Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

Diagnosis Kidney Function Test Using Machine Learning Algorithm Based on Runge Kutta Method

Dr. K. Kartheeban
Department of Computer Applications
Kalasalingam Academy of Research and Education
Krishnan Koil, TamilNadu, India
k.kartheeban73@gmail.com, hodmca@klu.ac.in
Dr. D. Kalpanadevi
Department of Computer Applications
Kalasalingam Academy of Research and Education
Krishnan Koil, TamilNadu, India
kalpanapani@gmail.com, dkalpanadevi@klu.ac.in

Article Info
Page Number: 87-98
Publication Issue:
Vol 71 No. 3s2 (2022)

Article History
Article Received: 28 April 2022
Revised: 15 May 2022
Accepted: 20 June 2022
Publication: 21 July 2022

Abstract
A main objective of the research to analyse the diagnosis of Kidney function test by the fluctuation of parameter in learning level. In this research work focuses on the machine learning techniques which treat to implement by supervised learning method by the comparison of three algorithm, such as Hybrid Method- AdaBoost with C4.5 algorithm using Runge Kutta method, 10-fold- cross validation in C4.5 and Random Forest under bagging method. Finally, an experimental performance can be analysed by applying machine learning technique to diagnosis patients' medical data and analyse the work performance of these learning algorithm, which one is better in performance analysis.

Keywords: - Hybrid Method- AdaBoost with C4.5 algorithm using Runge Kutta method, 10-fold- cross validation in C4.5 and random forest under Bagging Method, Diagnosis of Kidney Function test

I. INTRODUCTION

From existing method, several data sets can predict kidney failure with limited attributes. Some authors can suggest hybrid approach for getting performance accuracy for kidney data sets for the relevant causes of function test can be diagnosis. In this proposed work, machine learning algorithm can be used to finding the pattern of health level among the collection of kidney function tests for diagnosis and decision making.

The supervised learning machine learning techniques can be applying several techniques for comparing such as

- Hybrid Method- AdaBoost with C4.5 algorithm using Runge Kutta method
- 10-fold- cross validation in C4.5 and
- Random Forest under bagging method

In order to collect the patients functioning test predominantly focused on and predicting the functioning test level under the comparison of these three algorithms.

II. LITERATURE REVIEW

From the survey, the National foundation of kidney centre, the kidney disease can be mostly affected to American adults. The period of forecasting, function of kidney tests an approximate CARG is 5.7% [44]. Around 30.3 million people in the U.S. [38], have diabetes, the report was taken from National Diabetes Statistics (2017). Hence, the global kidney function test market owing to increasing prevalence of diabetes can be expected by Asia Pacific in rising geriatric population, high blood pressure for unhealthy lifestyle. An improper functioning of the kidney can be suffered from the affection of high blood pressure which can damage the blood vessel of the kidney [38-39]. In the year 2017, 70 million of people in Europe suffered from some or the other type of kidney disease which can be surveyed from European Renal Association and the European Dialysis and Transplant Association [36][37]. The diagnostic measures can be used to developing the machinery and devices that need for further use to help people who suffer from kidney disease which can help to give several treatments for avoiding the severe condition of patients much more quickly.

III. RESEARCH METHODOLOGY

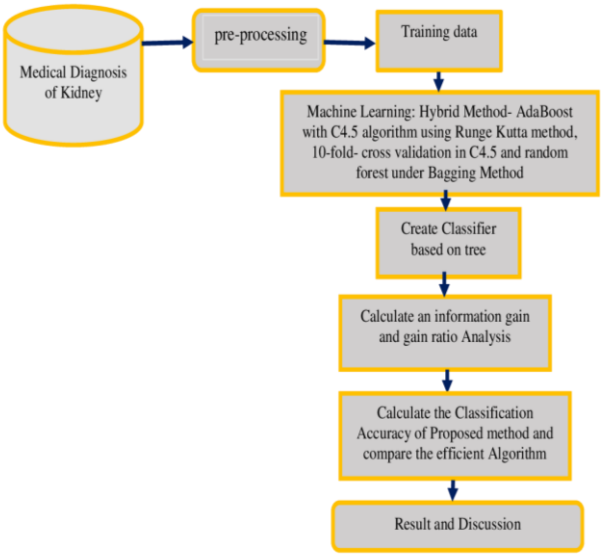

Figure.1 Frame work Architecture

By making a series of optimum decision about which attributes to use for partitioning the data. Greedy method can be used for constructing the decision tree for given set of attributes. Typically, machine learning techniques can be implemented for obtaining high accuracy of classification, time efficiency of medical diagnosis data.

The performance can be identified and diagnosis for kidney function test under level of blood urea, uric acid. Figure.1, represents the given parameters which has to measure the medical diagnosis of kidney function test data. For the process of classification process, pre-processing can be taken for converting the raw data of numerical data into nominal data. After the pre-process, apply three algorithms to analyse the accuracy of classification process.

Random Forest Under Bagging Method

1. For producing a predicted class able at input point ‘a’, let us assume the classifier as A (T, a) based on the training data S. To draw bootstrap samples like T^{*1}, T^{*D} each of size M with

replacement from the training data and include into the bag A. Then to find the majority vote as assignment of Abag (x) =Majority Vote {A (T∗b, y)} B where b=1. Here bagging can reduced the variance of unstable data in tree which can lead to im[rove the prediction efficiently [5 – 16].

2. To split the node iteratively when the subset of N feature can randomly select.

3. When the tree is grown in large, this algorithm can improve on bagging by “de-correlating” the trees.

4. Repeatedly above steps can follow and prediction is evolved from n number of trees based on aggregation as shown in figure 2.

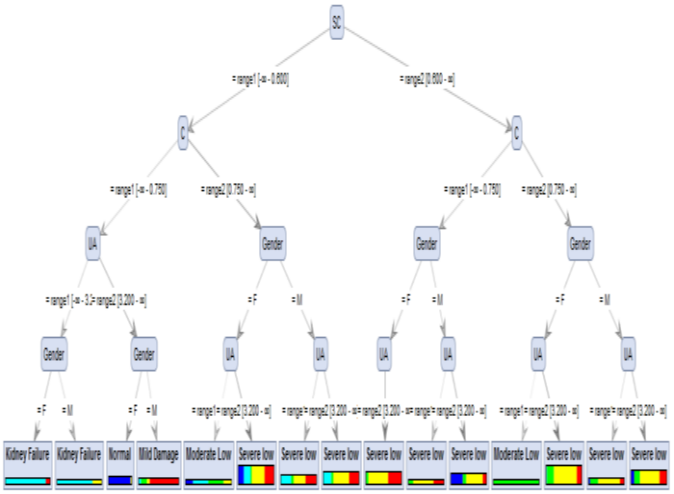

Figure 2 Tree view using Bagging with Random Forest

5. After training the decision tree, the new data can be classified through majority rule. From the above figure 2 represents the decision tree makes to predicts the data by selecting the class label that received more votes by making decision. The bagging classifier would have a lower variance like it is less over-fitting than an individual decision tree and the result can be found less complex decision boundary.

C4.5 With 10-Fold- Cross Validation

Cross-Validations, K- fold performance can be followed for estimating the unseen data performance of a classifier. The labelled data set split randomly into K equal partitions [17- 28].

- The remaining K-1 partitions can be trained on each partition and be tested from that partition.
- Getting average of all K accuracies can be analysed in the tree view as shown in figure 3.

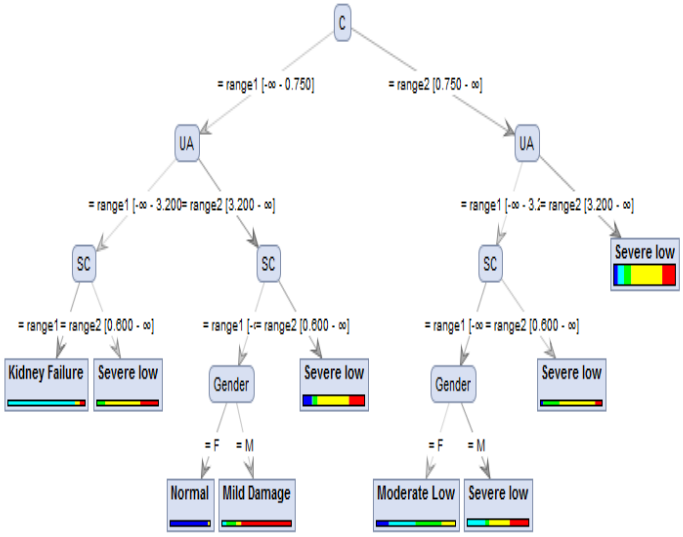

Figure 3 Tree view using 10-fold- cross Validation in 4.5

The rule base has written for the classify the prediction of class using bagging with C4.5 Algorithm, the results as shown a tree in graph view as shown in figure 3.

CAL = Normal
| SC = Normal: Normal (62.0)
| SC = Low: Mild Damage (27.0/12.0)
| SC = High: Normal (0.0)
CAL = Low
| TP = Normal: Severe low (334.0/143.0)
| TP = Low: Kidney Failure (19.0/1.0)
| TP = High
| | C = Normal: Severe low (0.0)
| | C = Low: Severe low (44.0/21.0)
| | C = High: Mild Damage (14.0/6.0)

Hybrid Method-Adaboost with C4.5 Algorithm Using Runge Kutta Method

In the proposed research work, the whole data set can be Hybrid Method- AdaBoost classification algorithm which is performed to identify the kidney patients' factor who are suffering from fluctuation of ranges by the collection of Urea nitrogen, Creatine, total protein, Phosphate, Albumin, Calcium, Bi- Carbonate, Potassium, Sodium, Uric acid in gender wise, Serum in gender wise. This research should be possible to design interventions to prevent the disease and its complications [29-36].

- For the binary classification, AdaBoost was developed for boosting algorithm.
- Here Runge- Kutta boosting can be applied in machine learning for understanding the boosting task in AdaBoost method.
- To overcome the problem of the Euler's method, Runge-Kutta method is the best choice of a sufficiently small step size can reach a reasonable accuracy in the specific problem in the parameter.
- In this research work, Runge-Kutta of fourth-order method (RK4) method can be implement for boosting purpose by applying the differential equation of

$Z' = f(s, y)$ where $y(s_0) = y_0$ as follows

in recursion formula:

Zn+1=Zn+1/6(k1+2k2+2k3+k4) h, where
1/6(k1+2k2+2k3+k4) =weighted average slope
k1=f (sn, Zn)
k2=f (sn+h2, Zn+h2k1)
k3=f (sn+h2, Zn+h2k2)
k4=hf (sn+h, Zn+hk3)
h=step size

- AdaBoost is created a short decision tree, an instance is used to analyse weight the performance of the tree on each training instance.
- By updating the weights on the training instance, models are created sequentially. Then an accuracy can be predicted after all the trees are built successfully.

Implementation: Steps of the System

1. To select the dataset, for giving an input to the algorithm for classification process.
2. To choose the classifier for calculate an entropy, gain ratio and information gain.
3. To apply an improved C4.5 algorithm process with integration of Runge Kutta method for processing the given dataset.
4. Tree generator can be generating the C4.5 decision tree for the inputted mechanism and rule set can be formed on decision tree.

- At the initial state each path can be evaluate the condition and outcome can be placed on the leaf node by the simplification of rule. The steps can be follows repeatedly when the condition comes not satisfied, it automatically terminates from the condition.
- To integrate C4.5 algorithm combines with AdaBoost and integrate Runge kutta method for tuning the parameter for reducing the variance of error when occur error on classifier.
- In training set of $(a_1, b_1), (a_2, b_2), \&(a_n, b_n)$, each a_i represents instance space A and each label b_i, contains set B= {-1, +1}. Although Hybrid AdaBoost using Runge- Kutta method can assign a weak learning parameter repeatedly in a series of round j= 1,2,3…. j, the weigh on the training example i on round s is denoted as $D_j(i)$.
- The same weight to be set at the initial point ($D_1(i)=1/N$, i=1,2,...N). Then the weight can be misclassified and concentrate on the hard examples in the training set. The integrated algorithm is given below,

Step1: Assign N example

$(a_1, b_1), (a_2, b_2), \&(a_n, b_n); a_i \in A, b_i \in \{-1, +1\}$

Step 2: Initialize the weight of

$W_1(i)$=Zn+1/6(k1+2k2+2k3+k4) h where n, i=1, …. N

Step 3: For j=1, 2, …J

Step 4: Using distribution D_j to train the weak leaner

Step5: Get weak hypothesis h_j:a->R with its error:

$\varepsilon_j = \sum_{i=h_j(a_i)^1 b_i} D_j(i)$

Step 6: Choose $\varepsilon_j = R$

Step 7: Update $D_{j+1}(i)=\frac{D_j(i)\exp(-\alpha_j b_j J_j(a_j))}{z_j}$

Step 8: The output of final hypothesis to apply Runge kutta method for tuning the step size parameter and adjust the weight of the parameter. Then it changes improving and solve the overfitting and finally generalizing the result.

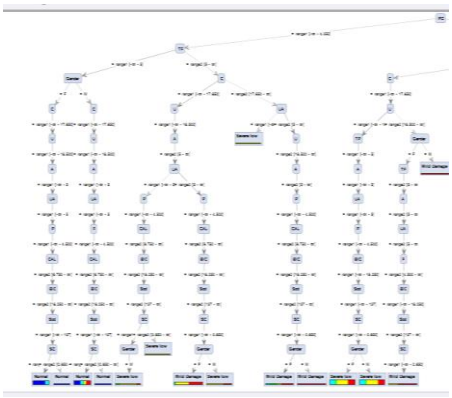

Figure 4 Tree view of Hybrid Method – Adaboost with C4.5 algorithm Using Runge Kutta Method

The main idea of Hybrid Method - adaptive boosting with C4.5 using Runge-Kutta method, weight of the parameter can be tunned well on each boosting iteration during the classifier construction. Thus, as a cost of this focus, training data correctly classified in a previous iteration will be slightly more likely to be mis-classified in the current iteration, the tree view can be shown in figure 4.

IV. EXPERIMENTAL ANALYSIS

The accuracy of 10-fold- cross validation in C4.5 is 61.2% and finally mean absolute error, Relative absolute error, Root relative squared error, Kappa statistic, Root mean squared error can be measured as shown in table.1.

Table.1 Experiment Analysis of Random Forest under bagging method

Algorithm	**Classification Instance Accuracy**	**Inclassified Instance Accuracy**	**Mean absolute error**	**Relative absolute error**	**Root relative-squared error**	**Kappa - statistic**	**Root - mean squared error**
Random Forest under bagging method	60.4%	39.3%	0.223	73.76 %	86.03 %	0.3832	0.3148

The accuracy of Random Forest under bagging method is 60.4% and finally mean absolute error, Relative absolute error, Root relative squared error,

Kappa statistic, Root mean squared error can be measured as shown in table 1.

Table 2 Experiment Analysis of 10-fold- cross validation in C4.5

Algorithm	Classification Instance Accuracy	In classified Instance Accuracy	Mean absolute error	Relative absolute error	Root relative squared error	Kappa statistic	Root mean squared error
10-fold-cross validation in C4.5	61.8%	39.68%	0.2021	73.97 %	87.32 %	0.3932	0.3278

The accuracy of 10-fold- cross validation in C4.5 is 61.8% and finally mean absolute error, Relative absolute error, Root relative squared error, Kappa statistic, Root mean squared error can be measured as shown in table 2.

Table. 3 Experiment Analysis of Hybrid Method- AdaBoost with C4.5 algorithm using Runge Kutta method

Algorithm	Classification Instance Accuracy	In classified Instance Accuracy	Mean absolute error	Relative absolute error	Root relative squared error	Kappa statistic	Root mean squared error
Hybrid Method-AdaBoost with C4.5 algorithm using Runge Kutta method	95.2%	8.9%	0.0125	8.04 %	30.51 %	0.947	0.0762

A Classified accuracy attained 95.2% from training dataset using Hybrid Method- AdaBoost with C4.5 algorithm using
Runge Kutta method decision Tree as shown in table 4.10. The fine rule base is generated below,

Sod < 136.25
| SC = Normal
| | CAL < 6.75
| | | UA <3.75 : Moderate Low (2/0)
| | | UA >= 3.75 : Mild Damage (1/0)
| | CAL >= 6.75 : Normal (14/0)
| SC = Low
| | Sod < 134.25
| | | Gender = M
| | | | UA < 3.55

| | | | | UA <3.25 : Kidney Failure (6/3)
| | | | | UA >= 3.25 : Kidney Failure (2/0)
| | | | UA >= 3.55
| | | | | UA <4.05 : Severe low (1/0)
| | | | | UA >= 4.05 : Severe low (5/2)

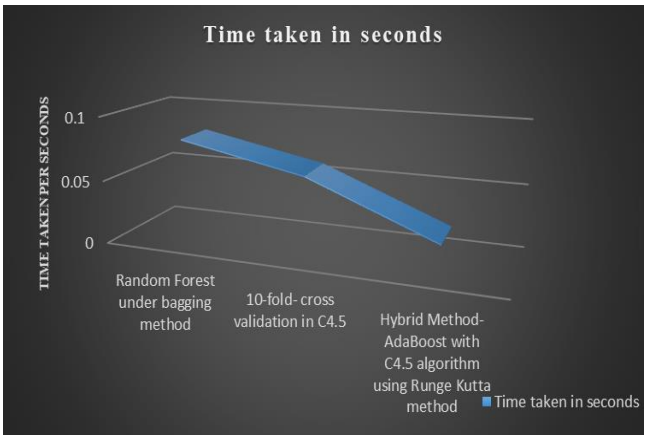

Figure 5 Time taken of each Algorithm

Figure 5 represents the time taken by each algorithm during execution of run time during classification. From this analysis, Random Forest under bagging method takes 0.08 seconds, 10-fold- cross validation in C4.5 can takes 0.06 and Hybrid Method- AdaBoost with C4.5 algorithm using Runge Kutta method can takes 0.02 seconds as shown in table 4.

Table 4 Time efficiency for Kidney Function Test

Algorithm	**Time taken in seconds**
Random Forest under bagging method	0.08
10-fold- cross validation in C4.5	0.06
Hybrid Method- AdaBoost with C4.5 algorithm using Runge Kutta method	0.02

From this analysis, Hybrid Method- AdaBoost with C4.5 algorithm using Runge Kutta method gives better result when compared with another algorithm as shown in figure 5.

VI CONCLUSION

In this work, by learning the level of fluctuation of parameter, to diagnosis the kidney function based on patient health conditions.

- Random Forest under bagging method has been implemented where bagging is integrating with random forest for avoiding the over fit with data includes in noisy. But the drawback occurring required careful tuning of different hyper-parameters.
- The validation process is taken over 10-fold- cross validation in C4.5 can be averaged to produce a single estimation and acquire less accuracy in classification which requires careful tuning of different hyper-parameters.
- When implemented Hybrid Method- AdaBoost with C4.5 algorithm using Runge Kutta method improves tuning the step size parameter and adjust the weight of the parameter. Then it changes and solve the overfitting and finally generalizing the result. Finally, Hybrid Method- AdaBoost with C4.5 algorithm using Runge Kutta method gives better result when compared with another algorithm

REFERENCE

1. Akter, S. , Shekhar, H. and Akhteruzzaman, S. (2021) Application of Biochemical Tests and Machine Learning Techniques to Diagnose and Evaluate Liver Disease. Advances in Bioscience and Biotechnology, **12**, 154-172. doi: 10.4236/abb.2021.126011.
2. Zhang, Y., Yang, D., Liu, Z. et al. An explainable supervised machine learning predictor of acute kidney injury after adult deceased donor liver transplantation. J Transl Med **19,** 321 (2021). https://doi.org/10.1186/s12967-021-02990-4.
3. Dovgan E, Gradišek A, Luštrek M, Uddin M, Nursetyo AA, Annavarajula SK, et al. (2020) Using machine learning models to predict the initiation of renal replacement therapy among chronic kidney disease patients. PLoS ONE 15(6): e0233976. https://doi.org/10.1371/journal.pone.0233976
4. Improta, G, Mazzella, V, Vecchione, D, Santini, S, and Triassi, M (2020). Fuzzy logic–based clinical decision support system for the evaluation of renal function in post-transplant patients. Journal of Evaluation in Clinical Practice. 26, 1224-1234. https://doi.org/10.1111/jep.13302
5. Rajabi, M, Sadeghizadeh, H, Mola-Amini, Z, and Ahmadyrad, N. (2019) . Hybrid adaptive neuro-fuzzy inference system for diagnosing the liver disorders. Available https://arxiv.org/abs/1910.12952
6. Anu Chaudhary, Puneet Garg, Detecting and Diagnosing a Disease by Patient Monitoring System, International Journal of Mechanical Engineering And Information Technology, Vol. 2 Issue 6 //June //Page No: 493-499, 2014.
7. Ashfaq Ahmed K, Sultan Aljahdali and Syed Naimatullah Hussain,(2013) "Comparative Prediction Performance with Support Vector Machine and Random Forest Classification Techniques", International Journal of Computer Applications Volume 69– No.11, May page no 12-16, 2013.Breiman, L . "Random Forests". Machine Learning 45 (1): 5–32, 2001.
8. N. Afhami, Prediction of Diabetic Chronic Kidney Disease Processing Using Data Mining Techniques", International Journal of Computer Science Engineering, Vol. 7 No. 02 PP- 35-40, ISSN: 2319- 7323, 2018.
9. Basma Boukenze, et.al," Performance of Data Mining Techniques to Predict in Healthcare Case Study : Chronic Kidney Failure Disease", International Journal of Database Management Systems (IJDMS) Vol.8, No.3, PP: 1 to 4, 2016
10. Breiman, L, "Bagging Predictors", Machine Learning, 24(2): 123-140, 1996.
11. L. Breiman, J. Friedman, R. Olshen and C. Stone. "Classification and Regression Trees", Wadsworth International Group, Belmont, CA, 1984.

12. Dietterich, T. G., “An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting and randomization”. Machine learning, 40: 139-157, 2001.
13. Endo, A, Shibata, T and Tanaka, H (2008) “Comparison of Seven Algorithms to Predict Breast Cancer Survival”, Biomedical Soft Computing and Human Sciences, 13(2), pp.11-16, 2008.
14. Freund, Y. Schapire, R. (1996). “Experiments with a new boosting algorithm”, In Proceedings of the Thirteenth International Conference on Machine Learning, 148-156, 1996.
15. T.F. Gonzales. Clustering to minimize the maximum inter cluster distance. Theoretical Computer Science, 38(2-3):293-306, 1985.
16. Hall, Mark, et al. "The WEKA data mining software: an update." ACM SIGKDD Explorations Newsletter 11(1), 10-18, 2009.
17. M., Hand, D. J and Steinberg, D, “Top 10 Algorithms in Data Mining”, Knowledge and Information Systems, 14 (1): 1-37, 2008.
18. D.A. Keim, 2002, “Information visualization and visual data mining”, IEEE Transactions on Visualization and Computer Graphics, 8(1):1–8, 2002.
19. Lakshmi. K.R, Nagesh. Y and VeeraKrishna. M, (2014) Performance Comparison of Three Data Mining Techniques For Predicting Kidney Dialysis Survivability, International Journal of Advances in Engineering & Technology, Mar., Vol. 7, Issue 1, pg no. 242-254, 2014.
20. D.Lavanya, K.Usha Rani, “Performance Evaluation of Decision Tree Classifiers on Medical Datasets”, International Journal of Computer Applications (0975 – 8887) Volume 26– No.4, 2011.
21. McLernon DJ, Donnan PT, Sullivan FM, et a, “ Prediction of liver disease in patients whose liver function tests have been checked in primary care: model development and validation using population-based observational cohorts”, BMJ;4:e004837. doi:10.1136/bmjopen- 2014-004837, 2014.
22. Nazmun Nahar and Ferdous Ara, “Liver Disease Prediction By Using Different Decision Tree Techniques”, International Journal of Data Mining & Knowledge Management Process,Vol.8, No.2, PP-1-9, DOI: 10.5121/ijdkp.2018.8201, 2018.
23. Nongyao Nai-arun, et.al, “Comparison of Classifiers for the Risk of Diabetes Prediction”, Procedia Computer Science 69, 132 – 142, 2015.
24. Opitz, D and Maclin, R, "Popular Ensemble Methods: An Empirical Study", 11: 169-198, 1999.
25. Oded Maimon and Lior Rokach, 2010, Data Mining and Knowledge Discovery Handbook, Second Edition, Springer, ISBN 978-0-387-09822-7 e-ISBN 978-0-387- 09823-4, DOI 10.1007/978-0-387-09823-4
26. Pugh RN, Murray-Lyon IM, Dawson JL, et. al. Transection of the oesophagus for bleeding oesophageal varices. Br J Surg.; 60:646, 1973.
27. J. Pradeep Kandhasamy, S. Balamurali, “Performance Analysis of Classifier Models to Predict Diabetes Mellitus”, Procedia Computer Science Issue 47 PP(45 – 51), doi: 10.1016/j.procs.2015.03.182, 2015.
28. Quinlan, J. R. , “Bagging, Boosting and C4.5”, AAAI/IAAI, 1: 725-730, 1996.
29. Schapire, R., “The strength of weak learnability”, Machine Learning, 5(2): 197-227. 1990.
30. Sajida perveena et. Al, “Performance Analysis of Data Mining Classification Techniques to Predict Diabetes”, Procedia Computer Science Elsevier , 82 (2016) 115 – 121.
31. K.Swapna and Prof. M.S. Prasad Babu,, “A Critical Study on Cluster Analysis Methods to Extract Liver Disease Patterns in Indian Liver Patient Data”, International Journal of

Computational Intelligence Research, Volume 13, Number 10, pp. 2379-2390, ISSN 0973-1873, 2017.
32. Tapas Ranjan Baitharua , Subhendu Kumar Panib, ”Analysis of Data Mining Techniques For Healthcare Decision Support System Using Liver Disorder Dataset”, Procedia Computer Science volume- 85, 862 – 870, doi: 10.1016/j.procs.2016.05.276, 2016.
33. Trey C, Burns DG, Saunders SJ. Treatment of hepatic coma by exchange blood transfusion. NEJM. 1966; 274:473. PMID: 5904286.
34. J.Vijayalakshmi, Kidney Failure Due to Diabetics – “Detection using Classification Algorithm in Data Mining”, International Journal of Data Mining Techniques and Application, Volume: 06, Issue: 02, , Page No.62-64 ISSN: 2278-2419, 2017.
35. Dr. S. Vijayarani, Mr.S.Dhayanand, “Data Mining Classification Algorithms For Kidney Disease Prediction”, International Journal on Cybernetics & Informatics, Vol. 4, No. 4, pp: 13 to 25 DOI: 10.5121/ijci.2015.4402, 2015.
36. Wolpert, D. (1992). “Stacked generalization”, Neural Networks, 5: 241-259.
37. Ajitha, P.Sivasangari, A.Gomathi, R.M.Indira, K."Prediction of customer plan using churn analysis for telecom industry",Recent Advances in Computer Science and Communications,Volume 13, Issue 5, 2020, Pages 926-929.
38. "Sivasangari A, Ajitha P, Rajkumar and Poonguzhali,” Emotion recognition system for autism disordered people”, Journal of Ambient Intelligence and Humanized Computing (2019)."
39. Ajitha, P., Lavanya Chowdary, J., Joshika, K., Sivasangari, A., Gomathi, R.M., "Third Vision for Women Using Deep Learning Techniques", 4th International Conference on Computer, Communication and Signal Processing, ICCCSP 2020, 2020, 9315196
40. Sivasangari, A., Gomathi, R.M., Ajitha, P., Anandhi (2020), Data fusion in smart transport using convolutional neural network”, Journal of Green Engineering, 2020, 10(10), pp. 8512–8523.
41. A Sivasangari, P Ajitha, RM Gomathi, "Light weight security scheme in wireless body area sensor network using logistic chaotic scheme", International Journal of Networking and Virtual Organisations, 22(4), PP.433-444, 2020
42. Sivasangari A, Bhowal S, Subhashini R "Secure encryption in wireless body sensor networks",Advances in Intelligent Systems and Computing, 2019, 814, pp. 679–686
43. Sindhu K, Subhashini R, Gowri S, Vimali JS, "A Women Safety Portable Hidden camera detector and jammer", Proceedings of the 3rd International Conference on Communication and Electronics Systems, ICCES 2018, 2018, pp. 1187–1189, 8724066.
44. Gowri, S., and J. Jabez. "Novel Methodology of Data Management in Ad Hoc Network Formulated Using Nanosensors for Detection of Industrial Pollutants." In International Conference on Computational Intelligence, Communications, and Business Analytics, pp. 206-216. Springer, Singapore, 2017.
45. Gowri, S. and Divya, G., 2015, February. Automation of garden tools monitored using mobile application. In International Confernce on Innovation Information in Computing Technologies (pp. 1-6). IEEE.