented. This work is mainly used to focus on the survey of efficient CVD predictions that are done by the MR model with various classifiers. The performances are tabulated with a various literatures to summarize the best methods and the accuracy. This paper is organized as the problem definition in the section 2 and the MR models are discussed in the section 3. In the section 4, the general description for the IoT based MR model is explained in it. The section 5 carries the literature survey methods that are done by the various researchers and also tabulated with its performances. Next the summary of the work is discussed in section 6 and finally the paper is concluded in the section 7.

II. PROBLEM STATEMENT

For the easy and accurate prediction of CVD diseases, the large set of data is so complicated to process it in a required time period. Further the maximum amount of time is required to extract the data from a big dataset. These are the main concerns that are needed for every conventional analysis of the data. Sometimes it may provide an incomplete or inaccurate prediction that reduces the robustness of data mining model.

By considering all these issues, the MR algorithm is suggested to provide a higher robust performance from a huge dataset. This MR model is used to minimize the task complexity by

applying a Machine Learning (ML) model to include in it. This Hybrid model of MR with ML techniques is obtained an average time consumption that may obtain a higher precise solution. Also this MR model can be provided a huge data storage and unlimited power processing [12].

III. MAP REDUCED (MR) MODEL

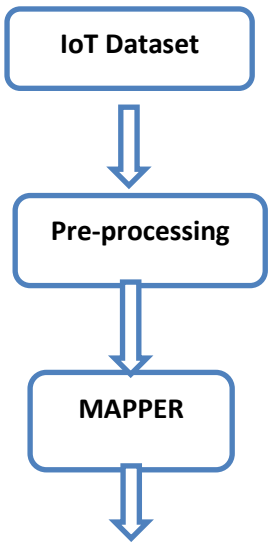
The MR model is used to access a set of hardware product in the various programming platform namely java or python. The MR model included a few significant functions where the first one is Map (M) function and another one is Reduce (R) function. The function of M is the process of combining a key-value for providing a few pairs of intermediate key-values. The function of R is used to participate an intermediate values completely that are relevant with its corresponding key. The beneficial model is programmed for a parallel form automatically and that can provide practical trade goods strategies. These are used to monitor and observe an input files division and organise to the inter-machine interactions. This function may provide a parallel and partition to obtain resources of large distributed systems quantity.

In the MR model, the Google file systems (GFS) are used as a primary storage and analyse an input data and stores an output data [14]. This GFS are based on the blocks and used to provide a system files. This may use to support for an error-tolerance that the data is utilized for a division and repetition. The external sources of MR model is the Apache Hadoop [15]. The Hadoop is a system that included a two section of layers such as Hadoop Distributed File System (HDFS) and Hadoop Map Reduce Framework (HMF). The HDFS is used to store a data and HMF is based on the information set.

The HDFS is a data analysis that can be done by a set of programming model to process a cluster formation. Each cluster node is inbuilt to the Hadoop that executes a Master Node (MN) or Slave Node (SN) model. The MN contains the job tracker node name and the SN provides both the Task Tracker and Data Node [17]. The task tracker function is to be done a more iteration in each cluster.

IV. IOT BASED MAP-REDUCE MODEL

In this section, the IoT based MR model is presented to predict the CVD in earlier stage. The general block diagram of the IoT based MR model is given in the figure 1. This block consists of IoT dataset, pre-processing, Mapper, Reducer, classifier and prediction solutions.



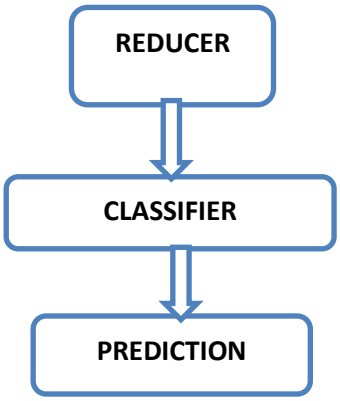


Figure 1 General block of IoT based MR model

From the figure 1, the initial step of the block is the data collection from the CVD which is based on the IoT devices. Next, the IoT big data are taken in to the process of pre-processing steps. This pre-processing is used to convert the dataset in to the accepted format and organise the data for a further steps. Then these data is given as input to Hadoop MR Model. This MR function is used to partition the data into different blocks. The MR model is attained into two stages namely Map stage that is followed by a next stage of Reduce stage. The next function is the Mapper phase that can be used to process an instance of each input and create a new pair of key-value.

The next block is a Reducer stage that can be used to receive an input from the mapper. The pair of intermediate key-value is given as input to the reducer. This reducer is the final task of the HDFS methodology that organise a similar key values and form a same group. This reducer output is isolated in HDFS and send to the classifier for the prediction. Next the output of hadoop reducer is given as an input to the classifier. The classification method is based on the Machine Learning (ML) and Deep Learning (DL) methods are used for the prediction. Some of the ML methods are Random forest, Naïve baiyes, Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Convolution Neural Network (CNN), Long Short Term Method (LSTM) and so on. By using these kind of advanced learning methods, the output for a CVD prediction can be predicated with a better accuracy. The results are executed and shown in the prediction stage in terms of various parameters such as precision, accuracy, F1 score respectively.

V. RELATED STUDY

In this section, the related literatures based on the IoT based MR model are discussed in the following. Syed Umar et al. [18] have presented a Back Propagation (BP) methodology to test and learn the Neural Network (NN). To obtain an optimal result the genetic algorithm (GA) is implemented with a NN. This method is well performed to predict a HD and also avoided a risk factor based on mortality. The NN Weight and Bias are used to optimise with GA and provided accuracy in the prediction as 96.2% in training set and 89% for testing set.

Shilaskar et al [19] have developed a hybridized forward selection (FS) method for CVD diagnosis. This method is used for a Features selection that is ranked with distance evaluation. These distances are calculated on measuring a forward selection, forward inclusion, and backward elimination to

determine the subset. Thus the accuracy of classification is improved and further increases the accuracy of diagnosis compared to forward inclusion and back-elimination techniques.

Purusothama et al. [20] have presented a several classification methods for HD prediction. This work has two models that are presented namely primary model or single model and secondary model or combined model. These primary and secondary are hybrid to train the data and predict the accurate result for HD. These models are compared with the methods of decision Tree (DT) method, KNN, association rule, ANN and Naive Bayes (NB). This method has a better accuracy of 96% and good performance to obtain a HD diagnosis.

Miao et al [21] developed an Ensemble ML based learning and prediction model from the dataset of Cleveland Clinic Foundation (CCF), Long Beach Medical Centre (LBMC), Hungarian Institute of Cardiology (HIC) and Switzerland University Hospital (SUH) for CVD diagnosis. The accuracies of CCF, LBMC, HIC and SUH are 80.14%, 77.78%, 89.12% and 96.72% respectively.

Dbritto et al. [22] explored an efficient medical data based decision support model depend on data mining. This model is focused to predict an appropriate classifier of ML such as NB, SVM and Logistic Regression (LR). These models obtained accuracy with 75%, 80% and 79% respectively. KaanUyar et al [23] discussed a natural selection model with a NN for CVD prediction. It is observed a 297 patient data and trained a 252 and tested 45 from it. This model attained an accuracy as 95.78%and also in future it could be combined with a meta-heuristic model.

Vishruti et al [24] presented a Hadoop based conventional Relational Database Management System (RDBMS). This RDBMS was a large dataset that may require a huge time to extract. To overcome these time consumption, the Hadoop is applied to extract the data which is provided an effective prediction in a minimum time. Some of the tool based on hadoop is applied in it such as Map Reduce, Hive and Pig.

Z. Lu et al [25] discussed a MR model that can provide an effective model to handle a large dataset. This work can be accessed and verified using the IoTDeM, this model can be reduced the time consumption by including a MR model. It also obtained a low error rate 10% from the previous work. Nagamani et al [26] described a MR model for CVD prediction with the mortality and morbidity rate. This model mainly focused on a earlier detection where the HD is in the fatal stage. The data mining is used to process the training and learning for prediction. The MR model result is compared with the ML methods of SVM, KNN, DT, NB and NN. The overall performance of MR model is better in accuracy of 98.12%.

Mylsami et al [27] presented the MR model for a massive web data to develop and predict in minimum time intervals. Therefore these big data is suggested to handle with an efficient MR model. This model is used to arrange, organise and sort the difficulties that are occurred in web. It also used to provide results from the web in an effective manner. Fathimabi et al [28] presented a MR model based HD Prediction System (MRHDP). The DT is included with a MR model using the Hadoop platform. The MR models are used for distributed algorithms, and handle massive data sets. The presented MRHDP model are exhibited a better prediction results for HD and showed a better execution with minimum time versatility and productivity.

Gupta et al [29] developed an ML based model to detect heart diseases. This method presented the KNN to detect the HD with the comparison of other methods like RF, DT, SVM and NB. The method is used a set of sensors for the prediction of HD. Thus, the KNN method provided a health based monitoring with a higher accuracy of 88.52% respectively.

Ganesan et al [30] presented an efficient CVD prediction based on IoT data using ML that are collected from the UCI dataset. The sensors are used to predict the disease and collected the data and predict the HD. Therefore, the classification methods are provided for the patient data classification to identify the CVD. The classifier is used to train the data from benchmark dataset in training phase and an actual patient data is to determine the CVD prediction in testing phase.

Verma et al [31] explored a CVD model for monitoring and diagnosing using an IoT based cloud data. This model is used to predict the seriousness of the HD. This model created the health estimation for the client based health evaluations. The classification model is achieved a parameters that are carried in terms of F-measure, specificity as well as sensitivity.

Gelogo et al [32] developed IoT data that are collected a data of human health. This data are accessed by the u-healthcare service. Therefore the IoT based healthcare services are presented another structure that is helpful for the IoT based u-healthcare service.

Gope et al [33] described Body Sensor Network (BSN) based IoT therapeutic devices that are used to collect a data. These data are collected by sensors where the patients are checked with a various powered sensors. The Healthcare based IoT is used to observe the health data and it is also included a watermarking method. This method is used for improving a signal for a higher prediction and execute in a less time 0.007655 ms.

Table 1. Performance analysis of various MR model

Si.No	Author	Method	Determination	Performance
1	Syed Umar et al [18]	Hybrid NN based GA	CVD prediction	Accuracy rate as 96.2%
2	Shilaskar et al [19]	forward selection (FS)	CVD prediction	Better than forward inclusion, and backward selection
3	Purusothaman [20]	Primary and secondary model	CVD prediciton	Accuracy 96 % with the ML techniques
4	Vishruti et al [24]	MR Hadoop based RDBMS	Heart care system	effective prediction in a minimum time
5	Z. Lu et al [25]	MR based IoTDeM	Big data extraction	Minimum time consumption and low error rate as 10%
6	Nagamani et al [26]	MR model	CVD prediction	Accuracy rate of 98.12% than the ML methods
7	Mylsami et al [27]	MR model	Massive data	Consume a

			accessing	minimum time, sort out web page errors
8	Fathimabi et al [28]	MRHDP	CVD prediciotn	Better accuracy, minimum time versatility and productivity.
9	Gupta et al [29]	KNN	CVD prediction	Accuracy rate of 88.52%
10	Ganesan et al [30]	IoT platform based ML	CVD prediction and diagnosis	Efficient and fast classification using IoT
11	Verma et al [31]	IoT based cloud	CVD and Health care	F-measure and specificity are higher
12	Gelogo et al [32]	IoT data	Healthcare	Effective and better accuracy
13	Gope et al [33]	Body Sensor Network based IoT	Healthcare	Less time consumption of 0.007655ms

VI. CONCLUSION

In this work, the survey of IoT based MR model for CVD prediction is presented. According to the survey of previous work for CVD prediction and health care system, the MR based IoT can be achieved and processed in a required time and accuracy. From the results, the MR model can be achieved a maximum accuracy of 98.12% in [26] and the IoT data is performed well and also extract a massive dataset in a less time and low error rate of 10% that stated in [25]. Therefore the IoT based MR model is efficient and better in accuracy, low error rate, minimum time consumption for the CVD and healthcare prediction.

REFERENCES

1. World Health Organization. Cardiovascular Diseases. http://www.who.int/cardiovascular_diseases
2. Qi, J., Yang, P., Min, G., Amft, O., Dong, F., & Xu, L. (2017). Advanced internet of things for personalized healthcare systems: A survey. Pervasive and Mobile Computing, 41, 132–149. https://doi.org/10.1016/j.pmcj.2017.06.018
3. Marjani, M. et al. (2017). Big IoT data analytics: Architecture, opportunities, and open research challenges. IEEE Access, 5, 5247–5261. https://doi.org/10.1109/ACCESS.2017.2689040
4. Suzuki, L. R. (2017). Smart cities IoT: Enablers and technology road map. In. Springer Optimization and Its Applications. Springer, 167–190. https://doi.org/10.1007/978-3-319-61313-0_10
5. Talavera, J. M., Tobón, L. E., Gómez, J. A., Culman, M. A., Aranda, J. M., Parra, D. T., Quiroz, L. A., Hoyos, A., & Garreta, L. E. (2017). Review of IoT applications in agro-industrial and environmental fields. Computers and Electronics in Agriculture, 142, 283–297. https://doi.org/10.1016/j.compag.2017.09.015
6. Sinha, A., Kumar, P., Rana, N. P., Islam, R., & Dwivedi, Y. K. (2019). Impact of internet of things (IoT) in disaster management: A task technology fit perspective. Annals of Operations Research, 283(1–2), 759–794. https://doi.org/10.1007/s10479-017-2658-1

7. Elhoseny, M., Ramírez-González, G., Abu-Elnasr, O. M., Shawkat, S. A., Arunkumar, N., & Farouk, A. (2018). Secure medical data transmission model for IoT-based healthcare systems. IEEE Access, 6, 20596–20608. https://doi.org/10.1109/ACCESS.2018.2817615
8. Mehmood, Y., Ahmad, F., Yaqoob, I., Adnane, A., Imran, M., & Guizani, S. (2017). Internet-of-things-based smart cities: Recent advances and challenges. IEEE Communications Magazine, 55(9), 16–24. https://doi.org/10.1109/MCOM.2017.1600514
9. Musleh, A. S., Debouza, M., & Farook, M. (2017). Design and implementation of smart plug: An Internet of things (IoT) approach. In International Conference on Electrical and Computing Technologies and Applications (ICECTA) (pp. 1–4). IEEE Publications.
10. Munir, A., Kansakar, P., & Khan, S. U. (2017). IFCIoT: Integrated Fog Cloud IoT: A novel architectural paradigm for the future Internet of Things. IEEE Consumer Electronics Magazine, 6(3), 74–82. https://doi.org/10.1109/MCE.2017.2684981
11. 4. S, Gomathy, V., Sheeba Rani, S., & Sujatha.K. (2018). Bhagavathy. Early Heart Disease Detection Using Data Mining Techniques with Hadoop Map reduce, International Journal of Pure and Applied Mathematics, 119(12), 1915–1920.
12. Kamilaris, A., Kartakoullis, A., & Prenafeta-Boldú, F. X. (2017). A review on the practice of big data analysis in agriculture. Computers and Electronics in Agriculture, 143(September), 23–37. https://doi.org/10.1016/j.compag.2017.09.037
13. Regha, S., & Manimekalai, M. (2015). Hadoop VS big data.
14. Lu, Z., Wang, N., Wu, J., & Qiu, M. (2018). IoTDeM: An IoT Big Data oriented Map Reduce performance prediction extended model in multiple edge clouds. Journal of Parallel and Distributed Computing, 118, 316–327. https://doi.org/10.1016/j.jpdc.2017.11.001
15. White, T. (2010). Hadoop: The definitive guide, Yahoo! Press.
16. Khanam, Z., & Agarwal, S. (2015). Map-reduce implementations: Survey and performance comparison. International Journal of Computer Science and Information Technology (IJCSIT), 7(4), 119–126. https://doi.org/10.5121/ijcsit.2015.7410
17. Buda, T. S., Salama, H. A. A. A., Xu, L., O'Sullivan, P. J., Thorpe, C., & Almeida, L. Data replication in a distributed file system. Google Patents 26-Sep-2019.
18. Amin, S. U., Agarwal, K., & Beg, R. (2013). Genetic neural network based data mining in prediction of heart disease using risk factors, IEEE Conference in Information and Communication Technologies (ICT), 1227–1231.
19. Shilaskar, S., & Ghatol, A. (August 2013). Feature selection for medical diagnosis: Evaluation for cardiovascular diseases. Expert Systems with Applications, 40(10), 4146–4153. https://doi.org/10.1016/j.eswa.2013.01.032
20. Purusothaman, G., & Krishnakumari, P. (2015). A survey of data mining techniques on risk prediction: Heart disease. Indian Journal of Science and Technology, 8(12), 1. https://doi.org/10.17485/ijst/2015/v8i12/58385
21. Miao, K. H., Miao, J. H., & Miao, G. J. (2016). Diagnosing coronary heart disease using ensemble machine learning. International Journal of Advanced Computer Science and Applications, 7(10), 30–39.
22. Dbritto, R., Srinivasaraghavan, A., & Joseph, V. (2016). Comparative analysis of accuracy on heart disease prediction using classification methods. International Journal of Applied Information Systems–ISSN, 11(2), 22–25. https://doi.org/10.5120/ijais2016451578
23. KaanUyar, A.İ. 'Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks', 9th.
24. Vishruti, P., & Ingle, M. D. (2018). Disease analytics in healthcare system using Hadoop. Int. J. Mod. Trends Sci. Technol., 4.
25. Lu, Z., Wang, N., Wu, J., & Qiu, M. (2018). IoT DeM: An IoT Big Data oriented Map Reduce performance prediction extended model in multiple edge clouds. Journal of Parallel and Distributed Computing, 118, 316–327. https://doi.org/10.1016/j.jpdc.2017.11.001
26. Nagamani, T., Logeswari, S., & Gomathy, B. (2019). Heart disease prediction using data mining with Mapreduce algorithm. p.p. no. 3, 137–140.
27. Mylsami, T., & Shiva kumar, B. L. (2017). Role of Map reduce Algorithm to Improve the Web data Access, 5(11) (pp. 47–51).

28. Fathimabi, S., Jangam, E., & Srisaila, A. (2021). Map Reduce based Heart Disease Prediction System 8th International Conference on Computing for Sustainable Global Development (INDIA Com), 2021 (pp. 281–286). https://doi.org/10.1109/INDIACom51348.2021.00049
29. Gupta, A., Yadav, S., Shahid, S., & V. U. (2019). Heart Care: IoT based heart disease prediction system International Conference on Information Technology (ICIT), 2019 (pp. 88–93). https://doi.org/10.1109/ICIT48102.2019.00022
30. Ganesan, M., & Sivakumar, N. (2019). IoT based heart disease prediction and diagnosis model for healthcare using machine learning models IEEE International Conference on System, Computation, Automation and Networking (ICSCAN), 2019 (pp. 1–5). https://doi.org/10.1109/ICSCAN.2019.8878850
31. Verma, P., & Sood, S. K. (2018). Cloud-centric IoT based disease diagnosis healthcare framework. Journal of Parallel and Distributed Computing, 116, 27–38. https://doi.org/10.1016/j.jpdc.2017.11.018
32. Gelogo, Y. E., Hwang, H. J., & Kim, H. (2015). Internet of things (IoT) framework for u-healthcare system. International Journal of Smart Home, 9(11), 323–330. https://doi.org/10.14257/ijsh.2015.9.11.31
33. Gope, P., & Hwang, T. (2016). BSN-Care: A secure IoT-based modern healthcare system using body sensor network. IEEE Sensors Journal, 16(5), 1368–1376. https://doi.org/10.1109/JSEN.2015.2502401
34. Hussain, A., Wenbi, R., da Silva, A. L., Nadher, M., & Mudhish, M. (2015). Health and emergency care platform for the elderly and disabled people in the smart city. Journal of Systems and Software, 110, 253–263. https://doi.org/10.1016/j.jss.2015.08.041
35. Yeh, K. (2016). A secure IoT-based healthcare system with body sensor networks. IEEE Access, 4, 10288–10299. https://doi.org/10.1109/ACCESS.2016.2638038
36. Chen, M., Hao, Y., Hwang, K., Wang, L., & Wang, L. (2017). Disease prediction by machine learning over big data from healthcare communities. IEEE Access, 5, 8869–8879. https://doi.org/10.1109/ACCESS.2017.2694446
37. Boukenze, B., Mousannif, H., & Haqiq, A. (2016). Predictive analytics in healthcare system using data mining techniques. Computer Science and Information Technology, 01–09.
38. Mitra, S., Bande, S., Kindle, S., & Kulkarni, A. (2014). Efficient FP growth using Hadoop—(improved parallel FP-growth). International Journal of Scientific and Research Publications, 4(7), 1–3.
39. Qiu, L., & Cheng, X.-H. (2012). The research of decision tree mining based on Hadoop. Proceedings of the of the 9th IEEE International Conference on Fuzzy Systems and Knowledge Discovery (pp. 798–801).
40. Kumar, D. S., Sathyadevi, G., & Sivanesh, S. (2011). Decision support system for medical diagnosis using data mining. International Journal of Computer Science Issues, 8(3), 147–153.