IJERMS

International Journal of Engineering Researches and Management Studies

.

PREDICT CHRONIC KIDNEY DISEASE USING DATA MINING ALGORITHMS IN HADOOP

Guneet Kaur*[1] & Er. Ajay Sharma[2]

*[1]M.Tech Student, Department of Computer Science and Engineering, Amritsar College of Engineering and Technology, Amritsar

[2]A.P. of CSE dept. Amritsar College of Engineering and Technology, Amritsar

ABSTRACT

This paper introduced the chronic kidney disease prediction with data mining algorithms. In the 21st century, chronic kidney diseases are growing so rapidly and it plays a key role in an individual's life. To obtain the hidden information from the given dataset, data mining is used to make the decisions. Big data is the latest technology used to store and process the voluminous data and that data can be structured data, unstructured data and semi-structured data . In this paper, to predict or detect the chronic kidney disease , KNN (K-nearest neighbor) and SVM (Support Vector Machine) data mining algorithms are used. From the given dataset, six statistical parameters are generated which are as follows:

i. Accuracy %
ii. Error %
iii. Precision
iv. Recall
v. F1 error
vi. Elapsed Time

The whole research has done in the layered form in order to enhance the above statistical parameters that have mentioned to predict the chronic kidney disease. MATLAB is a tool used to perform to prediction of chronic kidney disease by accessing Hadoop in itself.

Keywords: Big Data , Chronic Kidney Disease , Data Mining , Hadoop , KNN , MATLAB , SVM

Nomenclature:

ANN	Artificial Neural Network
CKD	Chronic Kidney Disease
KNN	K- Nearest Neighbor
MATLAB	Matrix Laboratory
SVM	Support Vector Machine

1. INTRODUCTION

In present scenario, computers have enhanced the technology in a way that leads to the formation of very large amount of data. In real time applications, we can easily have this much of data in a huge form that can further be used in the functionality of any social networking site or elsewhere. Data mining is an area of research to identify the necessary hidden information from a dataset that is being used. Now a days, many medical databases have created on the basis of the advancement of the healthcare database management system. Complex medical data can easily be maintained by using data mining techniques.

Data mining has discovered or either emerged the non-identical direction in the healthcare organization. Data can be available in many formats like text, signals or graphics.

IJERMS
International Journal of Engineering Researches and Management Studies

Data mining may have many approaches such as Naive Bayes , J48 , KNN , SVM , Supervised or Unsupervised learning and many more approaches. Data mining here , comes up with a number of technique that gives knowledge to the healthcare industry which when applied to the processed data. Domains of data mining are used for decision making. It acts as an integral part of Big Data in the current time.

In data science , Big data is the advanced technology that plays a vital role in analyzing the medical data. Big Data is the growing aspect of medical sciences. Data can be structured (strict format) data , semi-structured (XML documents) data and unstructured (simple documents as text) data.

Chronic Kidney Disease is referred to Chronic Renal Failure or Chronic Kidney Failure. It is developed in a crucial way in which functionality of kidney goes down and is a slow progressive process usually takes place in a continuity. It remains over a several years. When this disease occur in human's body what happens is that kidneys got damaged and in turn toxics cannot be filtered out easily from our body. In most of the cases , people realize this disease only when the function of kidney lowers down by 25% as compared to the normal functionality of the kidneys. Therefore, chronic kidney disease acts as a global issue in an individual's life.

In the existing work , to maintain the accuracy , KNN and Naive Bayes data mining algorithms are ignored to attain the best results. Even for the process of feature extraction , no description method is ye used by any of the researcher. To achieve the elapsed time , dimensionality reduction algorithms like as PCA (Principal Component Analysis) and ICA (Independent Component Analysis) are not in usage.

In our proposed work , KNN and SVM data mining algorithms are applied in MATLAB by accessing hadoop in itself to predict the chronic kidney disease.

These two algorithms deals with the statistical parameters in a way to give accuracy percentage , error percentage , precision , recall , F1 error and elapsed time taken by a specific dataset. For classify the data KNN is used and SVM is used for the feature extraction.

Technology Used

MATLAB and Hadoop are the technologies which are used in our work for predicting the chronic kidney disease in human's body . In dataset , number of input parameters resides in it . To attain the objective of predicting the chronic kidney disease , MATLAB itself is accessed by hadoop.

1) MATLAB :

MATLAB is a fourth generation programming language developed by Cleve Molar at the University of New Mexico. It has started in the late 1970's. It is accessed by engineering departments, sciences and economics in many ways. In present era, MATLAB is used in education, linear algebra, numerical analysis and in image processing also.

MATLAB was first developed by researchers and then used by the practitioners in control engineering. MATLAB provides an interactive environment for problem solving and design. It provides various tools for building an applications with custom graphical interfaces.

2) Hadoop:

Hadoop is the technology used to process the Big Data. Hadoop is an open-source framework written in java. Hadoop is founded by the Apache foundation. It was developed by Doug cutting , from his kid's yellow colored elephant toy. Data can easily be distributed across multiple clusters using Hadoop in Big Data. Therefore, it provides a large storage for the distributed data.

For dialysis, diagnosis and analysis of a chronic kidney disease dataset, MATLAB is used in the Big Data. There are variety of modules present in MATLAB used to mine the data by applying data mining algorithms. Prediction is the initial step for the identification of symptoms and to classify the data or to detect the chronic kidney disease.

IJERMS
International Journal of Engineering Researches and Management Studies

2. LITERATURE SURVEY

Many classifiers of data mining are used by different researchers to predict and detect the chronic kidney disease problem. But no data mining classifier is used by any of the researcher in the Big Data research area to predict the chronic kidney disease.

A.Sai Sabitha et al.(2016) applied ANN(Artificial Neural Network) and Naive Bayes data mining approaches. In their work, Rapid miner tool is used for the prediction and diagnosis of chronic kidney disease. Results comes out as 100% accuracy is possessed by Naive Bayes and 72.73% accuracy is possessed by Artificial Neural Network.

M.Archana Bakare et al.(2016) used Multibank algorithm for the prediction of asthma disease. Many mining techniques are also used fir this purpose. In their results, 80% accuracy has obtained.

Prasan Kumar Sahoo et al.(2016) attained 98% accuracy. They predicted asthma and cancer. Cloud type environment is created in their proposed work of the paper named as " Analyzing healthcare big data with the prediction for future health condition".

Lambodar Jena et al.(2015) applied J48, Naive Bayes, SVM(Support Vector Machine), Multi Layer Perceptron, conjunctive Roole and decision table to diagnose the chronic kidney disease. WEKA data mining tool is used in their proposed work and they observed that multi layer perceptron attained more accuracy.

Anu Chadhary et al.(2014) predicted heart and kidney failure disease. They applied K means and Apriori algorithms with 42 attributes from their dataset. In their experimental results, distribution statistics and machine learning tools are analyzed.

Dr.S.Vijayarani et al.(2014) applied Naive Bayes and SVM(Support Vector Machine) classifiers for the heart disease prediction. In their experimental results, performance of SVM is better than Naive Bayes.

Tommaso Di Noia et al.(2014) developed a software tool and used ANN(Artificial Neural Network) classifier which is used to predict the end stage probability of a disease. They predicted the ESKD(end stage kidney disease). In 38 years, at the University of Bari, this research provided with ten networks and can also be utilized as an android phone applications or an online web applications. This research plays a vital role in healthcare industry.

Solanki A.V.et al.(2014) predicted the sickle cell disease and they applied WEKA data mining tool . In their experimental results, they obtained the numerical computations.

David et al. (2013) used KNN(K Nearest Neighbor), J48, Bayesian network, random tree algorithm results for the prediction of Leukemia disease. In their results, accuracy is maintained by partitioned all these data mining classifiers.

Durairaj and Ranjani (2013) predicted the cancer and heart disease, AIDS, Diabetes, Kidney failure , IVF and so on. In their work, classifiers like J48, Naive BAYES, KNN are used. In their experimental results, 97.77% accuracy is obtained in cancer disease whereas IVF attained 70% accurate results.

Lakshmi K.R. et al.(2013) used Tanagra data mining tool with approaches like as AA, Logistics reasoning, decision tree, supervised machine learning algorithms for the dialysis of kidneys. In order to classify the data 10 fold cross validations are used. In their results, they observed that ANN performed better than decision tree and Logical regression algorithm .

Vijayarani S.et al.(2013) created a cloud network for the prediction of heart disease. In their experimental results, ANN(Artificial Neural Network) is applied.

International Journal of Engineering Researches and Management Studies

Giovanni Caocci et al.(2012)applied ANN(Artificial Neural Network) to predict the kidney rejection. In their work, they compared the sensitivity and specificity of logistic regression. For the prediction of kidney transplantation, ANN(Artificial Neural Network) and Logistics regression is discriminated.

Xun.L.et al.(2010) used ANN(Artificial Neural Network) and Naive Bayes to predict the kidney disease where Naive Bayes acquired less accuracy as of ANN.

Table 1. Data mining techniques for predicting multiple diseases

Name of the author & Year	Techniques/Methodology	Disease
A.Sai Sabitha et al. (2016)	Naive Bayes and ANN	Kidney
M.Archana Bakare et al.(2016)	Multirank	Asthma
Prasan Kumar Sahoo et al. (2016)	Map Reduce(Big Data)	Multiple
Lambodar Jena et al.(2015)	WEKA data mining tool	Kidney
Anu Chadhary et al.(2014)	Apriori and K-means clustering	Heart and Kidney
Dr.Vijayarani et al.(2014)	ANN and SVM	Kidney
Tommaso Di Noia et al.(2014)	ANN	Kidney
Solanki A.V. et al.(2014)	WEKA data mining tool	Sickle cell
David et al.(2013)	KNN, J48, Bayesian network and Random tree	Leukemia
Durairaj and Ranjani(2013)	J48, KNN, Naive Bayes, and C4.5	Multiple
Lakshmi K.R. et al.(2013)	ANN, Decision tree and Logistic regression	Kidney
Vijayarani .S. et al.(2013)	ANN	Heart
Giovanni Caocci et al.(2012)	ANN AND Logistic regression	Kidney
Xun.L.et al.(2010)	Naive Bayes and ANN	Kidney

IJERMS
International Journal of Engineering Researches and Management Studies

3. DESIGN & IMPLEMENTATION

The proposed design of the disease prediction model utilizes the multiple classification algorithms in the unique combination with automatic variable selection based upon variance and significance. The proposed model is designed to handle the kidney disease data on the basis of combination of SVM and KNN classifiers. The data of 400 patients has been used to evaluate the performance of the proposed model. The dataset contains the several variables, which includes age, blood pressure, blood cell count, WBC, RBC, etc. The decision is taken on the basis of few variables with higher significance, which are selected in the first layer of classification model using k-nearest neighbor classifier. In this paper, the application of the multi-layered classification can be described from the following figure

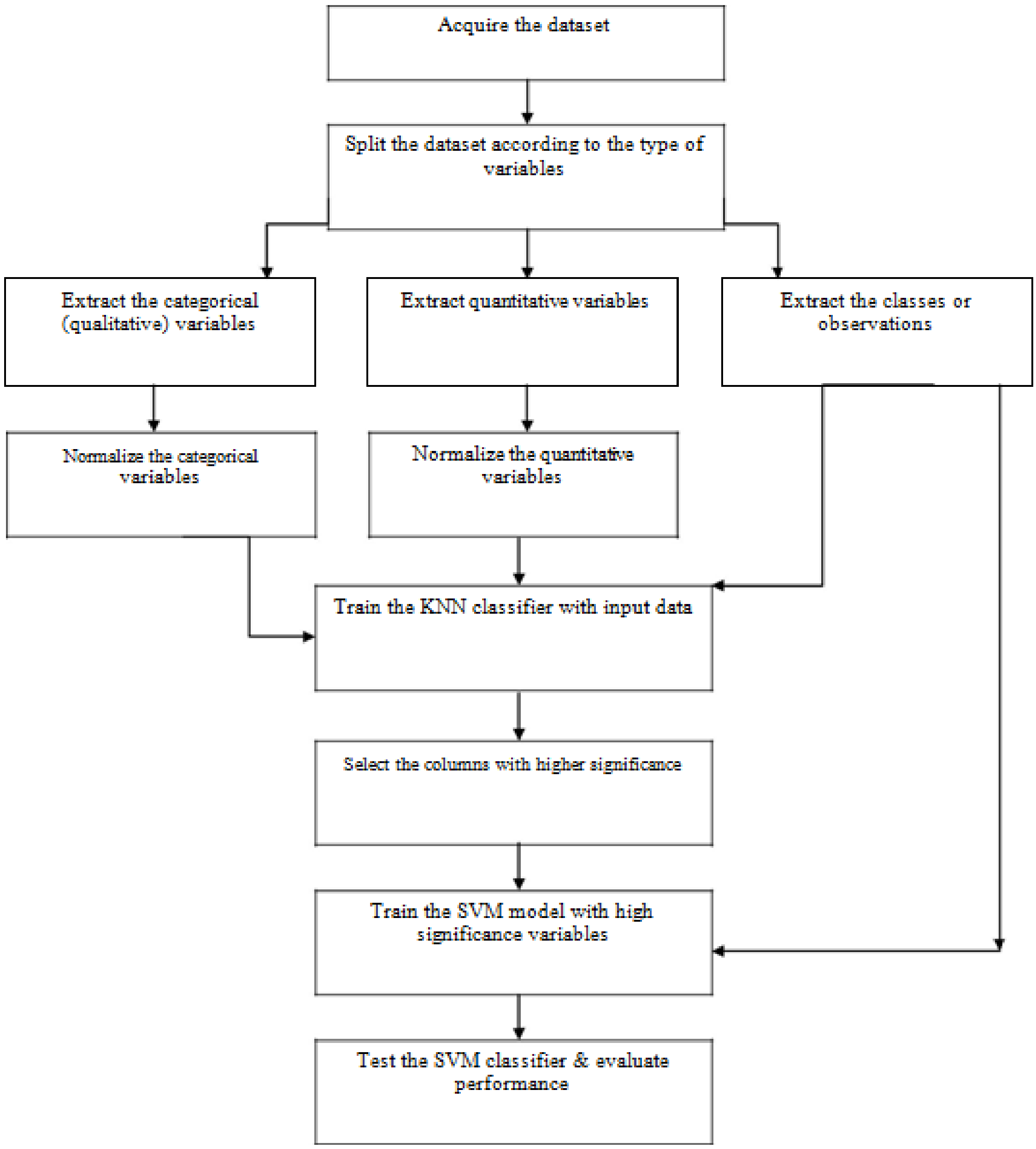

Fig 1. Flow Chart of Design & Implementation

IJERMS
International Journal of Engineering Researches and Management Studies

Algorithm 1: Dual Layered Classification for Disease Prediction

1) Acquire the dataset from the local disk
2) Convert the dataset acquisition to the data matrix
3) Split the different types of columns (variables) from the dataset to handle them individually
4) First, extract the categorical (or qualitative) variables
5) Run the iteration for each of the categorical variables

a. Extract the variable data from the matrix

b. Find the missing values in the extracted variable vector

c. Fill the missing values with another category, such as “other”

d. Convert the string variable options to numerical type

e. Transform the numerical data labels to dummy variables

f. Delete the last column of dummy variable to eliminate the chances of dummy variable trap, which can be also described as over fitting trap

6) Combine all of the dummy variable matrices to create the categorical feature matrix
7) Then, extract the quantitative variables
8) Run the iteration for each of the quantitative variables

g. Extract the variable data from the matrix

h. Find the missing values in the extracted variable vector

i. Fill the missing value with the column mean value

j. Scale the variable with the MinMax scaling method, which scales the variables to the range of 0 to 1

9) Combine all of the quantitative variable matrices to create the quantitative feature matrix
10) Combine quantitative and qualitative feature matrices to prepare the final feature vector
11) Initialize and configure the KNN classifier
12) Train the KNN with each individual variable
13) Store the accuracy for each of the individual variable
14) Sort the accuracy vector in descending order
15) Extract the N number of columns from the feature matrix according to the significance threshold
16) Split the training and testing data on the given ratio
17) Train the SVM classifier with training data
18) Test the SVM classifier with testing data
19) Compute the performance parameters from the SVM classification results
20) Return the performance parameters

4. RESULTS

The obtained experimental results are shown as two data mining algorithms which are KNN (K-Nearest Neighbor) and SVM (Support Vector Machine) . Feature evaluation or extraction and classification is done on the basis of the dataset that is being used. All the outcomes are discussed in the graphical sequence. So, in our proposed work, we calculate the performance measure of data mining algorithms used in this work based on the following statistical parameters :

IJERMS
International Journal of Engineering Researches and Management Studies

i. Accuracy%
ii. Error %
iii. Precision
iv. Recall
v. F1 error
vi. Elapsed time

All of the above mentioned parameters are compared in the tabular form for both the data mining classifiers, KNN and SVM are generated by using MATLAB itself.

1. Accuracy % :
It is observed that SVM data mining classifier gives 99.29% accuracy which is more as compared to KNN and KNN & SVM where KNN acquires 97.83% accuracy and KNN & SVM together acquires 98.93 accuracy .

Therefore , after SVM , KNN & SVM together is at second position which has less accuracy and SVM is at last position. . Many of the researchers concluded that SVM usually possess less accuracy , but in our work , SVM performed better than others in the given chronic kidney dataset. The following given table shows the comparison between KNN , SVM and KNN & SVM.

Table 2. Comparison of Accuracy % (KNN,SVM and KNN/SVM)

Id	KNN	SVM	KNN/SVM
1	100	100	98.387
2	97.058	97.014	98.529
3	100	100	100
4	100	100	100
5	96.774	100	100
6	100	100	100
7	95.238	100	100
8	98.214	100	100
9	100	100	100
10	98.275	100	100
11	98.113	100	98.437
12	96.875	100	98.333
13	95.161	96.610	100
14	95.161	100	95.238
15	93.650	100	98.245
16	100	98.484	96.721

17	100	100	98.412
18	95.238	96.875	100
19	100	98.529	98.039
20	96.969	98.387	98.360
Total	**97.83%**	**99.29%**	**98.93%**

Therefore, comparison is notified as follows :
SVM > KNN/SVM > KNN

2. Error % :
It is observed that KNN possess more error rate 2.1% as compared to SVM and KNN &SVM where SVM has 0.5% error rate and KNN & SVM contains 0.75% error.

Thus, SVM classifier has less error rate . The following table shows the comparison between error rate as follows:

Table 3. Comparison of Error % (KNN,SVM and KNN/SVM)

Id	KNN	SVM	KNN/SVM
1	0	0	2
2	2	2	1
3	1	0	0
4	0	0	0
5	4	0	0
6	3	0	0
7	5	0	0
8	2	0	0
9	1	1	0
10	2	0	0
11	2	0	1
12	2	0	1
13	3	2	0
14	3	0	4
15	5	0	1

IJERMS
International Journal of Engineering Researches and Management Studies

16	0	1	2
17	1	0	1
18	3	2	0
19	0	1	1
20	3	1	1
Total	**2.1%**	**0.5%**	**0.75%**

Therefore, comparison is notified as follows :
SVM < KNN/SVM < KNN

3. Precision :

It is observed that KNN & SVM possess more precision as compared to KNN and SVM data mining classifiers used to predict chronic kidney dataset. KNN & SVM possess 99.25% which is greater and SVM possess 99% and KNN has 97.9% precision. The following table shows the comparison of precision between KNN data mining classifier, SVM data mining classifier and KNN & SVM data mining classifier in together

Table 4. Comparison of Precision (KNN,SVM and KNN/SVM)

Id	KNN	SVM	KNN/SVM
1	100	100	98
2	98	98	99
3	99	100	100
4	100	100	100
5	96	100	100
6	97	100	100
7	95	100	100
8	98	100	100
9	99	99	100
10	98	100	100
11	98	100	99
12	98	100	99
13	97	98	100
14	97	100	96

15	95	100	99
16	100	99	98
17	99	100	99
18	97	98	100
19	100	99	99
20	97	99	99
Total	**97.8%**	**99%**	**99.25%**

Therefore, comparison is notified as follows :
KNN/SVM > SVM > KNN

4. Recall :
It is observed that KNN & SVM with 99.83% has more value in recall and SVM possess 99.76% recall value and KNN has 98.77 % recall . But in all, these classifiers possess approximate values to each other.

The following table shows the comparison between KNN , SVM and KNN & SVM data mining classifiers. Recall is the another parameter that has implemented in our work which is not done yet by any of the researchers. This is to check how many times repeatedly we contradict the chronic kidney disease in human body by using mining approaches.

Table 5. Comparison of Recall (KNN,SVM and KNN/SVM)

Id	KNN	SVM	KNN/SVM
1	100	100	98.387
2	100	97.014	100
3	98.360	100	100
4	100	100	100
5	96.774	100	100
6	95.454	100	100
7	96.774	100	100
8	98.214	100	100
9	98.412	98.305	100
10	98.275	100	100
11	98.113	100	100

IJERMS
International Journal of Engineering Researches and Management Studies

12	100	100	100
13	100	100	100
14	100	100	98.360
15	98.333	100	100
16	100	100	100
17	98.245	100	100
18	100	100	100
19	100	100	100
20	98.461	100	100
Total	**98.77%**	**99.76%**	**99.83%**

Therefore, comparison is notified as follows :
SVM > KNN & SVM > KNN

5. F1 error :

It is observed that SVM shows more results in F1 error whereas KNN & SVM in together shows less F1 error and KNN possess the least value as compared to both. SVM has 99.59% F1 error and KNN&SVM has 99.38 % F1 error and value of KNN is 98.23% .

Basically this statistical parameter F1 error determines that if there is chronic kidney disease in human body or elsewhere if not so by implementing the defined or undefined functions in itself using mining algorithms which in turn will provide the results in the form of F1 error .

The following table shows the comparison between F1 error in KNN , SVM and KNN&SVM data mining classifier

Table 6. Comparison of F1 error (KNN,SVM and KNN/SVM)

Id	KNN	SVM	KNN/SVM
1	100	100	98.371
2	98.507	98.484	99.259
3	100	100	100
4	96.774	100	100
5	97.674	100	100
6	96.000	100	100

7	98.214	100	100
8	98.214	100	100
9	99.200	99.145	100
10	98.275	100	100
11	98.113	100	100
12	98.412	100	99.159
13	97.520	98.275	96.774
14	97.520	100	99.115
15	95.935	100	98.333
16	100	99.236	99.200
17	99.115	100	100
18	97.561	98.412	99.009
19	100	99.259	99.173
20	97.709	99.187	99.290
Total	**98.23%**	**99.59%**	**99.38%**

Therefore, comparison is notified as follows :
SVM > KNN & SVM > KNN

6. Elapsed Time :

Elapsed time can be defined as the time to be noted from the beginning of some work till its end. This can be calculated on the daily basis in everyday human life. Elapsed time is considered as one of the significant skill in our lives.

It is observed that KNN & SVM takes more time to predict the chronic kidney dataset having value 0.9817% whereas SVM has 0.8045% elapsed time and time taken by KNN is 0.12558% .

IJERMS
International Journal of Engineering Researches and Management Studies

Table 7. Comparison of Elapsed time (KNN,SVM and KNN/SVM)

Id	KNN	SVM	KNN/SVM
1	0.391	0.782	1.012
2	0.121	0.786	1.004
3	0.120	0.814	0.867
4	0.121	0.781	0.929
5	0.125	0.821	1.088
6	0.130	0.783	1.047
7	0.124	0.946	0.918
8	0.122	0.799	0.959
9	0.123	0.767	0.941
10	0.125	0.781	0.992
11	0.123	0.785	0.908
12	0.124	0.783	1.066
13	0.123	0.773	1.102
14	0.122	0.800	1.159
15	0.122	0.815	0.927
16	0.123	0.789	0.964
17	0.126	0.783	1.040
18	0.122	0.774	0.858
19	0.123	0.783	1.020
20	0.123	0.777	0.945
Total	**0.13 sec**	**0.79 sec**	**0.98 sec**

Therefore, comparison is notified as follows :
KNN < SVM < KNN & SVM

ISSN: 2394-7659
IMPACT FACTOR- 3.775
IJERMS
International Journal of Engineering Researches and Management Studies

References

1. *A.Sai Sabitha, Abhay Bansal, Khushboo Chandel, Veenita Kunwar, Chronic Kidney Disease Analysis Using Data Mining Classification Techniques, 6th International Conference on Cloud System and Big Data Engineering, 2016.*
2. *Agarwal.Y, Pandey, H.M, Performance evaluation of different techniques in the context of data mining- A case of an eye disease. In Confluence The Next Generation Information Technology Summit(Confluence), 5th International Conference-IEEE, 2014.*
3. *Ahmed. S, Tanzir Kabir, M.Tanzeer Mahmood, N. Rahman, R.M., Diagnosis of kidney disease using fuzzy expert system. In Software, Knowledge Information Management and Applications(SKIMA), 8th International Conference- IEEE, 2014.*
4. *Alfisahrin, S.D.N.N, Mantoro, T, Data Mining Techniques for Optimization of Liver Disease Classification. In Advanced Computer Science Applications and Technologies(ACSAT), International Conference- IEEE, 2013.*
5. *Amin, S.U, Agarwal, K.Beg, R, Genetic Neural Network based data mining in prediction of heath disease using risk factors. In Information and Communication Technologies(ICT), IEEE Conference, 2013.*
6. *Andrew Kusiak, Bradley Dixonb, Shital Shaha, Predicting survival time for kidney dialysis patients: a data mining approach, Elsevier Publication, Computers in Biology and Medicine 35, page no 311-327, 2005.*
7. *Anu Chaudhary, Puneet Garg, Detecting and Diagnose Heart and Kidney disease by Patient Monitoring System, International Journal of Mechanical Engineering and Information Technology, Vol.2 Issue 6//June//Page No: 493-499, 2014.*
8. *David S.K., Saeb A.T., Al Rubeaan K., Comparative Analysis of Data Mining Tools and Classification Techniques using WEKA in medical Bioinformatics, Computer Engineering and Intelligent Systems, 4(13):28-38, 2013.*
9. *Dhamodharan S, Liver Disease Prediction Using Bayesian Classification, Special Issues, 4th National Conference on Advance Computing , Application Technologies, May 2014.*
10. *Vijayarani, S., & Dhayananad, M., S. Kidney Disease Prediction using SVM and ANN algorithms, unpublished.*
11. *Duraiaj M, Ranjani V, Data Mining applications in healthcare sector a study. Int . J. Sci. Technol. Res. IJSTR, 2(10), 2013.*
12. *Giovanni Caocci, Roberto Baccoli, Roberto Littera, Sandro Orru, Carlo Carcassi and Giorgio La Nasa, Comparison Between an ANN and Logistic Regression in Predicting Long Term Kidney Transplantation Outcome, Chapter 5, an open access article distributed under the terms of the Creative Commons Attribution License, http://dx.doi.org/10.5772/53104, 2012.*
13. *Su, J.L., Wu, G.Z., & Chao, I.P. The approach of data mining methods for medical database. In Engineering in Medicine and Biology Society, Proceedings of the 23rd Annual International Conference of the IEEE(Vol.4, pp.3824-3826), IEEE, 2001.*
14. *Joshi J, Rinal D, Patel J, Diagnosis and Prognosis of Breast Cancer using Classification Rules, International Journal of Engineering Research and General Science, 2(6): 315-323, October 2014.*
15. *Kumar M.N., Alternating Decision trees for early diagnosis of dengue fever .arXiv preprint arXiv: 1305.7331, 2013.*
16. *Lakshmi, K.R., Nagesh, Y., & VeeraKrishna, M, Performance Comparison of Three Data Mining Techniques for Predicting Kidney Dialysis Survivability, International Journal of Advances in Engineering & Technology, Mar., Vol.7, Issue 1, Page No: 242-254, 2013.*
17. *Lambodar Jena, Narendra Ku. Kamila, Distributed Data Mining Classification Algorithms for Prediction of Chronic Kidney Disease, International Journal of Emerging Research in Management & Technology ISSN: 2278-9359(Vol 4, Issue-11), IJERMT, 2015*
18. *Lee, H.G., Noh, K. Y., & Ryu, K.H, A data mining approach for coronary heart disease prediction using HRV features and carotid arterial wall thickness. In Biomedical Engineering and Informatics, BMEI, International Conference(Vol. 1, pp.200-206), IEEE- 2008.*
19. *M Archana Bakare, Prof. R.V. Argiddi, Prediction of Diseases using Big Data Analysis, International Journal of Innovative Research in Computer and Communication Engineering(Vol.4, Issue 4), ISSN, April 2016.*

ISSN: 2394-7659
IMPACT FACTOR- 3.775

IJERMS

International Journal of Engineering Researches and Management Studies

20. *M.A. Nishara Banu, Gomathy B., Heart Disease Prediction using Data mining Techniques . In Intelligent Computing Applications(ICICA), International Conference on (pp.130-133), IEEE, March 2014.*
21. *M.Cottle,W.Hoover, S. Kanwal, M. Kohn, T. Strome and N. W. Treister, Transforming Health Care Through Big Data, 2013.*
22. *Maskery, S., Zhang, Y., Hu, H., Shriver, C., Hooke, J., & Liebman, M. (2006, June), Caffeine intake, race, and risk of invasive breast cancer lessons learned from data mining a clinical database. In Computer- Based Medical Systems, CBMS , 19th IEEE International Symposium on (pp. 714-718), IEEE, 2006.*
23. *Prasan Kumar Sahoo, Suvendu Kumar Mohapatra, Shih-Lin Wu, Analyzing Healthcare Big Data with Prediction for Future Health Condition, Journal Publications , IEEE, 2016.*
24. *Shakil K.A. and Alam M., Data Management in Cloud Based Environment using K-median Clustering Technique, IJCA Proceedings on 4th International IT Summit Confluence - The Next Generation Information Technology Summit Confluence , 2013.*
25. *Vijayarani, S., Sudha, S., Comparative Analysis of Classification Function Techniques for Heart Disease Prediction, International Journal of Innovative Research in Computer and Communication Engineering, 1(3): 735-741, 2013.*
26. *Watanasusin, N., & Sanguansintukul, S., Classifying chief complaint in ear diseases using data mining techniques. In Digital Content, Multimedia Technology and its Applications (IDCTA), 7th International Conference on (pp. 149-153), IEEE, August 2011.*
27. *Xiong, X., Kim, Y., Baek, Y., Rhee, D. W., & Kim, S. H., Analysis of breast cancer using data mining & statistical techniques. In Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, First ACIS International Workshop on Self-Assembling Wireless Networks. 6th International Conference on (pp. 82-87), IEEE, May 2005.*
28. *Xun, L., Xiaoping, W., Ningshan, L., & Tanqi, L., Application of radial basis function neural network to estimate glomerular filtrationrate in Chinese patients with chronic kidney disease. In Computer Application and System Modeling (ICCASM), International Conference on (Vol. 15, pp. V15-332), IEEE, October 2010*