South Asian Journal of Engineering and Technology Research Article

Open Access Full Text Article 123(2022) 69-77 DOI: 10.26524/sajet.2022.12.41

DETECTING CHRONIC KIDNEY DISEASE BY USING A VOTING CLASSIFIER

M.Rama [1], G.Tejaswini [2], T.Niharika [3], K.Murari [4], M.Bhanu Prakash[5]

Assistant Professor (IT), Dept. of Information Technology, QIS College of Engineering & Technology, India(1).

Final Year (B.Tech), Dept. of Information Technology, QIS College of Engineering & Technology, India(2,3,4,5)

Corresponding author.
Correspondence: M.Rama
E-mail: ramaha@gmail.com

Article info
Received 12 th March 2022 Received in revised form 16 April 2022 Accepted 7 July 2022

Keywords
CKD,KNN,SVM

Abstract
Chronic Kidney Disease (CKD) is a global health problem with high morbidity and mortality rate, and it induces other diseases. Since there are no obvious symptoms during the early stage of CKD, patients often fail to notice the disease. Early detection of CKD enables patients to receive timely treatment to ameliorate the progression of this disease. Machine learning models can effectively aid clinicians achieve this goal due to their fast and accurate recognition performance. In this, we proposed a machine learning methodology for diagnosing CKD. The CKD data set was obtained from the University of California Irvine (UCI) machine learning repository, which has a large number of missing values. In general we have six machine learning algorithms like logistic regression, random forest, support vector machine, k-nearest neighbour, naive Bayes, classifier and feed forward neural network to establish models. In existing system random forest achieved the performance with 99.75% diagnosis accuracy. In proposed system, By analysing the misjudgments generated by the above model, we proposed an integrated model that combines logistic regression and random forest by using perceptron which could achieve an efficient accuracy of 99.83% after ten times of simulation. Hence, we speculated that this methodology could be applicable to more complicated clinical data for disease diagnosis.

Introduction:

1.Background

Chronic Kidney Disease is a global health problem affecting approximately 10% of the world population. The percentage of prevalence of CKD in China is 10.8% and the range of prevalence is 10%-15% in the United States. According to another study, this percentage has reached 14.7% in the Mexican adult general population. This disease is characterised by a slow deterioration in renal function, which eventually causes a complete loss of renal function. CKD does not show obvious symptoms in its early stages. Therefore, the disease may not be detected until the kidney loses about 25% of its function. In addition, CKD has high morbidity and mortality, with a global impact on the human body. It can induce the occurrence of cardiovascular disease. CKD is a progressive and irreversible pathologic syndrome. Hence, the prediction and diagnosis of CKD in its early stages is quite essential, it may be able to enable patients to receive timely treatment to ameliorate the progression of the disease.

Machine learning refers to a computer program, which calculates and deduces the information related to the task and obtains the characteristics of the corresponding pattern. This technology can achieve accurate and economical diagnoses of diseases; hence, it might be a promising method for diagnosing CKD. It has become a new kind of medical tool with the development of information technology and has a broad application prospect because of the rapid development of electronic health records. In the medical field, machine learning has been used to detect human body status, analyze the relevant factors of the disease and diagnose various diseases. For example, the models built by machine learning algorithms were used to diagnose heart disease, diabetes and retinopathy, acute kidney injury, cancer and other diseases. In these models, algorithms based on regression, tree, probability, decision surface and neural network were often effective. In the field of CKD diagnosis, Hodneland et al. utilized image registration to detect renal morphologic changes. Vasquez-Morales et al. established a classifier based on neural networks using large-scale CKD data, and the accuracy of the model on their test data was 95%. In addition, most of the previous studies utilized the CKD data set that was obtained from the UCI machine learning repository. Chen et al. used k-nearest neighbour (KNN), support vector machine (SVM) and soft independent modelling of class analogy to diagnose CKD, KNN and SVM achieved the highest accuracy of 99.7%. In addition, they used fuzzy rule-building expert system, fuzzy optimal associative memory and partial least squares discriminant analysis to diagnose CKD, and the range of accuracy in those models was 95.5%- 99.6%. Their studies have achieved good results in the diagnosis of CKD. In the above models, the mean imputation is used to fill in the missing values and it depends on the diagnostic categories of the samples. As a result, their method could not be used when the diagnostic results of the samples are unknown. In reality, patients might miss some measurements for various reasons before diagnosing. In addition, for missing values in categorical variables, data obtained using mean imputation might have a large deviation from the actual values. For example, for variables with only two categories, we set the categories to 0 and 1, but the mean of the variables might be between 0 and 1. Polat et al. developed an SVM based on feature selection technology, the proposed models reduced the computational cost through feature selection, and the range of accuracy in those models was from 97.75%-98.5%. J. Aljaaf et al. used novel multiple imputation to fill in the missing values, and then MLP neural network (MLP) achieved an accuracy of 98.1%. Subas et al. used MLP, SVM, KNN, C4.5 decision tree and random forest (RF) to diagnose CKD, and the RF achieved an accuracy of 100%. In the models established by Boukenze et al., MLP achieved the highest accuracy of 99.75% . The studies focus mainly on the establishment of models and achieve an ideal result. However, a complete process of filling in the missing values is not described in detail, and no feature selection technology is used to select predictors as well. Almansour et al. used SVM and neural network to diagnose CKD, and the accuracy of the models was 97.75% and 99.75%, respectively. In the models established by Gunarathne et al., zero was used to fill out the missing values and decision forest achieved the best performance with the accuracy was 99.1%.

To summarize the previous CKD diagnostic models, we find that most of them suffering from either the method used to impute missing values has a limited application range or relatively low accuracy. Therefore, in this work, we propose a methodology to extend application range of the CKD diagnostic models. At the same time, the accuracy of the model is further improved.

2.Problem Statement

Chronic kidney disease (CKD) is a global public health problem affecting approximately 10% of the world's population [1], [2]. The percentage of prevalence of CKD in China is 10.8% [3], and the range of prevalence is 10%-15% in the United States [4]. According to another study, this percentage has reached 14.7% in the Mexican adult general population [5]. This disease is characterised by a slow deterioration in renal function, which eventually causes a complete loss of renal function. 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 medicine.

3.Objectives

- To check weather the person is affected with Chronic Kidney Disease or not
- To check the stage of the affected person
- Comparing the model with another model

However, for the bias, when the updating method of (14) was used, the obtained decision line could classify the samples, but the line was located at the edge of the solution area, so it is not reliable. To solve this problem, a new bias adjustment strategy proposed in chapter 4 of the previous literature was referred to. The proposed models reduced the computational cost through feature selection, and the range of accuracy in those models was from 97.75%-98.5%. J. Aljaaf et al. used novel multiple imputation to fill in the missing values, and then MLP neural network (MLP) achieved an accuracy of 98.1%. Subas et al. used MLP, SVM, KNN, C4.5 decision tree and random forest (RF) to diagnose CKD, and the RF achieved an accuracy of 100%. In the models established by Boukenze et al., MLP achieved the highest accuracy of 99.75%. The studies focus mainly on the establishment of models and achieve an ideal result. However, a complete process of filling in the missing values is not described in detail, and no feature selection technology is used to select predictors as well. Almansour al. used SVM and neural network to diagnose CKD, and the accuracy of the models was 97.75% and 99.75%, respectively. In the models established by Gunarathne et al., zero was used to fill out the missing values and decision forest achieved the best performance with the accuracy was 99.1%.

Related Work:

Classification techniques, Feature selection,

and Ensemble model are the most significant and vital tasks in machine learning and data mining. A lot of research has been conducted to apply data mining and machine learning classification technique, feature selection method and ensemble model on different medical datasets to classify disease datasets. Many of them show good classification accuracy.

Polat, H et al. [6] Diagnosis chronic kidney disease using SVM and effective feature selection methods. They used wrapper and filter feature selection method to reduce the dimensionality of the feature. In their work they improve accuracy by implanting SVM without feature selection the accuracy rate was 97.75%, SVM with the classifier subset evaluator combine with greedy stepwise the accuracy rate was 98%, SVM with the wrapper subset evaluator combine with a best first search engine the accuracy rate was 98.25, SVM with the classifier subset evaluator combine with greedy stepwise the accuracy rate 98.25. And finally, SVM with the filter subset evaluator combine with best first search the accuracy was 98.5.

Bashir, S. et al [15] they proposed ensemble classifier which uses majority Vote Based framework for prediction of heart disease. They used five heterogeneous classifiers used to construct the ensemble model. The classifiers are Naïve Bayes, decision tree based on Gini Index, decision tree based on information Gain, memory based learner and SVM. After experiment using stratified cross-validation show that their MV5 framework has achieved an accuracy 88.5% with 86.96% sensitivity, 90.83% specificity and 88.85 F-Measure and they compare with the base classifiers show to increase the average accuracy of the ensemble model. They involved proposing the ensemble approach. The first approach generates the individual classifier decision and the second approach is combine the individual classifier decision correctly to create the new combine model.

Bashir, S., [33] proposed HMV medical decision support framework using multi-layer classifier for disease prediction. They proposed based on the optimal combination of the heterogeneous classifier to create the ensemble model. The classifiers are Naïve Bayes, Linear Regression, and quadratic discriminate analysis, KNN, SVM, Decision Tree using Gini Index, and Decision Tree using Information Gain. So, their HMV ensemble framework outperforms the other prediction models. HMV framework was proposed based on three modules. The first module was data acquisition and preprocessing. The second module was used to predict unknown class label for test set instances. The third module used to predict and evaluate the proposed HMV ensemble model. After applying all the selected data set the HMV ensemble model achieved highest accuracy disease classification and prediction.

Naghmeh Khajehali et al.[4] were presented by extracting factor affecting for pneumonia patients by using data mining techniques. They proposed modelling by using feature selection and classification with ensemble methods to preprocess, reduce dimensionality and classify the raw data. In their work, the design consists of different stages of preprocessing and used Bayesian Boosting method for constructed which identify factor related to patient LOS in hospital. The construction of modeling based on the data set SVM and ensemble method like AdaBoost, Vote, Stacking, Bayesian Boosting. Among these classifier techniques,

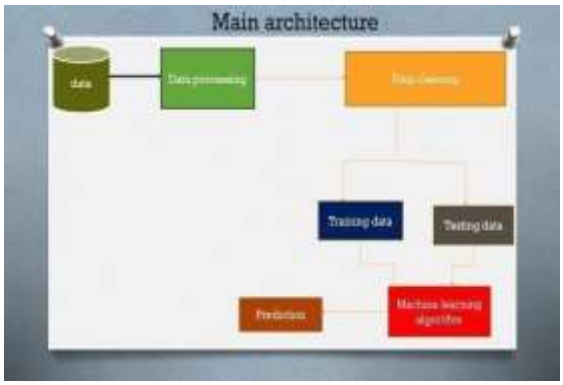
Bayesian Boosting used for analysis of data by using 10 fold cross-validation method. In this work the data set was divided into 10 subsets, the training subset participated 10 times. Out of 10 subsets 9 were classified as the training set. The result indicated that the Bayesian Boosting ensemble technique was scored a better result. The Bayesian boosting ensemble technique, accuracy was 97.17%, which is high performance used to predict pneumonia disease in anticipation of LOS.

Pritom, A. I et al. [12] applied a classification algorithm for Predicting Breast Cancer Recurrence by using SVM, Decision tree, Naïve Bays and C4.5. They enhanced the accuracy of each classifier with the help of effective feature selection methods. They improve the accuracy by using Info Gain attribute with ranker search engine. After implemented on weka tool the recurrence prediction accuracy has SVM achieved 75.75% accuracy, J48 achieved 73.73% and naïve bays achieved 67.17%. These are the original data set without feature selection. After carefully applied feature selection SVM enhanced bay 1.52%, C4.5 enhanced by 2.52% and Naïve Bays enhanced by 9.09%.

Dulhare, U. N. et al. [10] Built classification models, Used feature selection to extract an action rule and predict CKD by using naïve Bayes classifier and one R attribute selector to predict and classify the CKD and none CKD patients. These methods are Naïve bays with the wrapper subset evaluator combine with the best first search. After implemented on weka tool using wrapper subset evaluators combine with the best first search engine; Naïve Bay's classifier achieved 97.5 % accuracy rate.

Architecture:

LOG and RF were selected as underlying components to generate the integrated model to improve the performance of judging. The probabilities that each sample was judged as notckd in LOG and RF were used as the outputs of underlying components. These two probabilities of each sample were obtained and could be expressed in a two-dimensional plane. In the complete CKD data sets, the probability distributions of the samples in a two-dimensional plane are similar. Therefore, the probability distribution of samples when K equalling to 11 is shown in Fig. 3. It can be seen from Fig. 3 that the samples have different aggregation regions in the two-dimensional plane due to the different categories (ckd or notckd). In general, samples with ckd are concentrated in the lower left part, while the notckd samples are distributed in the top right part. Due to the fact that the results in the two models are different, some samples are located at the top left and lower right, and one of the two models makes the misjudgements. Perceptron can be used to separate samples of two categories by plotting a decision line in the two- dimensional plane of the probability distribution. Ciaburro and Venkateswaran defined perceptron as the basic building block of a neural network, and it can be understood as anything that requires multiple inputs and produces an output .



Objective Function: The perceptron used in this study is shown in Fig. 4. In Fig. 4, prob1 and prob2 are the probabilities that a sample was judged as notckd by LOG and RF,respectively. w0, w1 and w2 are the weights of input signals. w0 corresponds to 1, w1 corresponds to prob1 and w2 corresponds to prob2, respectively. y is calculated according to (7):

$$y = w0 + w1 \times prob1 + w2 \times prob2.$$

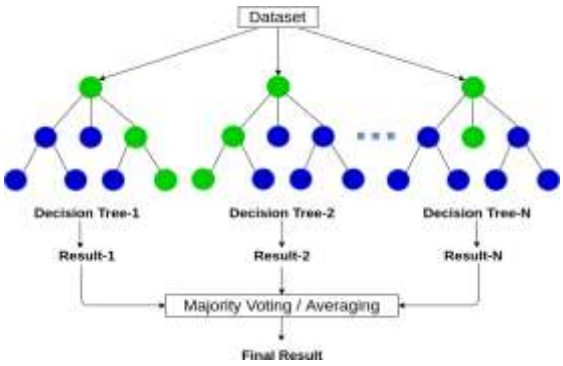
The probability distribution of the samples in the complete CKD data set (at K = 11) the horizontal axis and the vertical axis represent the probabilities that the samples were judged as notckd by the LOG and the RF, respectively. FIGURE 4. The structure of the perceptron used in this study. The input signal corresponding to the weight w0 is 1, which is a bias. The function of Signum is used to calculate output by processing the value of y as follows: If y > 0, then the output = 1, whereas if y < 0, then the output = −1. For the output, 1 corresponds to notckd, whereas -1 corresponds to ckd. A single perceptron is a linear classifier that can be used to detect binary targets. The weights are the core of the perceptron and adjusted in the training stage. y = 0 is the decision line, and this line can be described as (8):

prob2 = − w1 w2prob1 − w0 w2

ALGORITHM DESCRIPTION

Random forest

- Random Forest is a trademark term for an ensemble of decision trees.
- In Random Forest, we've collection of decision trees (so known as "Forest").
- Toclassify a new object based on attributes, each tree gives a classification and we say the tree "votes" for that class.
- The forest chooses the classification having the most votes (over all the trees in the forest).



Logistic Regression

- It is a classification not a regression algorithm.
- It is used to estimate discrete values based on given set of independent variable(s).
- it predicts the probability of occurrence of an event by fitting data to a logit function. Hence, it is also known as logit regression.

Module Descriptions:

Data Collection Module:

Be it the raw data from excel, access, text files etc., this step (gathering past data) forms the foundation of the future learning. The better the variety, density and volume of relevant data, better the learning prospects for the machine becomes.

Preparing the data Module:

Any analytical process thrives on the quality of the data used. One needs to spend time determining the quality of data and then taking steps for fixing issues such as missing data and treatment of outliers. Exploratory analysis is perhaps one method to study the nuances of the data in details thereby burgeoning the nutritional content.

Training a model:

This step involves choosing the appropriate algorithm and representation of data in the form of the model. The cleaned data is split into two parts – train and test (proportion depending on the prerequisites); the first part (training data) is used for developing the model. The second part (test data), is used as a reference.

Disease prediction Module:

Disease will be predicted based on the results of the algorithms Provided.

Data Set:

The CKD data set used in this study was obtained from the UCI machine learning repository.

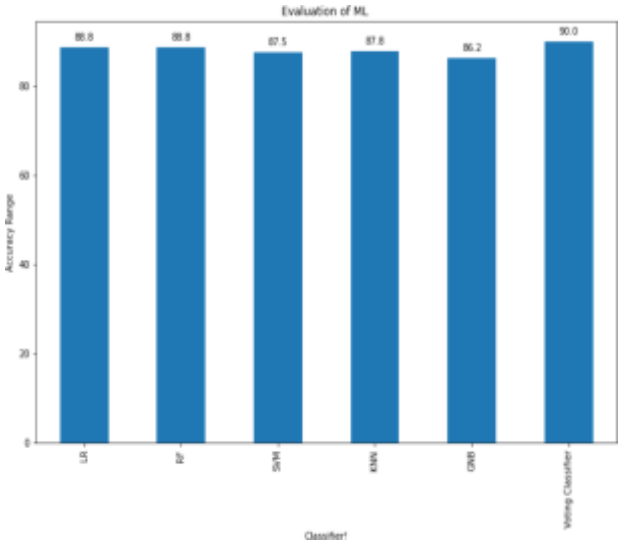
TABLE 1. Details of each variable in the original CKD data set.

Variables	Explain	Class	Scale	Missing Rate
age	Age	Numerical	age in years	2.25%
bp	Blood Pressure	Numerical	in mm/Hg	3%
sg	Specific Gravity	Nominal	(1.005,1.010,1.015,1.020,1.025)	11.75%
al	Albumin	Nominal	(0,1,2,3,4,5)	11.5%
su	Sugar	Nominal	(0,1,2,3,4,5)	12.25%
rbc	Red Blood Cells	Nominal	(normal,abnormal)	38%
pc	Pus Cell	Nominal	(normal,abnormal)	16.25%
pcc	Pus Cell clumps	Nominal	(present,notpresent)	1%
ba	Bacteria	Nominal	(present,notpresent)	1%
bgr	Blood Glucose Random	Numerical	in mgs/dl	11%
bu	Blood Urea	Numerical	in mgs/dl	4.75%
sc	Serum Creatinine	Numerical	in mgs/dl	4.25%
sod	Sodium	Numerical	in mEq/L	21.75%
pot	Potassium	Numerical	in mEq/L	22%
hemo	Hemoglobin	Numerical	in gms	13%
pcv	Packed Cell Volume	Numerical	-	17.75%
wbcc	White Blood Cell Count	Numerical	in cells/cumm	26.5%
rbcc	Red Blood Cell Count	Numerical	in millions/cmm	32.75%
htn	Hypertension	Nominal	(yes,no)	0.5%
dm	Diabetes Mellitus	Nominal	(yes,no)	0.5%
cad	Coronary Artery Disease	Nominal	(yes,no)	0.5%
appet	appet	Nominal	(good,poor)	0.25%
pe	Pedal Edema	Nominal	(yes,no)	0.25%
ane	Anemia	Nominal	(yes,no)	0.25%
class	Class	Nominal	(ckd,notckd)	0%

which was collected from hospital and donated by Soundarapandian et al. on 3rd July, 2015. The data set contains 400 samples. In this CKD data set, each sample has 24 predictive variables or features (11 numerical variables and 13 categorical (nominal) variables) and a categorical response variable (class). Each class has two values, namely, ckd (sample with CKD) and notckd (sample without CKD). In the 400 samples, 250 samples belong to the category of ckd, whereas 150 samples belong to the category of notckd. It is worth mentioning that there is a large number of missing values in the data.

The details of each variable are listed in Table.

Evaluation of All Models



Conclusion

The proposed CKD diagnostic methodology is feasible in terms of data imputation and sample diagnosis. After unsupervised imputation of missing values in the data set by KNN imputation, the integrated model could achieve satisfactory accuracy. Hence, we speculate that applying this methodology to the practical diagnosis of CKD would achieve a desirable effect. In addition, this methodology might be applicable to the clinical data of the other diseases in actual medical diagnosis. However, in the process of establishing the model, due to the limitations of the conditions, the available data samples are relatively small, including only 400 samples. Therefore, the generalization performance of the model might be limited. In addition, due to there are only two categories (ckd and notckd) of data samples in the data set, the model can not diagnose the severity of CKD.

In the future, a large number of more complex and representative data will be collected to train the model to improve the generalization performance while enabling it to detect the severity of the disease. We believe that this model will be more and more perfect by the increase of size and quality of the data.

References

1. K. Chandel, V. Kunwar, S. Sabitha, T. Choudhury, and S. Mukherjee, "A comparative study on thyroid disease detection using K-nearest neighbor and Naive Bayes classification techniques," CSI Trans. ICT, vol. 4, no. 2–4, pp. 313–319, Dec. 2016.
2. B. Boukenze, A. Haqiq, and H. Mousannif, "Predicting Chronic Kidney Failure Disease Using Data Mining Techniques," in Advances in Ubiquitous Networking 2, Springer, Singapore, 2017, pp. 701–712.
3. T. Shaikhina, D. Lowe, S. Daga, D. Briggs, R. Higgins, and N. Khovanova, "Decision tree and random forest models for outcome prediction in antibody incompatible kidney transplantation," Biomed. Signal Process. Control, Feb. 2017.
4. R. Ani, G. Sasi, U.R. Sankar, & O.S. Deepa, (2016, September). "Decision support system for diagnosis and prediction of chronic renal failure using random subspace classification – I EEE Conference Publication."[Online]. Available: http://ieeexplore.ieee.org/abstract/document/7732224/?reload=true. [Accessed: 15-Dec-2017].

5. L.-C. Cheng, Y.-H. Hu, and S.-H. Chiou, "Applying the Temporal Abstraction Technique to the Prediction of Chronic Kidney Disease Progression," J. Med. Syst., vol. 41, no. 5, p. 85, May 2017.
6. H. Polat, H. D. Mehr, and A. Cetin, "Diagnosis of Chronic Kidney Disease Based on Support Vector Machine by Feature Selection Methods," J. Med. Syst., vol. 41, no. 4, p. 55, Apr. 2017.
7. P. Pangong and N. Iam-On, "Predicting transitional interval of kidney disease stages 3 to 5 using data mining method," in 2016 Second Asian Conference on Defence Technology (ACDT), 2016, pp. 145–150.
8. K. R. A. Padmanaban and G. Parthiban, "Applying Machine Learning Techniques for Predicting the Risk of Chronic Kidney Disease," Indian J. Sci. Technol., vol. 9, no. 29, Aug. 2016.
9. S. Perveen, M. Shahbaz, A. Guergachi, and K. Keshavjee, "Performance Analysis of Data Mining Classification Techniques to Predict Diabetes," Procedia Comput. Sci., vol. 82, no. Supplement C, pp. 115–121, Jan. 2016.
10. U. N. Dulhare and M. Ayesha, "Extraction of action rules for chronic kidney disease using Naive Bayes classifier," in 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2016, pp. 1–5.
11. N. Borisagar, D. Barad, and P. Raval, "Chronic Kidney Disease Prediction Using Back Propagation Neural Network Algorithm," in Proceedings of International Conference on Communication and Networks, Springer, Singapore, 2017, pp. 295–303.

A. I. Pritom, M. A. R. Munshi, S. A. Sabab, and S. Shihab, "Predicting breast cancer recurrence using effective classification and feature selection technique," in 2016 19th International Conference on Computer and Information Technology (ICCIT), 2016, pp. 310–314.

12. S. Mishra, P. Chaudhury, B. K. Mishra, and H. K. Tripathy, "An Implementation of Feature Ranking Using Machine Learning Techniques for Diabetes Disease Prediction," in Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies, New York, NY, USA, 2016, p. 42:1–42:3.
13. D. Zufferey, T. Hofer, J. Hennebert, M. Schumacher, R. Ingold, and S. Bromuri, "Performance comparison of multi-label learning algorithms on clinical data for chronic diseases," Comput. Biol. Med., vol. 65, no. Supplement C, pp. 34–43, Oct. 2015.
14. S. Bashir, U. Qamar, F. H. Khan, and M. Y. Javed, "MV5: A Clinical Decision Support Framework for Heart Disease Prediction Using Majority Vote Based Classifier Ensemble," Arab. J. Sci. Eng., vol. 39, no. 11, pp. 7771–7783, Nov. 2014.
15. T.R. Baitharu, & S.K. Pani, (2016). "Analysis of Data Mining Techniques for Healthcare Decision Support System Using Liver Disorder Dataset- ScienceDirect." [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050916306263. [Accessed: 15-Dec-2017].
16. Z. Sedighi, H. Ebrahimpour-Komleh, and S. J. Mousavirad, "Feature selection effects on kidney disease analysis," in 2015 International Congress on Technology, Communication and Knowledge (ICTCK), 2015, pp. 455–459.
17. D.M. Filimon, & A. Albu, (2014, May "Skin diseases diagnosis using artificial neural networks - IEEE Conference Publication." [Online]. Available: http://ieeexplore.ieee.org/abstract/document/6840059/. [Accessed: 15-Dec-2017]
18. F. Ahmad, N. A. M. Isa, Z. Hussain, M. K. Osman, and S. N. Sulaiman, "A GA-based feature selection and parameter optimization of an ANN in diagnosing breast cancer," Pattern Anal. Appl., vol. 18, no. 4, pp. 861–870, Nov. 2015.
19. H. Asri, H. Mousannif, H. Al Moatassime, & T. Noel, (2016). "Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis ScienceDirect." [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050916302575. [Accessed: 15-Dec-2017].
20. B.L Deekshatulu, & P. Chandra, (2013 "Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm ScienceDirect." [Online].Available: http://www.sciencedirect.com/science/article/pii/S2212017313004945. [Accessed: 15-Dec-2017].

21. S. Ramya, and N. Radha. "Diagnosis of chronic kidney disease using machine learning algorithms." International Journal of Innovative Research in Computer and Communication Engineering vol. 4, no. 1 pp. 812-820. 2016.
22. R. Dhruvi, R. Yavnika, & R. Nutan, " Prediction of Probability of Chronic Diseases and Providing Relative Real-Time Statistical Report using data mining and machine learning techniques". International Journal of Science, Engineering, and Technology Research (IJSETR) vol. 5, no. 4. 2016.
23. Compound feature generation and boosting model for cancer gene classification Ibrahim, S. Jafar Ali Ibrahim., Affir, A.M., Thangamani, M. *International Journal of Engineering Trends and Technology, 2020, 68(10), pp. 48–51, Doi No:*doi:10.14445/22315381/IJETT-V68I10P208 *https://ijettjournal.org/Volume-68/Issue-10/IJETT-V68I10P208.pdf*
24. S. Vijayarani, S. Dhayanand, and M. Phil. "Kidney disease prediction using SVM and ANN algorithms." International Journal of Computing and Business Research (IJCBR) vol. 6, no. 2, 2015.
25. N. Chetty, S. V. Kunwar, and S. D. Sudarsan. "Role of attributes selection in the classification of Chronic Kidney Disease patients." In Computing, Communication and Security (ICCCS), 2015 International Conference on, pp. 1-6. IEEE, 2015.
26. http://www.datasciencecentral.com/profiles/blogs/python-resources-fortop-data-mining-algorithms
27. https://archive.ics.uci.edu/ml/datasets/chronic_kidney_disease
28. J. Radhakrishnan, and M. Sumit, "KI Reports and World Kidney Day." Kidney international reports vol.2, no. 2, pp. 125-126, Mar. 2017.
29. https://www.kidney.org/kidneydisease/global-facts-about-kidneydisease#_ENREF_1
30. P. Ahmad, Q. Saqib, and S. Q. A. Rizvi. "Techniques of data mining in healthcare: a review." International Journal of Computer Applications Vol. 120, no. 15 Jan. 2015.
31. B. R. Sharma, K. Daljeet, and A. Manju. "Review on Data Mining: Its Challenges, Issues and Applications." International Journal of Current Engineering and Technology vol. 3, no. 2 jun. 2013.
32. Bashir, S., Qamar, U., Khan, F. H., & Naseem, L. (2016). H. MV: a medical decision support framework using multi-layer classifiers for disease prediction. Journal of Computational Science, 13, 10-25
33. Ibrahim, S. Jafar Ali, and M. Thangamani. "Prediction of Novel Drugs and Diseases for Hepatocellular Carcinoma Based on Multi-Source Simulated Annealing Based Random Walk." *Journal of medical systems* 42, no. 10 (2018): 188. https://doi.org/10.1007/s10916-018-1038-y