

**DIGITAL ASSIGNMENT-2**

**CHRONIC KIDNEY DISEASE PREDICTION USING**

**ENSEMBLE TECHNIQUE**

**Under the guidance of**

**Dr. Agilandeeswari L**

**School of Information Technology and Engineering**

**For the course**

**ITE1015 – Soft Computing**

**by**

**B K ANIRUDH – 19BIT0348**

**Lanka Jaswanth – 19BIT0061**

**Sompalli Yeswin – 19BIT0134**

**ABSTRACT**

As of now, there are numerous individuals on the planet experiencing persistent kidney infections around the world. Because of the a few danger factors like food, climate and expectations for everyday comforts numerous individuals get infections out of nowhere without comprehension of their condition. Diagnosing of persistent kidney infections is for the most part obtrusive, expensive, tedious and frequently hazardous. That is the reason numerous patients arrive at late phases of it without treatment, particularly in those nations where the assets are restricted. Hence, the early recognition procedure of the illness remains significant, especially in non-industrial nations, where the sicknesses are for the most part analyzed in late stages. Finding an answer for previously mentioned issues and braving from impediments turned into a solid intention to lead this study.

**KEYWORDS**

Data Processing **.** Data Cleaning **.** Feature Extraction **.** Ensemble Technique **.** Model Training **.** Chronic Kidney Disease **.** Classification

**INTRODUCTION**

Technological development, including machine learning, has a huge impact on health through an effective analysis of various chronic diseases for more accurate diagnosis and successful treatment. Kidney disease is a major chronic disease associated with aging, hypertension, and diabetes, affecting people 60 and over.

Its major cause is the malfunctioning of the kidney in disposing toxins from the blood. This study analyses chronic kidney disease using machine learning techniques based on a chronic kidney disease (CKD). Chronic Kidney Disease (CKD) or chronic renal disease has become a major issue with a steady growth rate. A person can only survive without kidneys for an average time of 18 days, which makes a huge demand for a kidney transplant and Dialysis. It is important to have effective methods for early prediction of CKD. Machine learning methods are effective in CKD prediction. This work proposes a workflow to predict CKD status based on clinical data, incorporating data prepossessing, a missing value handling method with collaborative filtering and attributes selection. Out of the 11 machine learning methods considered, the extra tree classifier and random forest classifier are shown to result in the highest accuracy and minimal bias to the attributes.

In the paper [24} the authors has used ensemble technique where they have went with boosting approach and finally has predicted the chronic kidney disease with an good accuracy of 98% since the dataset is bit smaller comparitively so we believe that we can improve the accuracy more so we went with an ensemble technique of bagging where we considering the main classifiers namely SVM, Decision Tree, Random Forest, Logistic Regression and we are using all these models as estimators and going with voter classification to have best accuracy and in the paper [37] authors used gradient boosting ensembling technique where they used decision trees for the classification of chronic kidney disease.

**OBJECTIVE**

Chronic Kidney Disease (CKD) or chronic renal disease has become a major issue

with a steady growth rate. A person can only survive without kidneys for an

average time of 18 days, which makes a huge demand for a kidney transplant and

Dialysis. This study analyses chronic kidney disease using machine learning

techniques based on a chronic kidney disease (CKD).

It is important to have effective methods for early prediction of CKD. Machine

learning methods are effective in CKD prediction. The main objective is to predict

CKD in ealrlier stages. So we are implementing ensemble model considering the output of multiple models for better optimized prediction and to get best accuracy and the other metrics.

**MOTIVATION**

As of now, there are numerous individuals on the planet experiencing persistent

kidney infections around the world. Because of the a few danger factors like food,

climate and expectations for everyday comforts numerous individuals get

infections out of nowhere without comprehension of their condition. Diagnosing of

persistent kidney infections is for the most part obtrusive, expensive, tedious

and frequently hazardous. That is the reason numerous patients arrive at late phases

of it without treatment, particularly in those nations where the assets are restricted.

Hence, the early recognition procedure of the illness remains significant, especially

in non-industrial nations, where the sicknesses are for the most part analysed in

late stages. Finding an answer for previously mentioned issues and braving from

impediments turned into a solid intention to lead this study.

**LITERATURE SURVEY**

| **Authors and year** | **Methodologys** | **Advantages** | **Issues** | **Metrics** |
| --- | --- | --- | --- | --- |
| Abdulhamit Subas, Emina Alickovic, and Jasmin Kevric - 2017 | Chronic kidney disease data set, Artificial neural network(ANN), Support Vector Machine(SVM)  , Decision Tree, Random Frost | Having a clear core probability modeling | Probabilistc approaches report high performanes just in case when the primary  assumptions are accurate | Precision(),Re call(R),F- measure |
| Fuzhe Ma,Tao Sun,Lingyun Liu,Hongyu Jing -2020 | Algorithms:Base learning algorithm and back propagation algorithm, Support vector machine(SVM) | Using ultrasound imaging has pros of radiation free,low cost and convenience | Ultrasound for obese patients have wrong results and ultrasonics rely primarily on the skill of technician in accuracy | SBD-PCN,ANFIS,PDA-ADMI,PEFS,ANN-SVM,HMANN |
| Ahmed Abdelaziz, Ahmed S. Salama, A. M. Riad and Alia N. Mahmoud – 2018 | Decision tree,SVM,Naive Bayes,Fuzzy logic | Hybrid inteligent model that serves on cloud computing anywhere,anytime | Mysterious behaviour of the network | Accuracy,Precision,Recall, F1 Score and MSE values |
| Abdullah Al Imran, Md Nur Amin, Fatema Tuj Johora - 2018 | Missing value Imputation ,Logistic regression, Feedforward Neural network, wide and deep learning | Advantage of memorization and generalization | Class imbalance problem | Precision, Recall, F1 Score AUC Score |
| Pankaj Chittora, Sandeep Chaurasia and Prasun Chakrabarti - 2021 | ANN, C5.0 decision tree, Logistic regression, linear support vector machine,CHAID, K-nearest neighbors and random tree | Docter can detect the disease on time and highest accuracy of 98.86% | Non accurate results for bigger datasets | Accuracy, Precision, Recall, F-Measure, ROC and AUC, GINI Coefficient |
| Tahsin M. Rahman, Saima Siddiqua, Siam E.Rabby, Nahid Hasan - 2019 | MATLAB based application – Classification learner app, LSVM | Accuracy level of 97.6% | Loses accuracy for larege datasets | QT interval and RR interval |
| J Qin, Lin Chen, yuhua Liu and Bin chen - 2019 | Regression-based model:LOG, Tree based model :RF,  SVM,KNN,Forward neural networks | 99.75% diagnosis accuracy and 99.83% average accuracy | KNN imputation for missing values may not always give correct results | Accuracy, Recall, Specificity, Precision and F1 score |
| S.Revathy, B.Bharathi,P.Jayanthi and M.Ramesh - 2019 | Decision tree,Support vector machine, Random Forest | The random forest algorithm classified all positive and negative samples correctly | No imputation tech is used to fill out missing values if present | Accuracy |
| Sathiya Priya S and Suresh Kumar M - 2018 | Decision tree, Naive Bayes and dataset | High accuracy | No special classification techniques | Accuracy,Sensitivity and Specificity |
| Imesh Udara Ekanayake, Damayanthi Herath - 2020 | MCAR test, Logistic regression, l-Nearest Neighbors(KNN),SVC with linear kernal,Random forest,Classical neural network | Optimized by hyperparameter tuning from a genetic algorithm and grid search for training dataset | KNN imputation for missing values may not always give correct results | Precision, Recall and F-1 Score |
| Shanila Yunus Yashfi,Md Ashikul Islam,Pritilata - 2020 | Median for missing values,10-fold cross validation,Random forest,Decision tree,ANN | Accuracy of Artificial neural network | Extracting significat features from dataset may sometimes leads to wrong accuraacy | Total no.of Instances,Accuracy,Error,Recall,F1-score,Precision |
| Merve Dogruyol, Aydin Akan - 2017 | Reptree, BFTree, J48 Decision tree, Support vector machine, Adaboost Algorithm and Random Frost | Accuracy of J48 decision tree algorithem is high and above 98% | Imputation technique is not found | Accuracy |
| Shanila Yunus Yashfi, Md Ashikul Islam, Pritilata - 2020 | RandomFrost and Artificial neural newtwork(ANN),decision tree | Optimization of algorithm | Few features in the dataset are not relavent to objective | Accuracy,Error, Recall and F1-score |
| Gopi Battineni,  Getu Gamo Sagaro,  Nalini Chinatalapudi,  Francesco Amenta ,  2020 | SVM models,  K-Nearest Neighbor (KNN), and Naïve Bayes (NB) models,  LR,random forest (RF) | SVM models show better accuracy in predicting exacerbations and COPD detection. | Predective negativiy may cause a little problem | Accuracy |
| N.Ch.Sriman Narayana Iyengar  2017 | neuro-fuzzy systems and clustering | Helps in identifying relationships within diseases that patients have. | No better results for bigger datasets | Accuracy,sensitivity,specificity |
| Mohamed Elhoseny  K. Shankar &  J. Uthayakumar  2019 | A Cuckoo Search trained neural network (NN-CS) method | execute the DFS(density based feature selection) and ACO algorithms consecutively | Probability distribution changes by iteration | F-score |
| Huseyin Polat  Homay Danaei Mehr &  Aydin Cetin  2017 | Support Vector Machine(SVM) | Good classification accuracy models | Risk of over fitting, Computationally intensive | Accuracy |
| Slshanmukharajeswaran -2021 | Random forest decision tree | Capability toward work with imperfect knowledge  Fault tolerance | Mysterious behaviour of the network | Accuracy |
| Ogunleye, A.; Wang, Q  2020 | Gentic algorithm(GA),SelectKbest,  XGBoost model | The proposed full model has achieved an accuracy, sensitivity and specificity of 1.000, 1.000 and 1.000, r | Time complexity is high | Accuracy |
| L.Jerlin Rubini, Dr.P.Eswaran  2015 | .Logistic Regression,Multilayer Perceptron | multilayer perceptron classifier gave good accuracy. | There are lot of mismatched values on confusion matrix | Accuracy |
| K. M. Zubair Hasan and Md. Zahid Hasan  2019 | Adaptive Boosting (AdaBoost), Bootstrap Aggregating (Bagging), Extra Trees, Gradient Boosting, and Random Forest | proposed ensemble learning algorithm achieves 99% classification accuracy | Risk of over fitting | Accuracy,precision,f-score |
| Himanshu Kriplani, Bhumi Patel and Sudipta Roy  2019 | Randomforest, Support vector  machine,Gradient descent | The results Prediction of  would be more promising with increase in dataset. | Random forest has less accuracy | F-score |
| N V Ganapathi Raju , K Prasanna Lakshmi , K. Gayathri Praharshitha , Chittampalli Likhitha  2019 | Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbour, Naïve Bayes | SVM Given high accuracy with 98.5% | Random forest has the least accuracy over all the techniques | Accuracy |
| N.Ch.Sriman Narayana Iyengar  2017 | neuro-fuzzy systems and clustering | Helps in identifying relationships within diseases that patients have. | No better results for bigger datasets | Accuracy,sensitivity,specificity |
| K. M. Zubair Hasan and Md. Zahid Hasan, 2019 | Ensembled based technique | Ada boost has 99% accuracy | Gradient boosting has not performed well and has a least accuracy of 97% | Accuracy |
| Charleonnan, AFufaung, TNiyomwong, TChokchueypattanakit, W.Suwannawach, SNinchawee, N  2016 | K-nearest neighbors (KNN), support vector machine (SVM), logistic regression (LR), and decision tree classifiers. | SVM gives high accuracy of 98.2% | K-nearest nieghbours go for low values of accuracy | Accuracy,precision |
| Sedighi, ZeinabEbrahimpour-Komleh, HosseinMousavirad, Seyed Jalaleddin  2015 | AdaBoost,Naive Bayes,Genetic algorithm | Removes features that make noise or have low correlation with other characteristics | High running time | Accuracy |
| S.Gopika , Dr.M.Vanitha  2017 | K-means Algorithm,K-Medoids Algorithm ,Fuzzy clustering(FCM) | FCM given good accuracy over the results with 92% | K-means,K-mediods given accuracy less than 90% | Accuracy |
| Arulanthu, PramilaPerumal, Eswaran  2019 | decision tree (DT), SVM, ANN, linear regression (LR), KNN, NB | SVM given good accuracy with 97% | Loss of data due to undersampling method | Accuracy,F-score |
| S.Revathy, B.Bharathi,P.Jayanthi and M.Ramesh - 2019 | Decision tree,Support vector machine, Random Forest | The random forest algorithm classified all positive and negative samples correctly | No technique is used to fill out missing values if present | Accuracy |
| Ebrahime Mohammed Senan,  Mosleh Hmoud Al-Adhaileh,  Ahmed H. Alahmadi  2021 | algorithms: SVM, KNN, decision tree, and random forest to evaluate CKD dataset. | The random forest algorithm classified all positive and negative samples correctly | the SVM, KNN samples by 94.74%, 97.37%, and  respectively,SVM has no promisive results at higher datsets | Accuracy,precision,recall,F1-score |
| Marwa Almasoud,  Tomas E Ward  2019 | algorithms used are logistic regression (LR), support vector machines (SVM), random forest (RF), and gradient boosting (GB). | higher results were achieved using RF, and GB algorithm, | Remaining algorithms were not that accurate | Accuracy,precision,sensitivity |
| Zvi segal,  Dan kalifa,  Gideon koren,  2020 | gradient boosting tree algorithm (XGBoost implementation). | Negativi prediction is highly accurate | Predictive positivity is less so it may cause a problem | Accuracy,sensitivity |
| Gopi Battineni,  Getu Gamo Sagaro,  Nalini Chinatalapudi,  Francesco Amenta ,  2020 | SVM models,  K-Nearest Neighbor (KNN), and Naïve Bayes (NB) models,  LR,random forest (RF) | SVM models show better accuracy in predicting exacerbations and COPD detection. | Predective negativiy may cause a little problem | Accuracy |
| Vineeta Gulati,  Neeraj Raheja,  2021 | Support Vector Machine (SVM), Decision Tree (DT), Naïve Bayes (NB), Random Forest (RF), Recurrent Neural Network (RNF) etc | SVM and RNF shows the better accuracy in predicting CKD | factors such as prescribed drugs affecting blood pressure of the patient, we may have a high probability for a wrong prediction. | Accuracy |
| Anik saha,  Abir saha,  Tanni mitra  2020 | Random Forest, Naïve Bayes, Multilayer Perceptron, Logistic Regression and Neural Network optimized by Adam optimizer | Adam-Deep learning outperforms all the approaches by predicting accuracy of 97.34% | Using too many algorithm although the are not so accurate in finding the disease | accuracy |
| N radha,  S ramya,  2015 | Naive Bayes, Decision Tree, K-Nearest Neighbour and Support Vector Machine. | K-Nearest Neighbour algorithm gives better result with 98% accuracy | Remaining algorithms are not accurate in finding the disease | accuracy |
| Mohamed Alloghani,  Dhiya Al-Jumeily  Abir Hussain  2019 | decision tree boosted decision tree, and CN2 rule | Neural network and svm has high accuracy | Smoking and usage if drugs may change the results | Accuracy,efficiency |
| Hamid Mohamadlou,  Nicholas R. Saber,  Ritankar Das,  2018 | Gradiesnt boosted ensembles of decision trees, to train and prediction on retrospective data taken from UCI repository | The algorithm  has predicted who are having ckd but there are lot of misclassifications for who not having ckd | we cannot draw any conclusions about the impact the algorithm’s predictions will have on patient outcomes in a clinical setting | Accuracy,precison,recall,f1-score |
| S.revathy,  B.bharathi,  p.jeyanthi,  m. ramesh  2019 | KNN (K-Nearest Neighbor) and SVM (Support Vector Machine) random forest | Random Forest Classifier model better predicts CKD | Decision trees and Support Vector machines. Are little less accurate | accuracy |
| Jing xiao,  Xiulin xu,  Xinhui feng,  2019 | logistic regression, Elastic Net, lasso regression, ridge regression, support vector machine, random forest, XGBoost, neural network and k-nearest neighbor | Elastic Net, lasso regression, ridge regression and logistic regression showed the highest overall predictive power | Svm is not that usefull in this methodology in predicting the disease | AU-ROC, sensitivity (recall), specificity, accuracy, log-loss and precision |
| Geetanjali Bhola  Aman Garg  Manisha Kumari  2020 | logistic regression and random forest models are trained using real-world data collected from the USA in 2017 by BRFSS | Both the algorithms give highly accuarate prediction | selecting the appropriate model for each disease through comparative study | accuracy |
| Theyazn H.H Aldhyani,  Ali Saleh Alshebami,  Mohammed Y. Alzahrani  2020 | decision tree (DT), SVM, ANN, linear regression (LR), KNN, NB | Accurate prediction of disease | biggest challenges that we have faced within the implementation of the proposed system is the ambiguity embedded in the variable of the standard dataset | Accuracy,specificity,sensitivity,f1-score,precision |
| Abdullah Al Imran,Md Nur Amin,Fatema Tuj Johora | Logistic Regression, Feedforward Neural Network and Wide & Deep Learning | Good accuracy with neural network and accuracy of 99% | Logistic regression has not performed well and its accuarcy us only 97% | Accuracy |

**ISSUES IN EXISTING SYSTEM**

1.There is huge problem of overfitting where model cannot able to generalize the attributes.

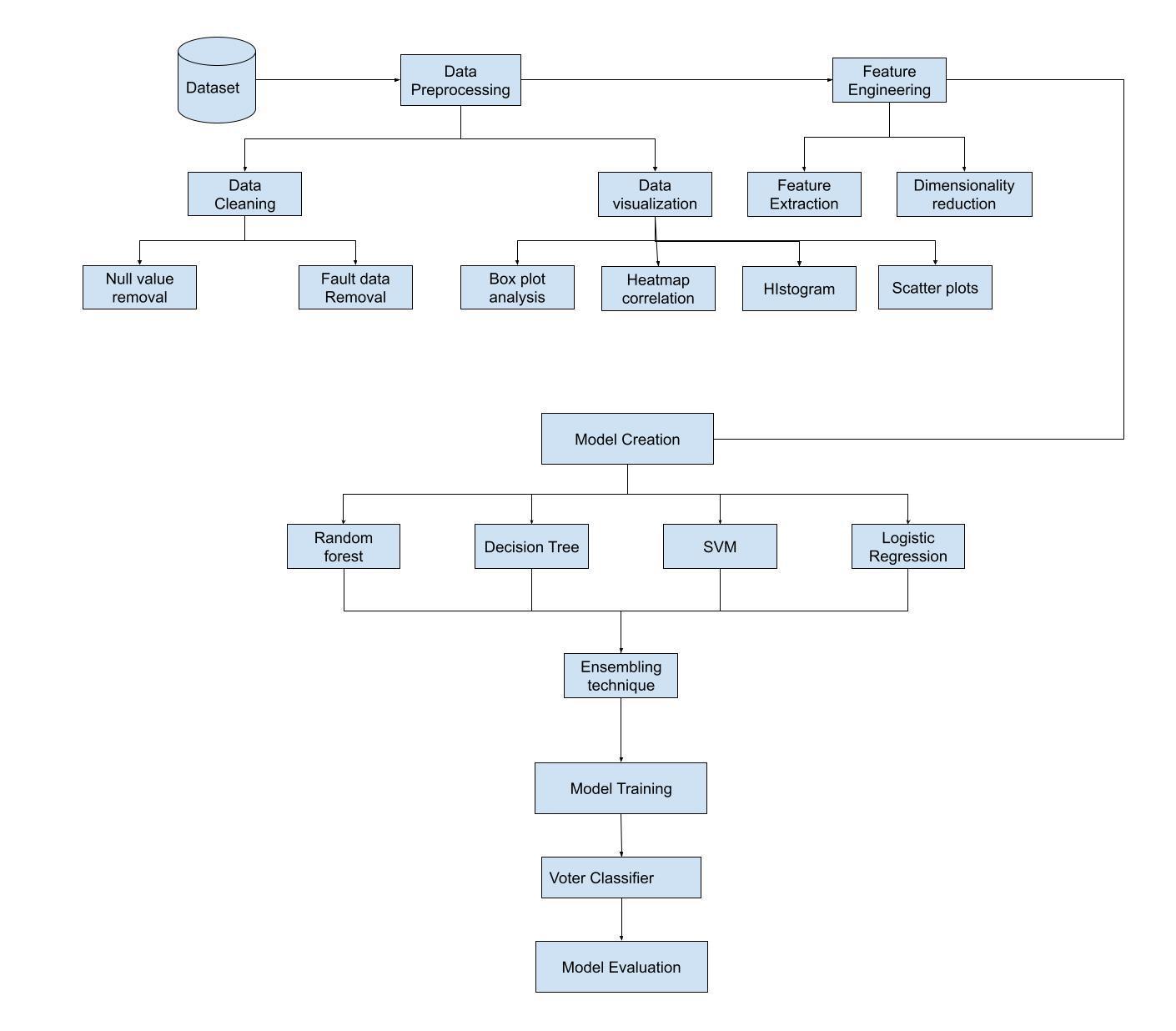
2. No promised results for larger datasets.

**ISSUES CONSIDERED IN OUR SYSTEM**

1. Solving the overfitting problems

2. Achieving maximum accuracy

**PROPOSED ARCHITECTURE**



**MODULE DESCRIPTION**

**DATA PREPROCESSING**

Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way. So, for this, we use data preprocessing task. It has steps like Getting the dataset, importing dataset, Finding the missing data, Encoding the categorical data, Splitting the dataset between training and testing.j**Data reduction**

We will use correlation between all the attributes which have high correlation will be dropped.

**Data Transmission**

We will change the data formats according to the required models.

**Feature Engineering**

Feature engineering refers to a process of selecting and transforming variables when creating a predictive model using machine learning or statistical modeling (such as deep learning, decision trees, or regression). The process involves a combination of data analysis, applying rules of thumb, and judgement. It is sometimes referred to as pre-processing, although that term can have a more general meaning. In this study we will extract the features by dimensionality reduction process.

**MODEL CREATION**

**Ensemble**  **Technique**

In this approach we considered multiple models namely decision tree,LogisticRegression,SVM,K-nearest nieghbours and RandomForest and made as estimaters now all these models will predict the output for the data which is present after training.

**Voting Classifier**

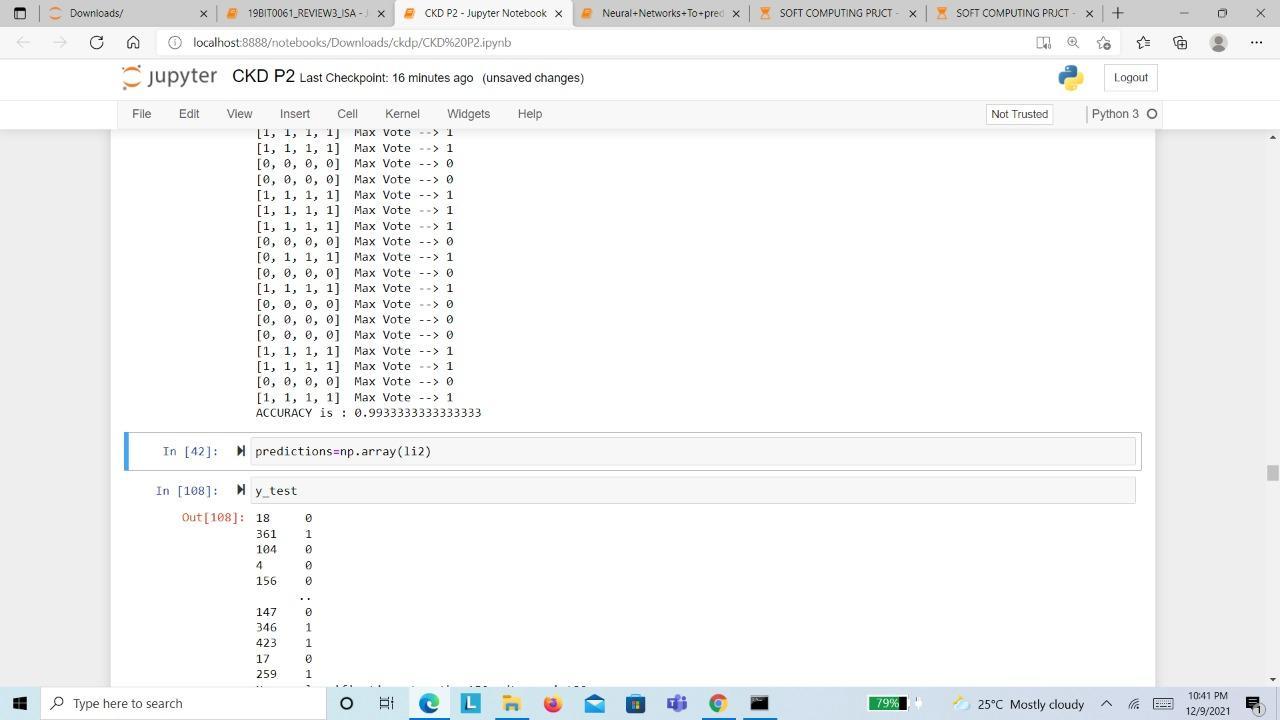
The max voting method is generally used for classification problems. In this technique, multiple models are used to make predictions for each data point. The predictions by each model are considered as a ‘vote’. The predictions which we get from the majority of the models are used as the final prediction.

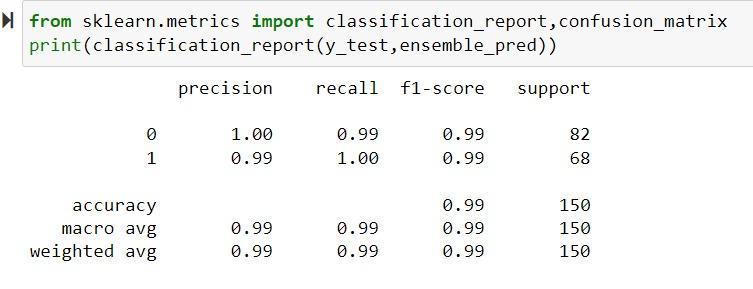
**DATASET DESCRIPTION**

The Dataset we are using is from UCI machine learning repository.The Dataset contains 24 health related attributes taken in 2-month period of 400 patients in which 11 numeric and 14 nominal attributes in which it consists of class label named ‘Class’ which classifies patients having disease and not present

<https://archive.ics.uci.edu/ml/datasets/chronic_kidney_disease>

**RESULTS**





*Accuracy : 99.33%*

**COMPARATIVE STUjDY**

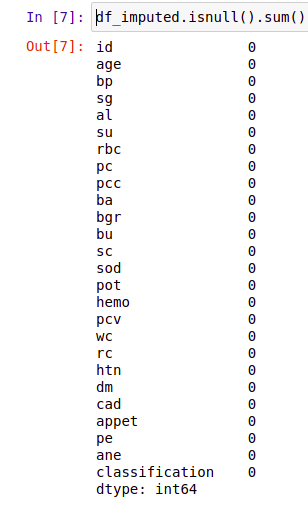
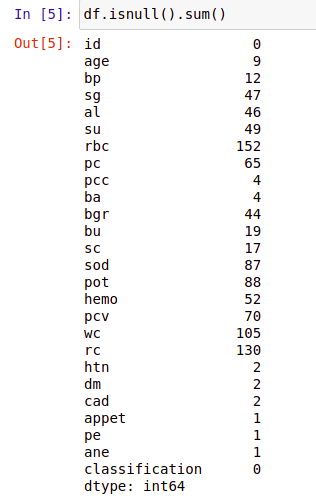
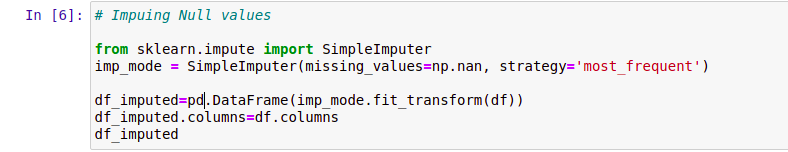
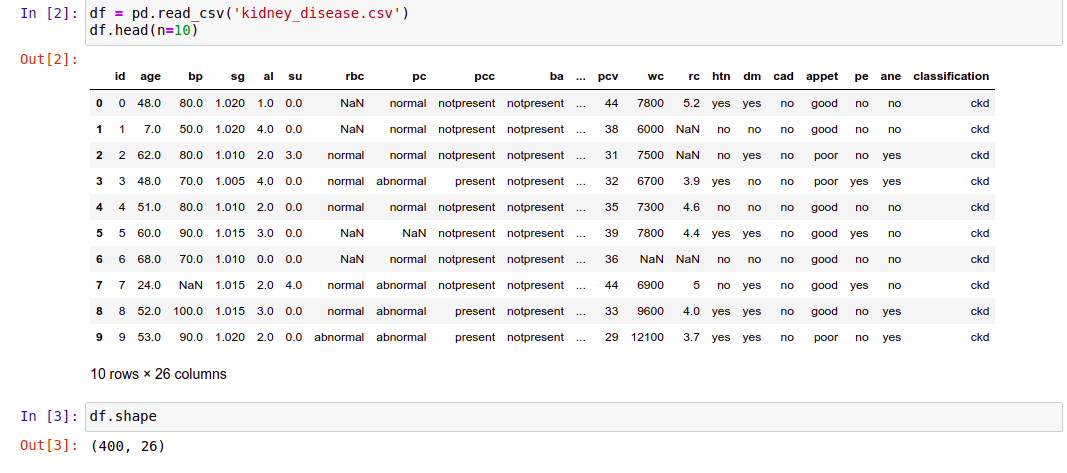
The paper [37] has used Ensemble technique which has an accuracy of 97% since in our proposed methodology we are using voting based ensemble technique we are taking this paper for analysis and have a good comparison over the classification and also we are taking paper [24] since this paper has done classification based on ensemble techniques so for the comparative results our proposed methodology also contains ensemble technique we have taken this paper.

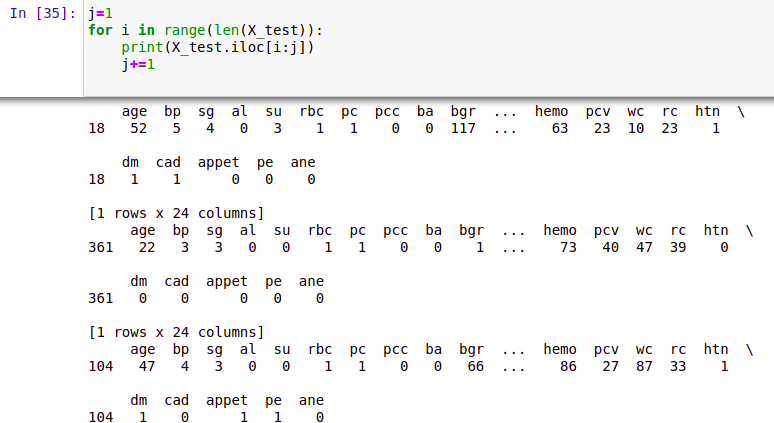
| Parameters | [37] | [24] | Ensemble Technique |
| --- | --- | --- | --- |
| Dataset | UCI repository | UCI repository | UCI repository |
| Accuracy | 0.97 | 0.99 | 99.33 |
| F1-score | 0.92 | 0.99 | 0.99 |
| precision | 0.97 | 0.99 | 0.99 |
| recall | 0.98 | 0.99 | 0.99 |

**INFERENCE**

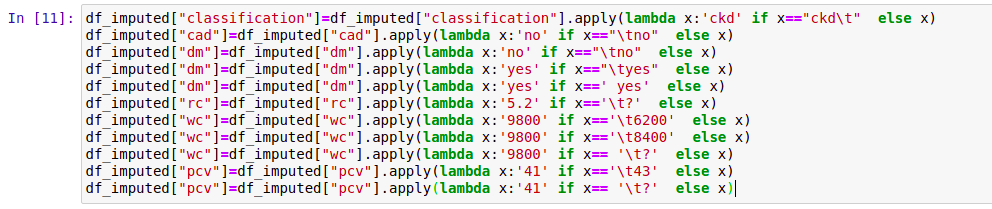
From the above table we can see the proposed methodology of voting based ensemble technique has improved accuracy of 99.33% which is higher than the accuracy in the paper[24] and paper[37] this is due to we have taken multiple classification models and predicted output with individual algorithm and we considered the predicted output as which has the highest votes and also our model has given good F1-score,precision and recall.

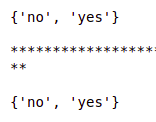
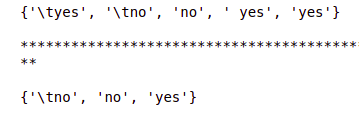
**WORKSPACE SNAPSHOTS**

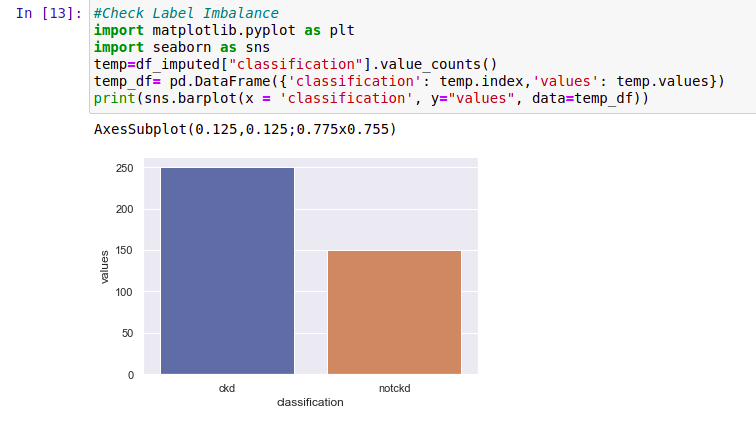


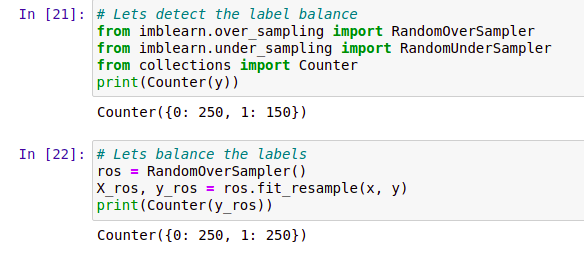


8

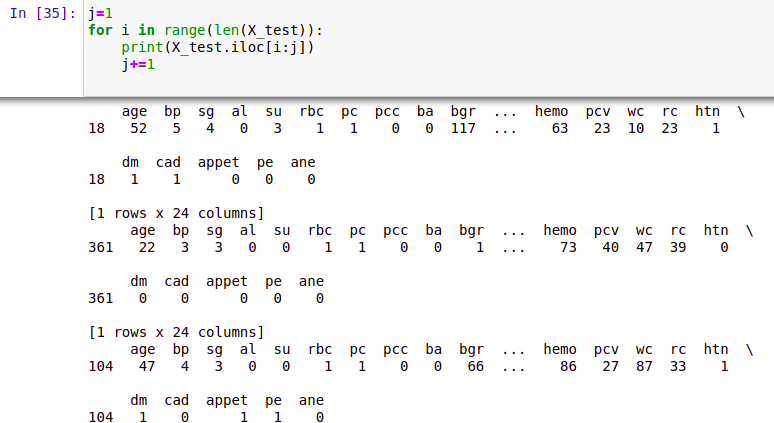


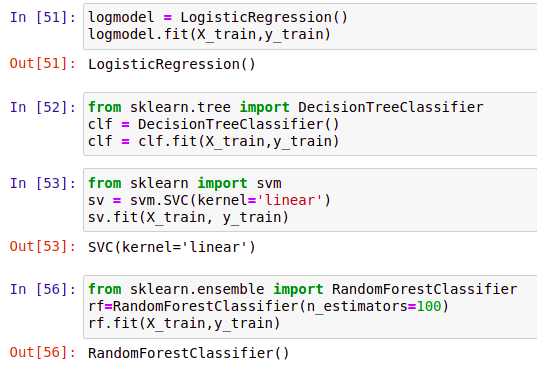




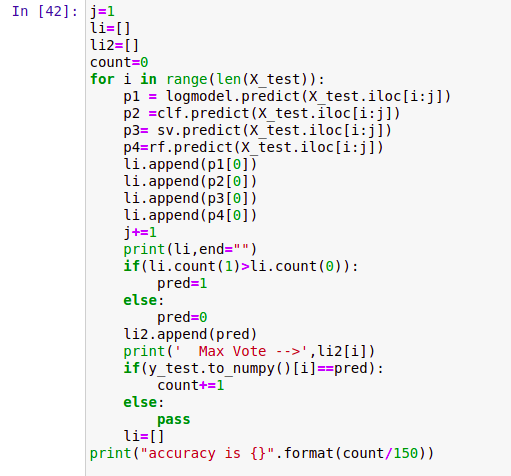


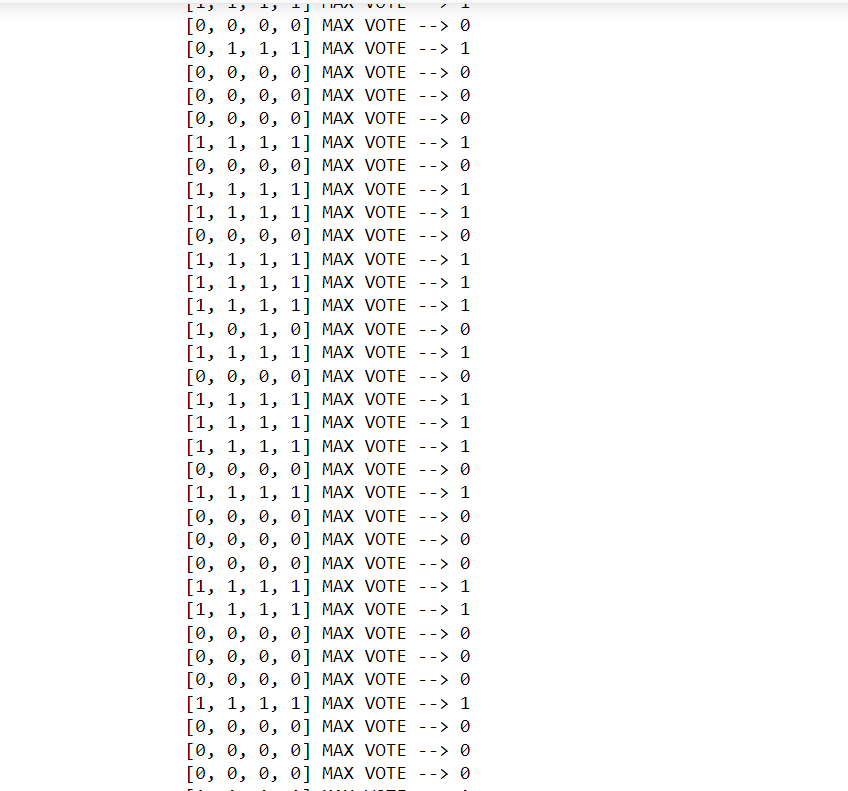
We are taking each individual row from the training features

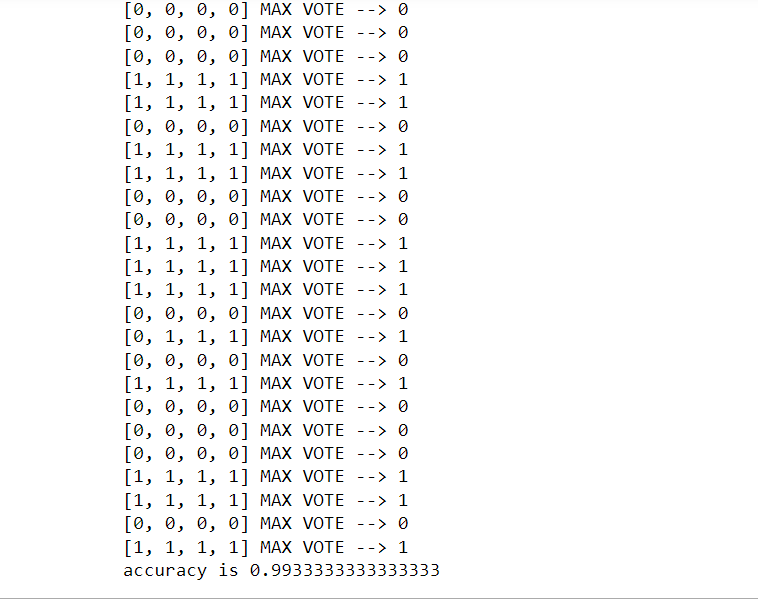




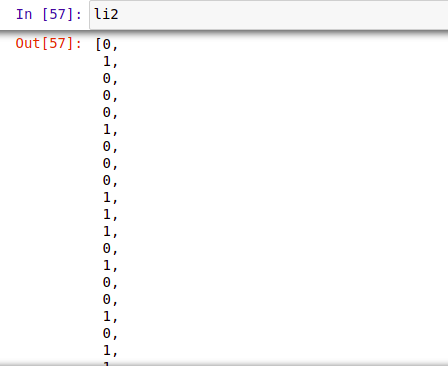
Ensemble technique: Here we are predicting the testing features with each individual algorithm and considering model output as the one which got maximum votes.

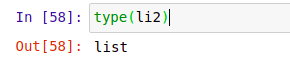




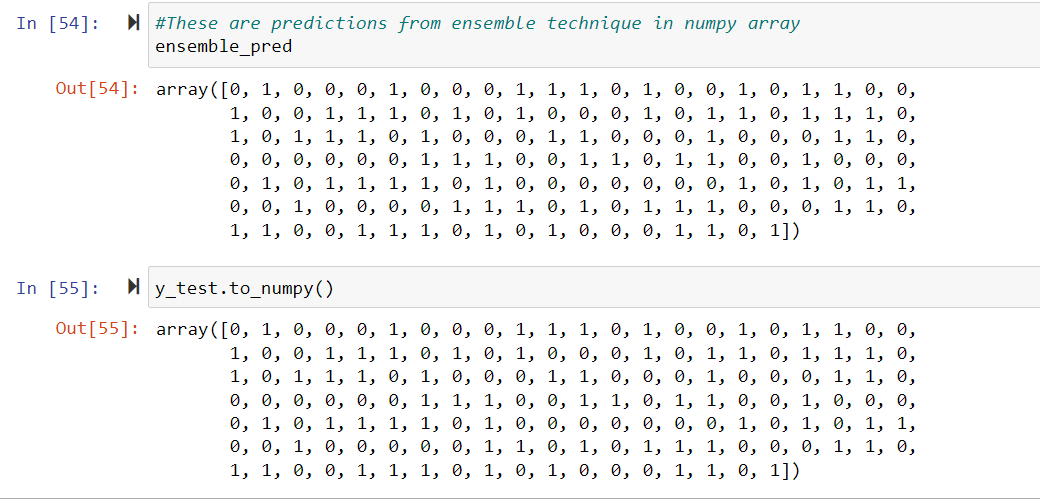


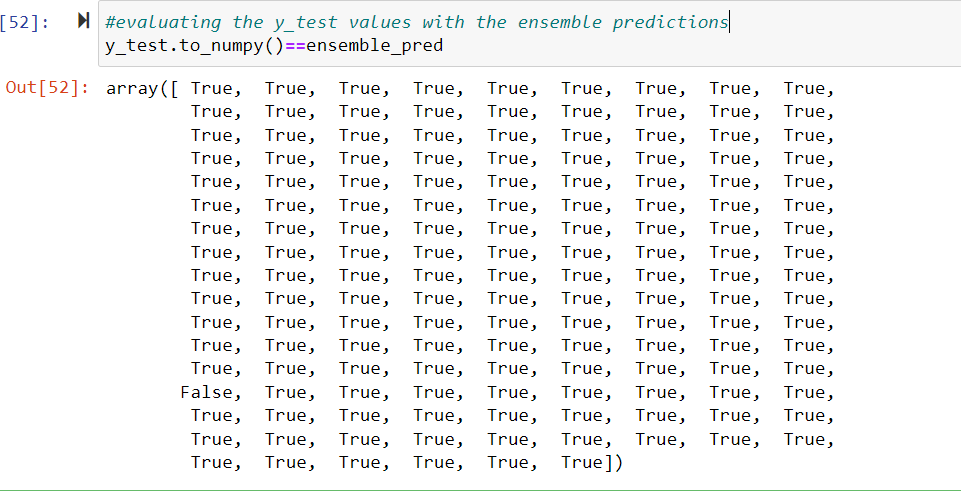
All the votes appended to the list li2 from the ensemble model



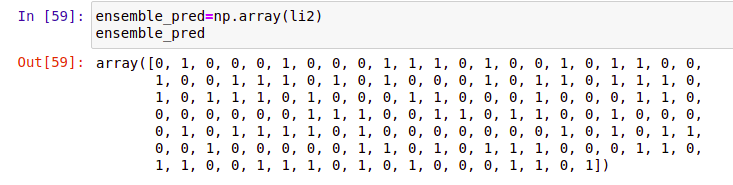


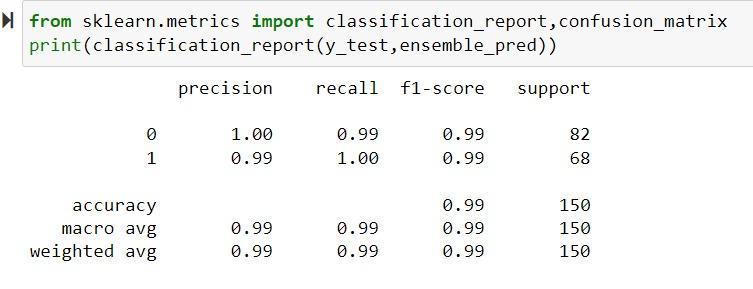
Here we are converting class labels of testing data into numpy array for the comparison of our ensemble predictions with the y test values

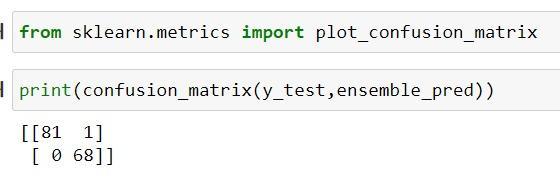


Above we can observe almost out of 150 records it predicted 149 correctly so the technique given really good prediction with 99.33 accuracy

Here we are converting our predictions in to numpy array for evaluation like generating confusion matrix







**CONCLUSION**

The proposed methodology has predicted the chronic kidney disease at higher accuracy of 99.33% with Ensemble technique and given better results than the existing systems and also our model has given good F1-score,precision and recall.

**FUTURE WORK**

we can improve ensemble technique by executing in pyspark or any other environment where the computational time is less and by repeating multiple similar models which is giving high accuracy.

**REFERENCES**

1. Abdulhamit Subas, Emina Alickovic, and Jasmin Kevric , Diagnosis of Chronic Kidney Disease by Using Random Forest, 2017

2. Fuzhe Ma,Tao Sun,Lingyun Liu,Hongyu Jing, Detection and Diagnosis of kidney disease using deep learning, 2020

3. Ahmed Abdelaziz, Ahmed S. Salama, A. M. Riad and Alia N. Mahmoud, A machine learning model for predicting chronic kidney disease, 2018

4. Abdullah Al Imran, Md Nur Amin, Fatema Tuj Johora, classification of kidney disease using logistic regression and feedforward neural network, 2018

5. Pankaj Chittora, Sandeep and Prasun Chakrabarti, Prediction of chronic kidney disease – a machine learning prespective, 2021

6. Tahsin M. Rahman, Saima Siddiqua, Siam E.Rabby, Nahid Hasan, Early dection of kidney disease using EBG signal with machine learning, 2019

7. J Qin, Lin Chen, yuhua Liu and Bin chen , A machine learning methodology for diagnosing kidney disease, 2019

8. S.Revathy, B.Bharathi,P.Jayanthi and M.Ramesh , Prediction of kidney disease using random frost, 2019

9. Sathiya Priya S and Suresh Kumar M, A machine learning methodology for predtion of kidney disease, 2018

10. Imesh Udara Ekanayake, Damayanthi Herath, Chronic kidney disease prediction using machine learning methods, 2020

11. Shanila Yunus Yashfi,Md Ashikul Islam,Pritilata, Risk prediction of chronic kidney disease, 2020 12. Merve Dogruyol, Aydin Akan, Detection of chronic kidney disease using ensemble classifiers, 2017

13. Shanila Yunus Yashfi, Md Ashikul Islam, Pritilata, Risk prediction of kidney disease using ML algorithms, 2020

14. Gopi Battineni, Getu Gamo Sagaro, Nalini Chinatalapudi, Francesco Amenta, Applications of machine learning predictive model for kidney disease prediction, 2020

15. N.Ch.Sriman Narayana Iye ngar, Supervised learning method for prediction of kidney disease ,2017

16. Mohamed Elhoseny K. Shankar & J. Uthayakumar, Intelligent diagnostic prediction and classification system for chronic kidney disease, 2019

17. Huseyin Polat Homay Danaei Mehr & Aydin Cetin, Diagnosis of chronic kidney disease based on support vector machine by feature selection method, 2017

18. Shanmugarajeshwari, V. Ilayaraja.M, Chronic kidney disease for healthcare data analytics using random forest classification algorithms, 2021

19. Ogunleye, A.; Wang, Q , XGBoost model for chronic kidney disease diagnosis, 2020

20. L.Jerlin Rubini, Dr.P.Eswaran, Generating comparative analysis of early stage kidney disease, 2015

21. K. M. Zubair Hasan and Md. Zahid Hasan, Performance evaluation of ensembled-based machine learning technique for prediction of kidney disease, 2019

22. Himanshu Kriplani, Bhumi Patel and Sudipta Roy, Prediction of chronic idney disease using deep artificial neural networks, 2019

23. N V Ganapathi Raju , K Prasanna Lakshmi , K. Gayathri Praharshitha , Chittampalli Likhitha, Prediction of kidney disease using data science, 2019

24. K. M. Zubair Hasan · Md. Zahid Hasan (B) Daffodil International University, 102 Sukrabadh, Mirpur Road, Dhaka 1207, Bangladesh, 2019

25. Emon, Minhaz Uddin Imran, Al Mahmud Islam, Rakibul Keya, Maria Sultana Zannat, Raihana Ohidujjaman, Performance analysis of chronic kidney disease through machine learning approaches, 2021

26. Charleonnan, A Fufaung, T Niyomwong, T Chokchueypattanakit, W. Suwannawach, S Ninchawee.N, Predictive analytics for chronic kidney disease using ,aching learning, 201

27. Sedighi, Zeinab Ebrahimpour-Komleh, Hossein Mousavirad, Seyed Jalaleddin, Feature selection effects on chronic kidney disease, 2015

28. S.Gopika , Dr.M.Vanitha, Machine learning approach on chronic kidney disease using clustering , 2017

29. Arulanthu, Pramila Perumal, Eswaran, Predticting kidney disease using various classfiers , 2019

30. S.Revathy, B.Bharathi,P.Jayanthi and M.Ramesh , Prediction of kidney disease

31. Ebrahime Mohammed Senan, Mosleh Hmoud AlAdhaileh, Ahmed H. Alahmadi,Chronic kidney disease prediction-a machine learning approach, 2021

32. Marwa Almasoud, Tomas E Ward, Prediction of chronic kidney disease using least nuber of predictors, 2019

33. Zvi segal, Dan kalifa, Gideon koren, Machine learning algorithm for early detection of end-stage renal disease, 2020

34. Gopi Battineni, Getu Gamo Sagaro, Nalini Chinatalapudi, Francesco Amenta, Applications of machine learning predictive model for kidney disease prediction, 2020

35. Vineeta Gulati, Neeraj Raheja, comparative analysis of machine learning techniques based on chronic kidney disease dataset, 2021

36. Anik saha, Abir saha, Tanni mitra. Multilayer perceptron based prediction system for chronic kidney disease, 2020

37. N radha, S ramya, Performance analysis of various machine learning algorithms for chronic kidney disease ,2015

38. Mohamed Alloghani, Dhiya Al-Jumeily Abir Hussain, Performance based prediction for chronic kidney disease using machine learning for high risk , 2019

39. Hamid Mohamadlou, Nicholas R. Saber, Ritankar Das, Prediction of acute kidney injury using ML algorithms , 2018

40. S.Revathy, B.Bharathi,P.Jayanthi and M.Ramesh , Prediction of kidney disease using random frost, 2019

41. Jing xiao, Xiulin xu, Xinhui feng, Comparing and Developing machine learning tools for prediction of chronic kidney disease, 2019

42. Geetanjali Bhola Aman Garg Manisha Kumari, Comparative study of machine learning techniques for chronic kidney disease, 2020

43. Theyazn H.H Aldhyani, Ali Saleh Alshebami, Mohammed Y. Alzahrani, Soft clustering for diagnosis of chronic kidney disease over machine learning , 2020

44. Zvi segal, Dan kalifa, Gideon koren, Machine learning algorithm for early detection of end-stage renal disease, 2020

45.Abdullah Al Imran,Md Nur Amin,Fatema Tuj Johora,Classification of Chronic Kidney Disease using Logistic Regression, Feedforward Neural Network and Wide & Deep Learning,2019