International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-7, Issue-3, May-2019
DOI: 10.21276/ijircst.2019.7.3.10

Chronic Kidney Disease Detection Using Micro-service Architecture

Champa M S, Rekha B S

***Abstract* — In this proposed project, the patients will perform the registration, once the patient is registered into the application then the patients are allowed to enter there scanning attributes of kidney diseases and then the user kidney disease stage level is determined by making use of C5.0 and Naïve Bayes supervised machine learning algorithm, the system provides, the user will get suggestions from the doctor at various level of kidney stage and if the user belongs to kidney stage 5 then patient will get suggestions patient will get suggestions as well as appointment request.**

***Keywords*- chronic kidney disease; C5.0; Naïve Bayes; scanning attributes**

I. INTRODUCTION

Chronic Kidney Disease (CKD) is considered as one of the major public health issue and is considered to be dangerous due to its high possibility of death. According to past data, the World Health Organization (WHO) came to the reference that South East Asia and the America has been affected by this disease with the highest annual rate i.e. around 14%, out of which 9-13% is adult population [1]. As the medical expenditure for treatment is high and also due to tight schedule of people, they fail to go for a medical checkup regularly, so this application would help to provide the users with an initial stage of medication. A person affected by CKD has a high chance of cardiovascular death than reaching end stage of renal blockage. Patients suffering from hypertension, diabetes mellitus and other metabolic symptoms are found to suffer more from chronic kidney disease. Classification is a considered to be an important data mining task and the main function of classification is to design a classification model in order to classify the dataset into specific class label.

Due to high medical cost and time restriction, it is not always possible for a patient to go and visit doctor and have proper medication. Also according to previous approaches, patients would book an appointment with the doctor clinic and patients are not classified according to the stage detected and patients who need immediate medical assistance. So, this application uses machine learning algorithm to predict the stage of chronic kidney disease and give suggestions based on the stage predicted. The training data that is present generally has missing attributes and uneven data. Hence it is required to either normalize the data in some cases or in some cases it is required to fill the missing values by taking the mean of the other values used in the data set in order to predict the stage of chronic kidney disease to give better accuracy and results. Sometimes the data set might not contain the target class and hence it is required to find the target class using data pre-processing algorithm like C5.0 algorithm and Naive Bayes algorithm is used to predict the CKD. The Medical Institutes and Government Organizations can have some statistic generated for a specific demography of people like age or gender based on chronic kidney disease affected. Also the whole application is built using micro-service architecture in order to decentralize the application.

Micro-service architecture is a software technique to structure the application as loosely coupled services. Micro-service architecture is an architectural style that designs the application in form of services, so that it is highly maintainable and testable.

II. RELATED WORKS

The technical review articles on Chronic Kidney Disease

- Kubra Eroglu et.al 2016 [1] suggested to compare the performance of six different machine learning classifier for detection of chronic kidney disease. K- nearest neighbor, Support Vector Machine, Naïve Bayes, J48, Decision tables, ensemble algorithms namely bagging, AdaBoost, random subspace. The result was determined based on three specific parameters namely Kappa, area under the ROC curve, and accuracy. The result obtained focused that J48 algorithm when used with bagging and random subspace provided an accuracy of 100 percent while classification.
- Veenita Kunwar et.al 2016 [2] compared three classification technique to predict chronic kidney disease. The classification techniques, used for experimental purpose was Artificial Neural Network (ANN) and Naïve Bayes. After successful conduction of the experiment it was concluded that Naïve Bayes gave better accuracy than Artificial Neural Network. Rapid Miner tool was used to conduct the experiment. One of the main reason behind conducting such experiments is to discover knowledge and

Manuscript received May 20, 2019

Champa M S, Information Science and Engineering, RV College of Engineering, Bengaluru, INDIA, +916360920131, (email: champams123@gamil.com)

Prof. Rekha B S, Information Science and Engineering, RV College of Engineering, Bengaluru, INDIA, +919945164208

find unique patterns in them while analyzing the huge un-mined data.

- K.R Lakshmi et.al 2014 [3] proposed that based on data collected from different diagnostic site for chronic kidney prediction, it is advisable to use Artificial Neural Network rather than Logistic Regression for better results in accuracy and performance. It was also concluded depending on the attributes and dataset size, suitable machine learning algorithm should be chosen.
- Naganna Chetty et.al 2014 [4] quoted that chronic kidney disease has become a common disease in everyday life and is measured in terms of glomerular filtration rate. Researchers in field of medicine is working to find best model that can be used for classification of CKD, so that it gets detected in initial stage of medication. There are several machine learning algorithms that can be used for classification. The main objective of this paper is to use wrapper subset evaluation to select the best suitable attribute for the classification model to classify CKD and non-CKD patients.
- Zeinab Sedighi et.al 2015 [5] applied classification algorithm on the UCI dataset to classify CKD and non-CKD patients. The dataset contained 400 instances out of which some instances had missing attribute or missing class, so it was important to fill the missing values in the attribute column. One of the methods could have been to drop the instance with missing value but it would result in less accuracy, so it was found KNN uses the best search method to find similar instance in the dataset to substitute the missing value.
- Ani R et.al 2016 [6] proposed that there are several stages in chronic kidney disease which will ultimately cause severe loss of kidney functions and result in kidney failure. It is fatal sometimes if it is detected in last stage so it is advisable to be detected in early stage in order to save human lives. The main objective of this research paper is to provide a clinical system that supports different machine learning algorithm. The following classification algorithm was used namely Naïve Bayes, neural network based back propagation, LDA classifier, tree based decision tree, lazy learner K- based nearest neighbor, and random subspace classification algorithm. On performing analysis, the accuracy for following algorithms was found to be 78%, 81.5%, 90%, 93%, 76%, and 94% on the UCI dataset. It was also concluded that random subspace with 10- fold cross validation gave better accuracy.
- Merve Dogruyol Basar et.al 2016 [7] proposed in this paper that AdaBoost ensemble algorithm will give better performance than individual classification algorithm performed on the classification model. The metrics on which the classification model is evaluated is accuracy, performance and area under curve, and also the accuracy is evaluated using root mean square error method.
- S. Vijayarani et.al 2015 [8] suggested a comparison between two algorithms namely support vector machine and artificial neural network based on accuracy and performance during runtime. After the experimental analysis, it was observed that depending on the dataset attributes, the selection of algorithm depends.
- S. Ramya et.al 2016 [9] performed four machine learning algorithm to give a proper overview about the accuracy and performance comparison. It choose decision tree, Naïve Bayes, Support Vector Machine and KNN (K Nearest Neighbor) for classification model. Experimental analysis displayed that KNN and SVM gave better accuracy for that particular attribute types in the dataset.
- Al-Hyari et.al 2016 [10] proposed a clinical system for predicting chronic kidney disease, which implemented classification models based on data mining algorithms namely Naïve Bayes, Artificial Neural Network and Decision Tree. The dataset was collected from some medical institute with 102 instances and was used for this experimental analysis. The result obtained was more accurate for decision tree.

III. EXISTING SYSTEM

In today world, as per the statistics of Information and Communication Technology forums for health more people have huge health issues especially in the corporate sector. More ever the people don't have much time due to the project deadlines and other issues. Currently the patients either book appointment or go to doctor and few stand in line at clinics. There are no specific criteria for the patients who critically suffer or high level of disease to have higher priority as compared to people who just want random health check.

A. Problems with Existing System

- The approach requires a user to fix the appointment by calling an admin of the hospital.
- The approach does not provide the priorities of which user is the best user suited for early treatment and prioritize the appointment.
- The previous approach cannot perform classification beyond two set of attributes.
- The previous approach requires more number of iterations for obtaining the best accuracy.
- The previous approach used was only Naïve Bayes algorithm.

IV. PROPOSED SYSTEM

In the proposed approach first preprocessing is performed by using C5.0 algorithm, to clean the unwanted data, following this process the best attributes are chosen based on gain computation. Classification of data set into chronic kidney disease into various stages in predicted by using Naïve Bayes Classifier. If the data set has the missing class then it is predicted using C5.0 Algorithm.

International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-7, Issue-3, May-2019
DOI: 10.21276/ijircst.2019.7.3.10

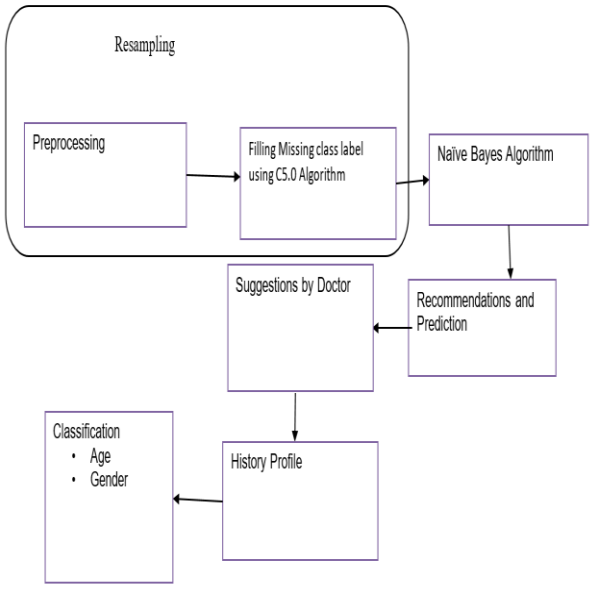

Fig 1: Methodology of the Application

A. Resampling

Obtain the data sets from UCM Library. This library is maintained by Indian medical association. The data set is a historic data of patients who had kidney disease. The data set will have various attributes like age, red blood cell count, hypertension, urea, potassium, sugar, appetite, albumin etc. Each row represents data for a patient with columns as attributes causing Kidney disease and one attribute among those columns represent the class (level of Kidney chronic disease). Among those rows few rows will have no class labels and few rows will have no class labels and few rows will have missing attributes and few rows will not have been normalized. Sampling process is performed in order to balance these scenarios and also cross validation is performed in order fill the class.

B. Preprocessing

If there are any missing values then the missing values are filled based on the mean of other values, if there are any missing values and class then data is discarded and if there are any missing class then cross validation is performed.

C. Cross Validation

This is a process in which if the class attribute is missing in the training data then we perform computations such as information gain, gain and entropy are performed and then form a decision tree. Once the decision tree is formed then the class is determined from the decision tree.

D. C5.0 and Bayes Classifier Algorithm

If the data set has the missing class then it is predicted using C5.0 Algorithm. Naive Bayes is used to construct classification model to assign class label to a particular instance of the dataset, and it predict the Kidney Stage Disease.

E. Recommendations and Prediction

Naïve Bayes Classifier with Resampling is used in order to classify the disease of the user and generate suggestions for each level of disease and appointment if the disease level is the highest

F. Suggestions by Doctor

This module is responsible for creating the suggestions for the end user. The doctor will select each kidney stage like 3, 4 or 5 and then provide the treatment plan for each of the kidney stages.

G. History Profile

Each time whenever the user takes the test the scanning data attributes are entered, then after predicting the kidney disease stage and it maintain the class. Like this n number times when the user takes the test then a graph is generated with the number of test taken as x axis as number of people and stage of the kidney disease as the y axis as the kidney disease stage. From the history profile one can come to know about the improvement in the Kidney stage disease.

H. Classification by Age and Gender

During registration the user will select the gender and also provides the age. Like this there will be many users who will be registered in the system. From the set of users the male and female users are found out and then a graph is generating based on how many males/females have Kidney stage 3 or Kidney stage 4 and Kidney stage 5. The age group is also divided into a set of multiple age groups and then the graphs are generate for age group 1 and level of kidney stage, age group 3 and level of kidney stage and finally age group 4 and level of kidney stage

V. RESULTS AND ANALYSIS

The application will create a login page by running the spring boot applications (Micro-service), need to open the server portal address and register or if you are already registered should login (figure 2). If you are a new user then register with the basic requirement (figure 3). Then login as the user and add the attributes according to the kidney test, in the application it will predict that you are in which stage you are. In the application if you go to the predict result then the patient will get the disease name and suggestion from doctor (figure 5). The application not only classify the stages it will also do classification based on the age and gender (figure 6,7). At last the accuracy is achieved 98% which have proposed using C5.0 and naïve based algorithm v/s only naïve algorithm. The Chronic kidney disease detection is more accurate by using both C5.0 and Naïve algorithm (Fig 7). Probability that the No Chronic Disease using attributes is computed using the formula:

$$P_{attribute} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(\frac{(\mu-T)^2}{2*\sigma^2}\right)}$$

Where,

$\sigma = s\tan dard\ \ deviation$

$\mu = mean$

$T = current\ \ value\ of\ \ attribute$

If the computed given data belongs to Chronic Disease is given by

$$Pchronic = Pchronic + (Pchronic + PNonChronic)$$

If the computed given data belongs to No Chronic Disease is given by

$$PNonChronic = PNonChronic + (PChronic + PNonChronic)$$

And in the time comparison the application is much faster (Fig 8), then the patient can view the suggestion from doctor if it in the stage 5 then the patient can book an appointment to doctor from the application no need to go for hospital.

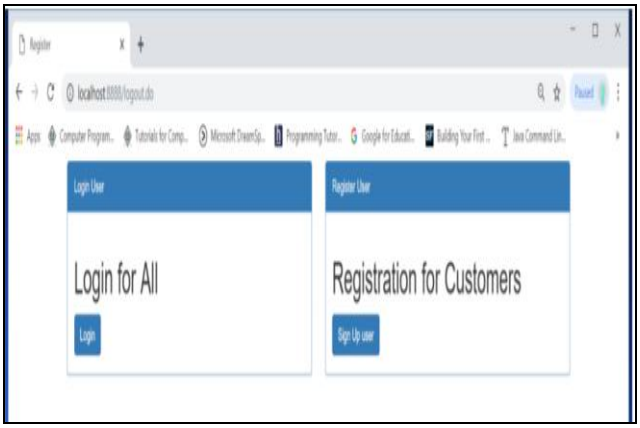

Fig 2: Login page

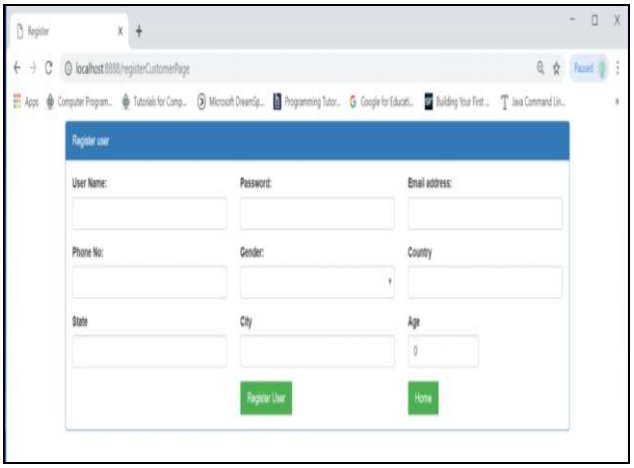

Fig 3: Registration page

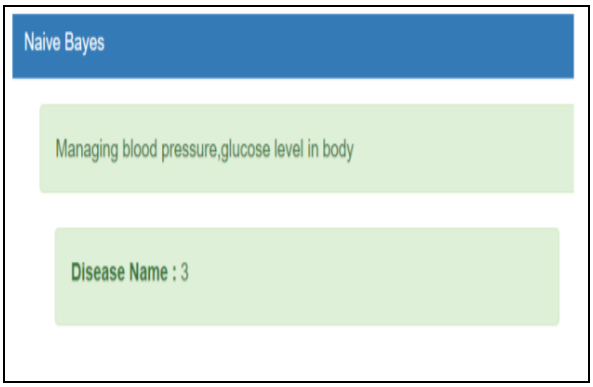

Fig 4: Prediction result with suggestion from doctor

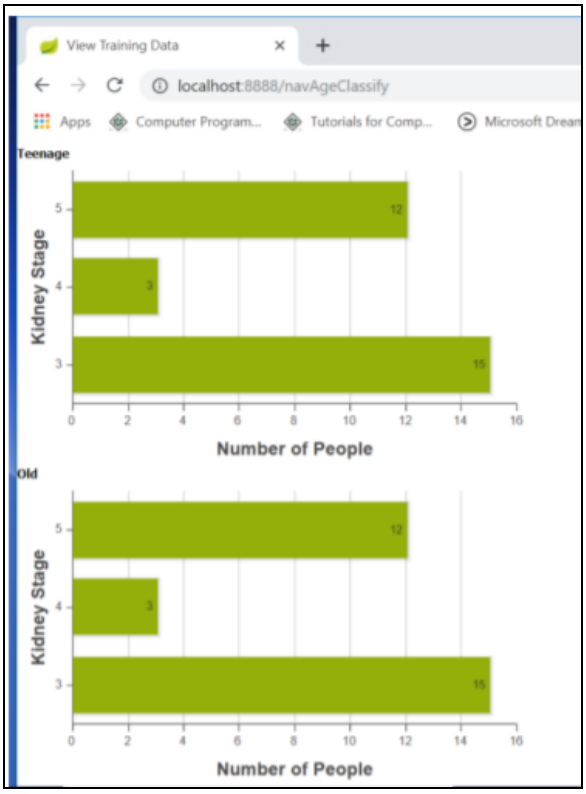

Fig 5: Classification based on age

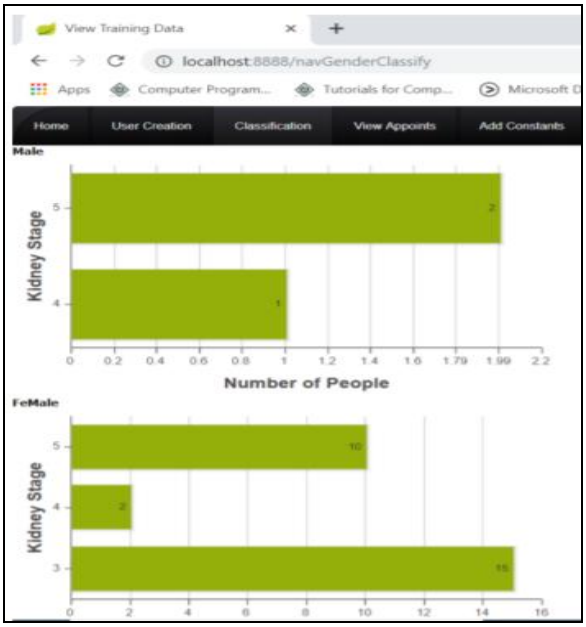

Fig 6: Classification based on gender

International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-7, Issue-3, May-2019
DOI: 10.21276/ijircst.2019.7.3.10

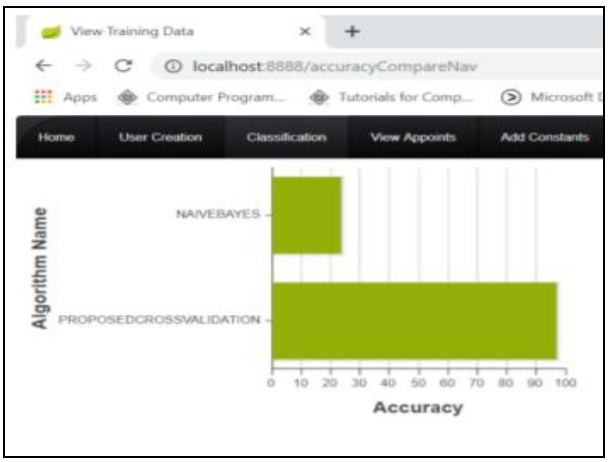

Fig 7: Accuracy comparison

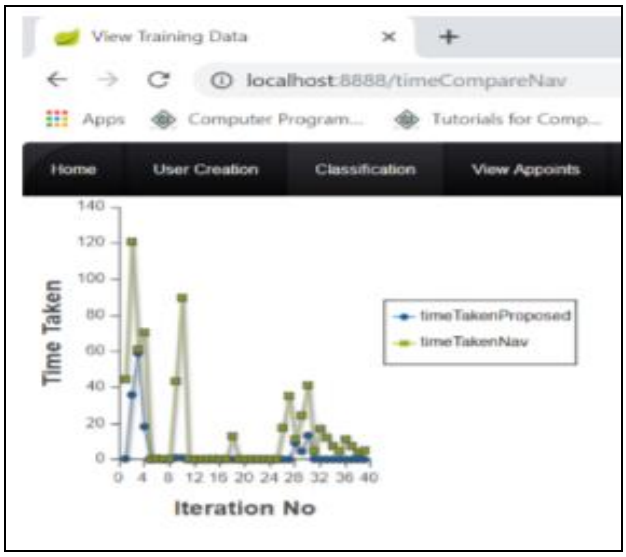

Fig 8: Time comparison

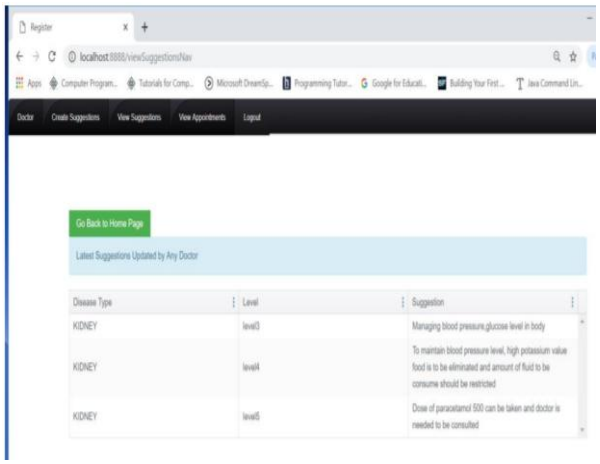

Fig 9: View of Suggestions

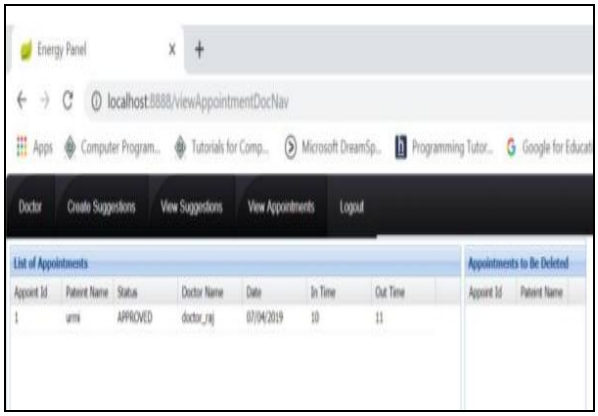

Fig 10: Appointment details of the Doctor

VI. Conclusion

Data mining gives a huge impact on analysis when implemented using proper tools and appropriate techniques. The project was developed as a real time application where the attributes for prediction was taken as input parameter and that data was used to predict the class label for the person suffering from chronic kidney disease or not and initial medication to the patient was provided and it also helped for early stage detection. Another module was added where the doctor appointment cycle was managed with suggestions regarding the stage predicted. The accuracy and performance of algorithm used can be enhanced using feature selection.

References

[1] Kubra Eroglu, Tugba Palabas," The Impact on the Classification Performance of the Combined Use of Different Classification Methods and Different Ensemble Algorithms in Chronic Kidney Disease Detection", Journal of Applied Engineering, vol. 53, no. 2, Feb., pp. 1144-1149, 2002.

[2] Veenita Kunwar 1 Khushboo Chandel2 A. Sai Sabitha3 Abhay Bansal4" Chronic Kidney Disease Analysis Using Classification Technique", in Proc. of the 2018 1st Int. Conf. on Information Technology, IT 2018, 19-21 May 2018, Gdansk, Poland[Online].

[3] K.R.Lakshmi1, Y.Nagesh2 and M.VeeraKrishna3, "Performance Comparison of Three Data Mining Techniques for Predicting Kidney Dialysis Survivability", International Journal of Advances in Engineering & Technology, Mar. 2014, Vol. 7, Issue 1, pp. 242-252

[4] Naganna Chetty, Kunwar Singh Vaisla, Sithu D Sudarsan,"Role of Attributes Selection in Classification of Chronic Kidney Disease Patients", in Proc. of the 2015 1st Int. Conf. on Applied Science, 2015, 21-25 june 2015, Russia [Online].

[5] Zeinab Sedighi, Hossein Ebrahimpour-Komleh, Seyed jalaleddin Mousavirad, "Featue selection effects on Kidney disease analysis," in Proc. of the 2016 7th Int.

Conf. on Formal Engineering Method, 2016, 20-25 May 2016, Washington, DC[Online].

[6] Ani R1, Greeshma Sasi1, Resmi Sankar U1, O.S Deepa, "Decision Support system for diagnosis and prediction of Chronic Renal Failure using Random Subspace Classification", 2016 Intl. Conference on Advances in Computing, Communications and Informatics (ICACCI), Sept. 21-24, 2016.

[7]]Miguel A. Estudillo-Valderrama, Alejandro TalaminosBarroso, Laura M. Roa," A Distributed Approach to Alarm Management in Chronic Kidney Disease", IEEE Journal of Biomedical and Health Informatics, November 2014,vol 18,no 16.

[8] Merve Doğruyol Başar1, Pelin Sarı1 , Niyazi Kılıç2 , Aydın Akan, "Detection of Chronic Kidney Disease by Using Adaboost Ensemble Learning Approach", in Proc. of the 2015 1st Int. Conf. on Data Mining, 2015, 17-20 September 2015, Singapore [Online].

[9] Anu Chaudhary, Puneet Garg,(2014) "Detecting and Diagnosing a Disease by Patient Monitoring System", International Journal of Mechanical Engineering And Information Technology, Vol. 2 ,Issue 6 ,June 2017,Page No: 493-499.

[10] 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.

ABOUT THE AUTHORS

Champa M S, M.Tech Scholar in department of Information Science and Engineering, RV College of Engineering, Bengaluru.

Prof. Rekha B S, Assistant Professor, Department of Information Science and Engineering, RV College of Engineering, Bengaluru.

