CHRONIC KIDNEY DISEASE ANALYSIS USING DATA MINING CLASSIFICATION TECHNIQUES

Veenita Kunwar [1]
ASET,CSE
Amity University Uttar Pradesh
Noida, India
veenitakunwar@gmail.com

Khushboo Chandel[2]
ASET, CSE
Amity University Uttar Pradesh
Noida, India
geetchandel92@gmail.com

A. Sai Sabitha[3]
ASET, CSE
Amity University Uttar Pradesh
Noida, India
saisabitha@gmail.com

Abhay Bansal[4]
ASET,CSE
Amity University Uttar Pradesh
Noida, India
abansal1@amity.edu

***Abstract*— Data mining has been a current trend for attaining diagnostic results. Huge amount of unmined data is collected by the healthcare industry in order to discover hidden information for effective diagnosis and decision making. Data mining is the process of extracting hidden information from massive dataset, categorizing valid and unique patterns in data . There are many data mining techniques like clustering, classification, association analysis, regression etc. The objective of our paper is to predict Chronic Kidney Disease(CKD) using classification techniques like Naive Bayes and Artificial Neural Network(ANN). The experimental results implemented in Rapidminer tool show that Naive Bayes produce more accurate results than Artificial Neural Network.**

Keywords— Data mining, Classification, Chronic Kidney disease, Naive Bayes, Artificial Neural Network.

I. INTRODUCTION

Data Mining is one of the most encouraging areas of research with the purpose of finding useful information from voluminous data sets. It has been used in many domains like image mining, opinion mining, web mining, text mining, graph mining etc. Its applications include anomaly detection, financial data analysis, medical data analysis, social network analysis, market analysis etc. It has become popular in health organization as there is a requirement of analytical methodology for predicting and finding unknown patterns and information in health data. It plays a vital role for discovering new trends in healthcare industry.

Data Mining is particularly useful in medical field when no availability of evidence favoring a particular treatment option is found. Large amount of complex data is being generated by healthcare industry about patients, diseases, hospitals, medical equipments, claims, treatment cost etc. that requires processing and analysis for knowledge extraction. Data mining comes up with a set of tools and techniques which when applied to this processed data, provides knowledge to healthcare professionals for making appropriate decisions and enhancing the performance of patient management tasks. Patients with similar health issues can be grouped together and effective treatment plans could be suggested based on patient's history, physical examination, diagnosis and previous treatment patterns.

Chronic kidney disease (CKD) has become a global health issue and is an area of concern. It is a condition where kidneys become damaged and cannot filter toxic wastes in the body. Our work predominantly focuses on detecting life threatening diseases like Chronic Kidney Disease (CKD) using Classification algorithms like Naive Bayes and Artificial Neural Network(ANN).

The remaining paper is organized as follows: Section II reviews some work related to medical field. Section III describes the research methodology. Section IV includes experimental setup. In Section V, the results are discussed and analysed. Section VI discusses case study. Finally, Section VII concludes the paper discussing future scope.

II. LITERATURE SURVEY

Nowadays, health care industries are providing several benefits like fraud detection in health insurance, availability of medical facilities to patients at inexpensive prices, identification of smarter treatment methodologies, construction of effective healthcare policies, effective hospital resource management, better customer relation, improved patient care and hospital infection control. Disease detection is also one of the significant areas of research in medical.

Data mining approaches have become essential for healthcare industry in making decisions based on the analysis

978-1-4673-8203-8/16/$31.00 ©2016 IEEE

of the massive clinical data. Data mining is the process of extracting hidden information from massive dataset. Techniques like classification, clustering, regression and association have been used by in medical field to detect and predict disease progression and to make decision regarding patient's treatment. Classification is a supervised learning approach that assign objects in a collection to target classes. It is the process which classifies the objects or data into groups, the members of which have one or more characteristic in common. The techniques of classification are SVM, decision tree, Naive Bayes, ANN etc. Clustering involves grouping of objects of similar kinds together in a group or cluster. Some of its techniques include K-means, K-medoids, agglomerative, divisive, DBSCAN etc. Association states the probability of occurrence of items in a set. Apriori is an example of association [36], [37], [38], [39], [40].

S.No	Author	Year	Disease	Technique
1	Ju-Hsin Tsai [1]	2008	Cancer breast	CLUSTERING (AGNES)
2	Mostafa Ghannad Rezaie et al[2]	2008	Temporal	SVM(classification)
3	Jenn-Lung Su et al. [3]	2001	breast tumour	Bayesian Network , DT,
4	Paolo Bonato et al. [4]	2004	Parkinson	clustering
5	S Wang et al,[5]	2005	breast cancer	decision tree
6	Yanwei Xing et al. [6]	2007	coronary heart	SVM ,ANN ,DT
7	Sellappan Palaniappan et al. [7]	2008	heart disease	(DT, naive bayes,ANN)
8	Heon Gyu Lee et al. [8]	2008	coronary heart	classification (SVM)
9	K.Srinivas et al. [9]	2010	heart disease	DT, Naive Bayes, ANN
10	Narin Watanasusin [10]	2011	ear	ANN,Naive Bayes
11	Debabrata Pal et al. [11]	2011	heart disease	Classification (DT)
12	T.John Peter [12]	2012	heart disease	DT, NB, K-NN and NN
13	Jenn-Long Liu [13]	2012	cardiac	GA, K-Means algorithm
14	Geeta Yadav [14]	2012	Parkinson	DT,Regression, SVM
15	M. Ilayaraja [15]	2013	multiple	Apriori algorithm
16	Sivagowry .S et al. [16]	2013	heart disease	Classification (DT, ANN)
17	K. Vasantha Kokilam[17]	2012	genetic	clustering &
18	Syed Umar Amin et al. [18]	2013	heart disease	genetic neural network
19	Girija D.K [19]	2013	fibroid	ANN
20	Juliet Rani Rajan [20]	2013	lung cancer	ANN
21	Sa'diyah Noor Novita Alfisahrin	2013	liver	DT,Naive Bayes
22	Ranganatha S. et al. [22]	2013	heart disease	ID3,Naive Bayes
23	Yukti Agarwal [23]	2014	eye disease	Fuzzy logic , ANN
24	M.A.Nishara Banu [24]	2014	heart disease	k-means,c4.5
25	X.Xiong et. al [25]	2005	Breast Cancer	DT,association rules
26	Susan Maskery et al. [26]	2006	Breast Cancer	Bayesian network
27	Menolascina F et al. [27]	2007	Breast Cancer	J48 and Naïve Bayes
28	Qi Fan,Chang-jie Zhu [28]	2010	Breast Cancer	Pre-classification method
29	Abdelaal [29]	2010	Breast cancer	Classification (SVM),DT
30	Vijayarani, S. Et al.[30]	2015	Kidney	SVM and
31	Chiu, R. K et al. [31]	2012	Kidney	ANN
32	Lakshmi, K. R et al. [32]	2014	Kidney	(classification)DT,ANN,
33	Xun, L et. Al [33]	2010	Kidney	ANN
34	Ravindra, B. V. et al. [34]	2014	Kidney	K-means clustering
35	Ahmed, S et al. [35]	2014	Kidney	Fuzzy Logic

Fig. 1.Data Mining Techniques used for Disease detection

Figure 1 describes about various data mining techniques used over last 15 years for investigating various diseases.

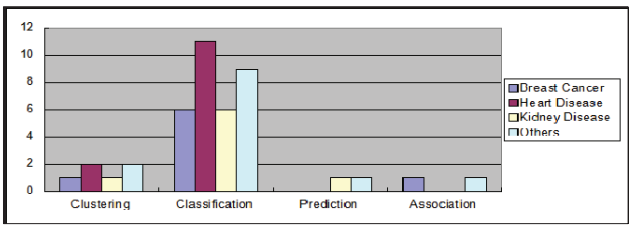

Fig. 2. Data Mining techniques used for disease detection

Figure 2 shows a potential use of data mining techniques like clustering, classification which includes DT, Naive Bayes, Neural Network , SVM etc. in predicting heart disease [6], [7], [8], [9], [11], [12], [13], [16], [18], [22], [24]. Classification, association and clustering techniques have also been adopted for breast cancer detection [1], [3], [5], [25], [26], [27], [28], [29]. Other diseases like lung cancer, liver cancer, diabetes, parkinson's disease etc. have also been studied, detected and diagnosed by data mining algorithms [2], [3], [4], [10], [14], [15], [17], [19], [20], [21], [23].

The present lifestyle of people, working environment and diet have given rise to many diseases, one of which includes chronic kidney disease. Chronic Kidney disease(CKD) is prevailing nowadays and has become a global health issue which must be timely detected and diagnosed. Kidneys are important organs of human body that eradicate toxic and unwanted waste from blood causing smooth functioning of body organs. CKD is a condition that describes loss of kidney function over time making it difficult for them to filter poisonous wastes from the body. Researchers in their recent study have addressed the use of data mining techniques for CKD detection [30], [31], [32], [33], [34], [35].

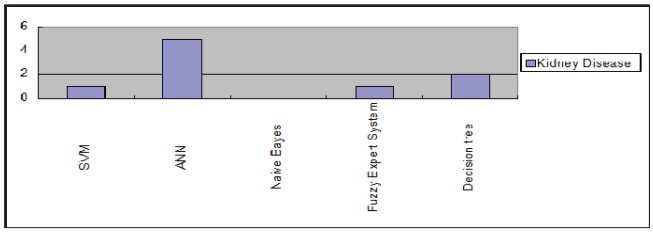

Fig. 3. Classification techniques used for detecting kidney disease

It has been observed that classification algorithms have widely been used for identifying and investigating kidney disease. Figure 3 shows that many research work has been conducted using ANN while other techniques like SVM, Fuzzy logic has been used the least. It has also been observed that Naive Bayes has rarely been used. In this research work Naive Bayes approach, an important classification algorithm which uses Bayes Theorem has been used. It is particularly suited when the dimensionality of inputs is high. In this work the dimensionality of dataset is 25. The performance of Naive Bayes has also been compared with ANN algorithm.

Naive Bayes is a probabilistic classifier based on Bayes theorem. It assumes variables are independent of each other. The algorithm is easy to build and works well with huge data sets. It has been used because it makes use of small training data to estimate the parameters important for classification. Bayes Theorem states the following:

$P(A|X) = P(X|A) \cdot P(A) / P(X)$.

$P(X)$ is constant for all classes.

$P(A)$ = relative frequency of class A samples a such that p is increased=c Such that $P(X|A)\ P(A)$ is increased

Problem: computing $P(X|A)$

ARTIFICIAL NEURAL NETWORK

The artificial neural network (ANN) is a computational model inspired by structure and function of biological neural network. It is an interconnection of artificial neurons that processes information using connected links. It has been used as it works well with noisy data and processes both numeric and categorical data. It is used for supervised learning and unsupervised clustering. Some of its key strengths include high compute performance when processing huge data, robustness and adaptability to varying inputs and outputs. All these encourages its use in clinical decision making.

This research work mainly focuses on chronic kidney disease detection using classification algorithms like Naive Bayes and ANN.

III. METHODOLOGY

Data Mining is one of the most significant stages of the Knowledge Data Discovery process. The process involves data collection from various sources with preprocessing of the chosen data. The data is then transformed into suitable format for further processing. Data Mining technique is applied on the data to extract valuable information and evaluation is done at the end.

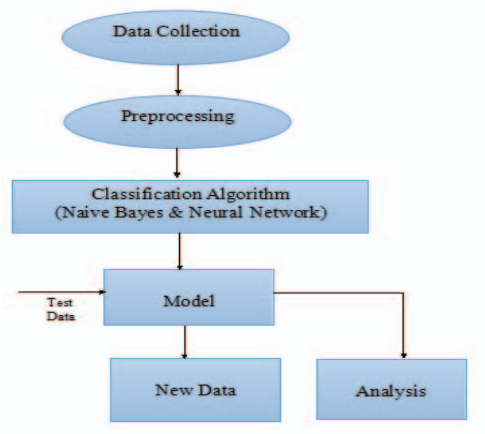

Fig. 4. Flowchart showing KDD

IV. EXPERIMENTAL SETUP

A. *Data Set*

The clinical data of 400 records considered for analysis has been taken from UCI Machine Learning Repository. The data obtained after cleaning and removing missing values is 220. The data has been implemented using Rapid Miner tool. There are 25 attributes in the dataset. The numerical attributes include age, blood pressure, blood glucose random, blood urea, serum creatinine, sodium, potassium, hemoglobin, packaged cell volume, WBC count, RBC count. The nominal attributes include specific gravity, albumin, sugar, RBC, pus cell, pus cell clumps, bacteria, hypertension, diabetes mellitus, coronary artery disease, appetite, pedal edema, anemia and class.

Number of Instances: 400
Number of Attributes: 25 Class: {CKD, NOTCKD}
Missing Attribute Values: yes
Class Distribution: [63% for CKD] [37% for NOTCKD]

B. *Model Construction*

This work has been performed in Rapidminer data mining tool. Following are Naive Bayes models :

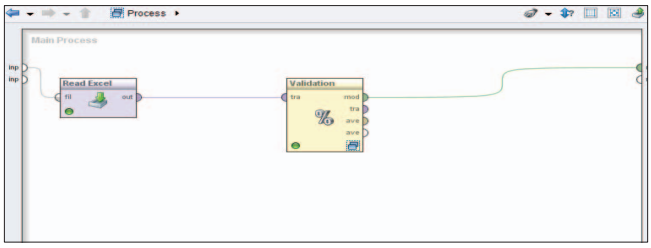

Fig. 5. Validation in Naive Bayes

Figure 5 shows the validation process which helps to examine the accuracy of fitted models and its performance on new data.

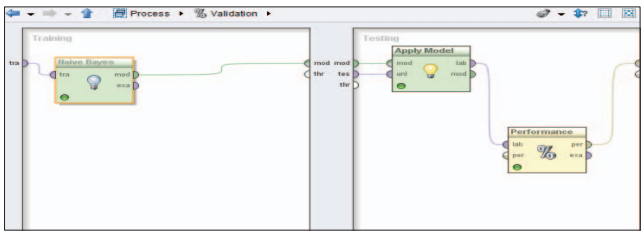

Fig. 6. Training and Testing Process in Naive Bayes

Figure 6 shows training data which is used to build a model and testing dataset is used to measures its performance.

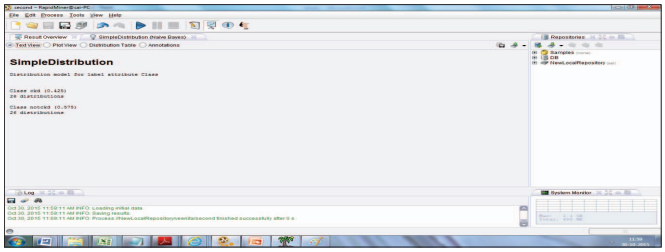

Fig. 7. Text View in Naive Bayes

Figure 7 shows distribution model for label attribute class. 0.425 have CKD while 0.575 do not have CKD.

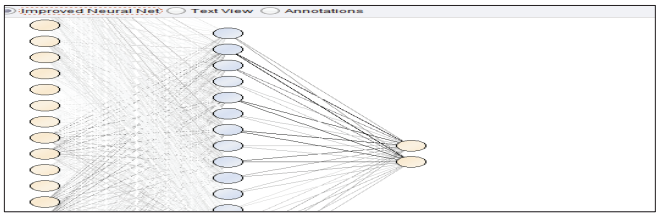

Fig. 8. ANN Model

Figure 8 shows model of Artificial Neural Network.

V. RESULTS AND ANALYSIS

The experimental comparison of Naive Bayes and ANN are done based on the performance vectors. It is statistical performance evaluation of classification tasks and contains list of performance criteria values.

A. *Performance Analysis(Naive Bayes vs ANN)*

PerformanceVector

```
PerformanceVector:
accuracy: 100.00%
classification_error: 0.00%
kappa: 1.000
ConfusionMatrix:
weighted_mean_recall: 100.00%, weights: 1, 1
spearman_rho: 1.000
kendall_tau: 1.000
absolute_error: 0.000 +/- 0.000
relative_error: 0.00% +/- 0.00%
relative_error_lenient: 0.00% +/- 0.00%
relative_error_strict: 0.00% +/- 0.00%
normalized_absolute_error: 0.000
root_mean_squared_error: 0.000 +/- 0.000
root_relative_squared_error: 0.000
squared_error: 0.000 +/- 0.000
```

Fig. 9. Performance Vector for Naive Bayes

Figure 9 shows performance vector containing list of performance criteria values. Accuracy refers to number of correct predictions or how precise the dataset is being classified. Kappa takes into account the correct predictions occurring by chance. It gives a quantitative measure of the magnitude of agreement between observers. It lies in the range -1 to 1, where 1 is perfect agreement, 0 is chance agreement, and negative values indicate agreement less than chance i.e disagreement between observers. The accuracy of Naive Bayes obtained is 100% and kappa value is 1 which indicates perfect agreement.

PerformanceVector

```
PerformanceVector:
accuracy: 72.73%
kappa: 0.455
spearman_rho: 0.542
kendall_tau: 0.542
absolute_error: 0.246 +/- 0.388
relative_error: 24.63% +/- 38.75%
relative_error_lenient: 24.63% +/- 38.75%
relative_error_strict: 646.39% +/- 1,799.03%
normalized_absolute_error: 0.493
root_mean_squared_error: 0.459 +/- 0.000
root_relative_squared_error: 0.918
squared_error: 0.211 +/- 0.363
```

Fig. 10 Performance Vector for Artificial Neural Network

Figure 10 shows performance of ANN with accuracy obtained as 72.73% and kappa value as 0.455 showing moderate agreement range.

B. *Accuracy of Naive Bayes vs ANN*

Multiclass Classification Performance | Annotations

Table View | Plot View

accuracy: 100.00%

	true ckd	true notckd	class precision
pred. ckd	11	0	100.00%
pred. notckd	0	11	100.00%
class recall	100.00%	100.00%	

Fig. 11. Accuracy of Naive Bayes

Figure 11 shows 100% accuracy of Naive Bayes algorithm. This shows it produces most accurate and correct results.

Multiclass Classification Performance | Annotations

Table View | Plot View

accuracy: 72.73%

	true ckd	true notckd	class precision
pred. ckd	11	6	64.71%
pred. notckd	0	5	100.00%
class recall	100.00%	45.45%	

Fig. 12 Accuracy of Neural Network

Figure 12 shows 72.73% accuracy obtained for ANN which shows that results obtained are not perfectly correct.

VI. CASE STUDY

Nowadays the working conditions, eating habits, pollution, environmental factors have caused stress and anxiety leading to diseases like diabetes, affecting young and old. Thus, following factors have been considered for case study :

1. Diabetes Mellitus
2. Age

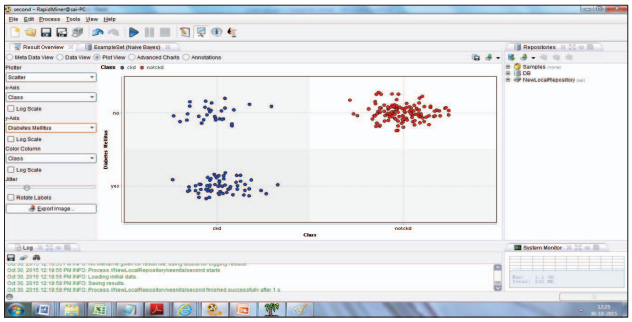

Fig. 13. Plot View showing Class with Diabetes for Naive Bayes

In Figure 13 the lower left corner has blue scatter plot which indicates diabetes with chronic kidney disease(CKD) . The upper left corner having blue scatter plot indicates CKD but no diabetes. The upper right corner with red scatter plot indicates no diabetes and no CKD. It has been observed from this figure that diabetes can cause kidney disease so it can be considered as one of the factors causing CKD but it is not the only contributing factor for CKD.

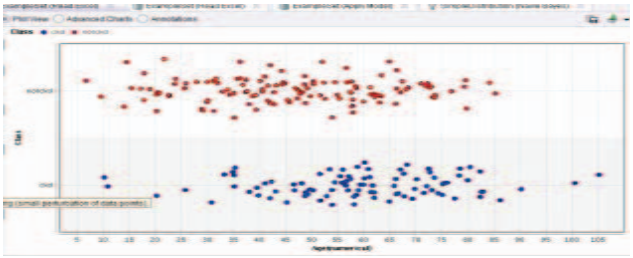

Fig. 14 Plot View for Naive Bayes showing age with class

Figure 14 shows Class with respect to age. It has been observed from the figure that most people have CKD with age between 38 to 82. There are also cases where people having age between 30 to 75 do not have CKD.

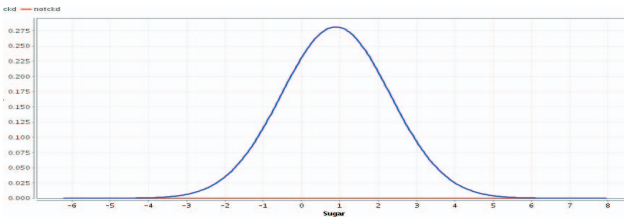

Fig. 15 Plot View for Naive Bayes for sugar as a parameter

Figure 15 shows Class with respect to sugar. CKD has been detected with increased sugar shown by blue curve. No CKD has been depicted by red line.

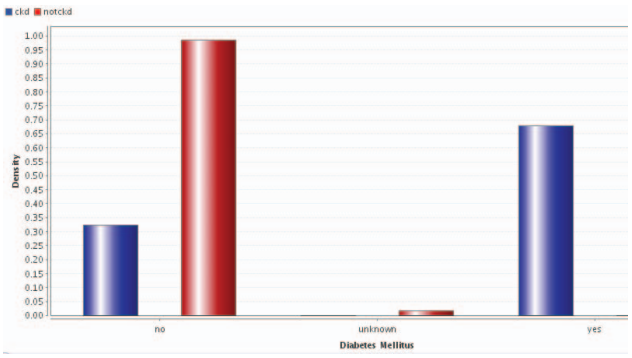

Fig. 16 Distribution Model of Naive Bayes taking Diabetes as a parameter

Figure 16 depicts CKD for range 0.00 to 0.30 with no diabetes. It has been found that diabetes and CKD do not occur for range 0.00 to 0.97 . No CKD is found for range 0.00 to 0.02 when diabetes parameter is unknown. It has also been observed that diabetes occur along with CKD for range 0.00 to 0.68. Thus, diabetes can be considered as one of the important factors for CKD but not the only factor.

From the experiment performed, the algorithm with higher accuracy has been considered as a good algorithm. Each classifier shows different accuracy rate. Naive Bayes has the 100% accuracy which indicates that it produces more accurate results than ANN, hence, it is considered as a good classification algorithm.

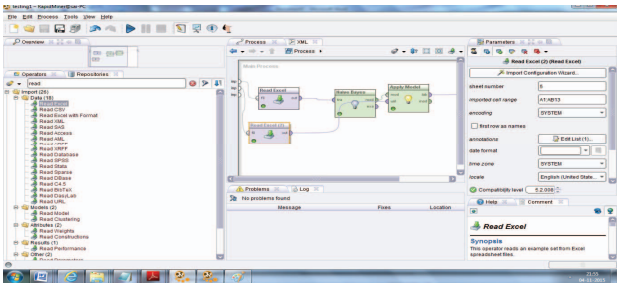

Fig. 17 Model for Naive Bayes

Figure 17 shows a new test dataset taken in Naive Bayes classifier.

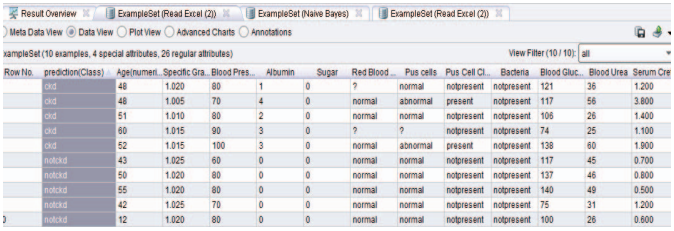

Fig. 18 Output for test data in Naive Bayes

As one of the primary component in data mining algorithm is model evaluation we have evaluated the model using a dataset of unknown class labels. Figure 18 shows the output obtained after taking new test data in Naive Bayes classifier and the results were found to be accurate.

VII. CONCLUSION

Chronic Kidney Disease has been predicted and diagnosed using data mining classifiers: ANN and Naive Bayes. Performances of these algorithms are compared using Rapidminer tool. The obtained results showed that Naive Bayes is the most accurate classifier with 100% accuracy when compared to ANN having 72.73% accuracy. In this research study, some of the factors considered were age, diabetes, blood pressure, RBC count etc. The work can be extended by considering other parameters like food type, working environment, living conditions, availability of clean water, environmental factors etc for kidney disease detection. Further studies can be conducted using other classifiers like Fuzzy logic, KNN.

ACKNOWLEDGMENT

The authors would like to express their gratitude to UCI Machine Learning Repository for providing the dataset and facilitating anonymised access to it. We would thank then for making this study possible. Special thanks to Dr.P.Soundarapandian, L.Jerlin Rubini and Dr.P.Eswaran for providing source information.

REFERENCES

[1] Tsai, J. H. *(2008). Data Mining for DNA Viruses with Breast Cancer and its Limitation.* INTECH Open Access Publisher.

[2] Ghannad-Rezaie, M., & Soltanian-Zadeh, H. (2008). *Interactive knowledge discovery for temporal lobe epilepsy*. INTECH Open Access Publisher.

[3] Su, J. L., Wu, G. Z., & Chao, I. P. (2001). The approach of data mining methods for medical database. In *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*(Vol. 4, pp. 3824-3826). IEEE.

[4] Bonato, P., Sherrill, D. M., Standaert, D. G., Salles, S. S., & Akay, M. (2004, September). Data mining techniques to detect motor fluctuations in Parkinson's disease. In *Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE* (Vol. 2, pp. 4766-4769). IEEE.

[5] Wang, S., Zhou, M., & Geng, G. (2005). Application of fuzzy cluster analysis for medical image data mining. *Mechatronics and Automation*, *2*, 631-636.

[6] Xing, Y., Wang, J., Zhao, Z., & Gao, Y. (2007, November). Combination data mining methods with new medical data to predicting

outcome of coronary heart disease. In *Convergence Information Technology, 2007. International Conference on* (pp. 868-872). IEEE.

[7] Palaniappan, S., & Awang, R. (2008, March). Intelligent heart disease prediction system using data mining techniques. In *Computer Systems and Applications, 2008. AICCSA 2008. IEEE/ACS International Conference on* (pp. 108-115). IEEE.

[8] Lee, H. G., Noh, K. Y., & Ryu, K. H. (2008, May). A data mining approach for coronary heart disease prediction using HRV features and carotid arterial wall thickness. In *BioMedical Engineering and Informatics, 2008. BMEI 2008. International Conference on* (Vol. 1, pp. 200-206). IEEE.

[9] Srinivas, K., Rao, G. R., & Govardhan, A. (2010, August). Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques. In *Computer Science and Education (ICCSE), 2010 5th International Conference on* (pp. 1344-1349). IEEE.

[10] Watanasusin, N., & Sanguansintukul, S. (2011, August). Classifying chief complaint in ear diseases using data mining techniques. In *Digital Content, Multimedia Technology and its Applications (IDCTA), 2011 7th International Conference on* (pp. 149-153). IEEE.

[11] Pal, D., Chakraborty, C., & Mandana, K. M. (2011, November). Data mining approach for coronary artery disease screening. In *Image Information Processing (ICIIP), 2011 International Conference on* (pp. 1-6). IEEE.

[12] Peter, T. J., & Somasundaram, K. (2012, March). An empirical study on prediction of heart disease using classification data mining techniques. In*Advances in Engineering, Science and Management (ICAESM), 2012 International Conference on* (pp. 514-518). IEEE.

[13] Liu, J. L., Hsu, Y. T., & Hung, C. L. (2012, June). Development of evolutionary data mining algorithms and their applications to cardiac disease diagnosis. In*Evolutionary Computation (CEC), 2012 IEEE Congress on* (pp. 1-8). IEEE.

[14] Yadav, G., Kumar, Y., & Sahoo, G. (2012, November). Predication of Parkinson's disease using data mining methods: A comparative analysis of tree, statistical and support vector machine classifiers. In *Computing and Communication Systems (NCCCS), 2012 National Conference on* (pp. 1-8). IEEE.

[15] Ilayaraja, M., & Meyyappan, T. (2013, February). Mining medical data to identify frequent diseases using Apriori algorithm. In *Pattern Recognition, Informatics and Mobile Engineering (PRIME), 2013 International Conference on*(pp. 194-199). IEEE.

[16] Sivagowry, S., Durairaj, M., & Persia, A. (2013, February). An empirical study on applying data mining techniques for the analysis and prediction of heart disease. In *Information Communication and Embedded Systems (ICICES), 2013 International Conference on* (pp. 265-270). IEEE

[17] Kokilam, K. V., & Latha, D. P. M. P. (2012, December). A review on evolution of data mining techniques for protein sequence causing genetic disorder diseases. In *Computational Intelligence & Computing Research (ICCIC), 2012 IEEE International Conference on* (pp. 1-6). IEEE.

[18] Amin, S. U., Agarwal, K., & Beg, R. (2013, April). Genetic neural network based data mining in prediction of heart disease using risk factors. In*Information & Communication Technologies (ICT), 2013 IEEE Conference on*(pp. 1227-1231). IEEE.

[19] Girija, D. K., Shashidhara, M. S., & Giri, M. (2013, October). Data mining approach for prediction of fibroid disease using neural networks. In *Emerging Trends in Communication, Control, Signal Processing & Computing Applications (C2SPCA), 2013 International Conference on* (pp. 1-5). IEEE.

[20] Rajan, J. R., & Chelvan, C. C. (2013, December). A survey on mining techniques for early lung cancer diagnoses. In *Green Computing, Communication and Conservation of Energy (ICGCE), 2013 International Conference on* (pp. 918-922). IEEE.

[21] Alfisahrin, S. D. N. N., & Mantoro, T. (2013, December). Data Mining Techniques for Optimization of Liver Disease Classification. In *Advanced Computer Science Applications and Technologies (ACSAT), 2013 International Conference on* (pp. 379-384). IEEE.

[22] Ranganatha, S., Raj, H. P., Anusha, C., & Vinay, S. K. (2013). Medical data mining and analysis for heart disease dataset using classification techniques.

[23] Agarwal, Y., & Pandey, H. M. (2014, September). 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), 2014 5th International Conference-* (pp. 72-76). IEEE.

[24] Banu, N., & Gomathy, B. (2014, March). Disease Forecasting System Using Data Mining Methods. In *Intelligent Computing Applications (ICICA), 2014 International Conference on* (pp. 130-133). IEEE.

[25] Xiong, X., Kim, Y., Baek, Y., Rhee, D. W., & Kim, S. H. (2005, May). Analysis of breast cancer using data mining & statistical techniques. In *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, 2005 and First ACIS International Workshop on Self-Assembling Wireless Networks. SNPD/SAWN 2005. Sixth International Conference on* (pp. 82-87). IEEE.

[26] 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, 2006. CBMS 2006. 19th IEEE International Symposium on* (pp. 714-718). IEEE.

[27] Menolascina, F., Tommasi, S., Paradiso, A., Cortellino, M., Bevilacqua, V., & Mastronardi, G. (2007, April). Novel data mining techniques in aCGH based breast cancer subtypes profiling: the biological perspective. In *Computational Intelligence and Bioinformatics and Computational Biology, 2007. CIBCB'07. IEEE Symposium on* (pp. 9-16). IEEE.

[28] Fan, Q., Zhu, C. J., & Yin, L. (2010, April). Predicting breast cancer recurrence using data mining techniques. In *Bioinformatics and Biomedical Technology (ICBBT), 2010 International Conference on* (pp. 310-311). IEEE.

[29] Abdelaal, M. M. A., Farouq, M. W., Sena, H. A., & Salem, A. B. M. (2010, October). Using data mining for assessing diagnosis of breast cancer. In*Computer Science and Information Technology (IMCSIT), Proceedings of the 2010 International Multiconference on* (pp. 11-17). IEEE.

[30] Vijayarani, S., & Dhayanand, M. S. KIDNEY DISEASE PREDICTION USING SVM AND ANN ALGORITHMS.

[31] Chiu, R. K., Chen, R. Y., Wang, S. A., & Jian, S. J. (2012, July). Intelligent systems on the cloud for the early detection of chronic kidney disease. In*Machine Learning and Cybernetics (ICMLC), 2012 International Conference on*(Vol. 5, pp. 1737-1742). IEEE

[32] Lakshmi, K. R., Nagesh, Y., & VeeraKrishna, M. (2014). Performance Comparison of Three Data Mining Techniques for Predicting Kidney Dialysis Survivability. *International Journal of Advances in Engineering & Technology (IJAET)*, *7*(1), 242-254.

[33] Xun, L., Xiaoming, W., Ningshan, L., & Tanqi, L. (2010, October). Application of radial basis function neural network to estimate glomerular filtration rate in Chinese patients with chronic kidney disease. In *Computer Application and System Modeling (ICCASM), 2010 International Conference on* (Vol. 15, pp. V15-332). IEEE.

[34] Ravindra, B. V., Sriraam, N., & Geetha, M. (2014, November). Discovery of significant parameters in kidney dialysis data sets by K-means algorithm. In*Circuits, Communication, Control and Computing (I4C), 2014 International Conference on* (pp. 452-454). IEEE.

[35] Ahmed, S., Tanzir Kabir, M., Tanzeem Mahmood, N., & Rahman, R. M. (2014, December). Diagnosis of kidney disease using fuzzy expert system. In*Software, Knowledge, Information Management and Applications (SKIMA), 2014 8th International Conference on* (pp. 1-8). IEEE.

[36] Han, J., Kamber, M., & Pei, J. (2006). *Data mining, southeast asia edition: Concepts and techniques*. Morgan Kaufmann.

[37] https://msdn.microsoft.com/en-us/library/ms167167.aspx

[38] http://www.tutorialspoint.com/data_mining/

[39] http://www.oracle.com/technetwork/database/enterprise-edition/odm-techniques-algorithms-097163.html

[40] http://dms.irb.hr/tutorial/tut_main.php