*Chronical Kidney Disease Prediction using*

*AdaBoost Classifier*

P.Niharika  
*Department of Computer Science and Engineering*  
*Bapatla Engineering College  
(Autonomous)  
(Affiliated to Acharya Nagarjuna University)*Bapatla, India  
niharikaputta10@gmail.com

V.Joshna Florence  
*Department of Computer Science and Engineering*  
*Bapatla Engineering College  
(Autonomous)  
(Affiliated to Acharya Nagarjuna University)*Bapatla, India  
joshnaflorence7@gmail.com

B.Gowtham  
*Department of Computer Science and Engineering*  
*Bapatla Engineering College  
(Autonomous)  
(Affiliated to Acharya Nagarjuna University)*Bapatla, India  
battinagowtham@gmail.com

S.Mohan  
*Department of Computer Science and Engineering*  
*Bapatla Engineering College  
(Autonomous)  
(Affiliated to Acharya Nagarjuna University)*Bapatla, India  
 mohan@gmail.com

*Abstract*—The large amount of data that are generated from healthcare are too complex to analyzed manually Legitimately analyzing clinical records approximately patients wellbeing expect the plausibility of an event of various infections. In expansion, Procuring data concerning specialists of that specific infection as per the prerequisite facilitates legitimate and proficient determination of disease. Nowadays doctors are more focused on corona disease by forgetting other diseases which are present. No proper treatments are given for chronic kidney disease which is also a major disease in the world. Early prediction and proper medication for betterment is the right way to help the people and physicians. By the use of Machine learning techniques data can be analysed easily to provide better treatments for patients. Analysis of data and prediction of disease can be early done through machine learning. Machine learning plays a vital role in the healthcare industry and helps in prediction of disease from the datasets by using machine learning techniques like Adaboost. Keywords—Machine Learning, Chronic Kidney Disease, AdaBoost.

# INTRODUCTION

Healthcare industry generates terabytes of data every year. The medical documents maintained are a pool of information regarding patients. The task of extracting useful information or quality healthcare is tricky and important. By analysing these voluminous data, we can predict the occurrence of the disease and safe guard people. Thus, an intelligent system for disease prediction plays a major role in controlling the disease and maintaining the good health status for people by providing accurate and trustworthy disease risk prediction.

Machine learning is field concerned with the study of large and numerous variable information. In Health Care discerning, Machine learning guarantees to help doctors to form perfect determination, suggests the leading medicines for the patient’s, spot patients at high-risk for pitiable results and particularly progressing patient’s physical condition whereas minimizing costs. Machine learning has demonstrated a victory in forecast and conclusion of different basic illness.

Chronic Kidney disease is worldwide health disease with higher burden with regard to the wellbeing within the show circumstance. Chronic Kidney infection is characterized as a glomerular filtration rate(GFR)<60mL/min or Kidney harm or both for at slightest a period of 3 months. End-stage renal illness is completely connected with mortality. Chronic Kidney is recognized with research facility tests. Major downside of this disease is, most of the time CKD is recognized at its last stage and which too leads to kidney failure. Within the early stages of chronic kidney illness, there will be few signs or side effects. CKD may not ended up clear until kidney work is altogether disabled.

Chronic kidney malady can be advance to conclusion organize of kidney failure, which is fatal without dialysis or a kidney transplantation. CKD is complicated illness by influencing the parts of the body by causing anemia, cardiovascular disease, Decreased Immune system, harm to central nervous system. It is exceptionally critical to urge check-up patient within short period of time.

# RELEATED WORK

Using a variety [3] of machine learning approaches, Farjana et al. look at the prediction of CKD. They assess the performance of several algorithms, such as LR, RF, DT, SVM, and K-NN. The dataset used by the authors contains clinical and laboratory information from people with CKD.

The prediction of CKD is investigated by Anil et al [6].utilising several ML techniques. They evaluate the effectiveness of several algorithms, including k-NN, SVM,RF, and LR. The authors make use of a dataset that includes clinical and laboratory data from individuals with CKD.A severe global health issue, chronic kidney disease(CKD) [1] affects millions of individuals globally. Early and accurate diagnosis of CKD is essential for effective management and therapy. Recently, machine learning algorithms have showed a lot of promise for aiding doctors in the diagnosis of CKD. This section discusses a variety of research that have looked at the use of machine learning algorithms for the prognosis and diagnosis of CKD in this review of the literature. The emphasis of Aruna and Sameerunnisa [13] is on the value of data pre-processing methods in CKD prediction. They look on how feature extraction and selection techniques affect CKD prediction model performance. For categorization, the authors use algorithms like Naive Bayes, Decision Trees, and SVM.

Performance criteria including accuracy, sensitivity, specificity, and F1-score are used in the research to assess the models. Kumari and Singh [14] provide an ensemble learning based strategy for precise CKD prediction. By mixing several ML methods, including RF, DT, SVM, and k-NN, they create an ensemble model. The dataset used by the authors includes clinical and laboratory information on persons with CKD .Raju et al. [20] concentrate on applying data science methods to forecast chronic renal disease. They analyse a dataset that contains the clinical and laboratory characteristics of CKD patients using ML methods including SVM, RF, and k-NN.Machine learning is a very popular tool for predicting various diseases based on data sets [22] [23] [24]. The AdaBoosting ensemble approach and K-fold cross-validation are suggested by Suganthi et al. [10] for the identification of CKD. They contrast the effectiveness of Ada Boosting with that of Decision Tree and SVM, two additional machine learning techniques. The authors make use of a dataset that includes clinical and laboratory data from individuals with CKD. Devika et al. [12] compare the performance of Naïve Bayes, KNN, and Random Forest algorithms as classifiers for the prediction of CKD. They use a dataset of CKD patient clinical and laboratory characteristics. A Multilayer Perceptron (MLP) model is used by Yildirim [15] to solve the problem of unbalanced data in CKD prediction. The goal of the research is to enhance the performance of prediction by optimizing the MLP architecture.

The clinical and laboratory characteristics of CKD patients are included in the dataset utilised in this investigation. Ekanayake and Herath [4]research the prediction of chronic kidney disease (CKD) using machine learning approaches. They assess the performance of several techniques, including Support Vector Machines (SVM), Random Forest, and k-Nearest Neighbours (k-NN),with the objective of predicting CKD. The authors' dataset contains the clinical and laboratory features of people with CKD. Performance is assessed using variables including accuracy, precision, recall, and F1-score. Bhowmick and V M[5] concentrate on utilising machine learning models to analyse and predict CKD. They assess how well CKD prediction methods like DT, RF, SVM, and Naive Bayes perform. The authors make use of a dataset that includes clinical and laboratory data from individuals with CKD. Metrics including accuracy, sensitivity, specificity, and F1-score are used to assess performance.

An intelligent clinical support system for the early identification of CKD is suggested by Swain et al [16]. To create a prediction model, they use ML techniques including DT, RF, and k-NN. The authors concentrate on the use of clinical and laboratory characteristics for CKD early identification. ML-based recommendations for an appropriate diet plan and the prediction of CKD are the main areas of interest for Maurya et al [21]. Based on clinical and laboratory characteristics, they use algorithms like Decision Tree and Naive Bayes to predict the presence of CKD. The writers also provide individual nutrition advice using a knowledge based methodology.

For the purpose of predicting CKD, Arif-Ul-Islam and Ripon [11] suggest using boosting classifiers, Ant-Miner, and J48 decision tree algorithms. To create prediction models, they concentrate on rule induction techniques. The authors make use of a dataset that includes clinical and laboratory data from individuals with CKD. The performance assessment of ML classification algorithms for illness categorization and CKD predicting is the main topic of Gunarathne et al [17].They evaluate the effectiveness of several algorithms, including Decision Tree, k-NN, and SVM. The authors make use of a dataset that includes clinical and laboratory data from individuals with CKD. Sisodia and Verma [18] assess the effectiveness of individual and group learners' predictions for CKD. They contrast several algorithms, including Random Forest, SVM, k-NN, and ensemble approaches. The authors make use of a dataset that includes clinical and laboratory data from individuals with CKD. Research Statement: It is observed from this literature survey that ML models are very useful in predicting CKD. But still this disease is uncontrolled because of less prediction accuracy of ML models. So there is a need to do further research to improve the accuracy of ML models for prediction of CKD so that appropriate actions can be taken timely to avoid CKD.

# EXISTING SYSTEM

The existing system develops the most accurate model for

predicting chronic kidney disease (CKD) by assessing

various machine learning algorithms predictive capacities.

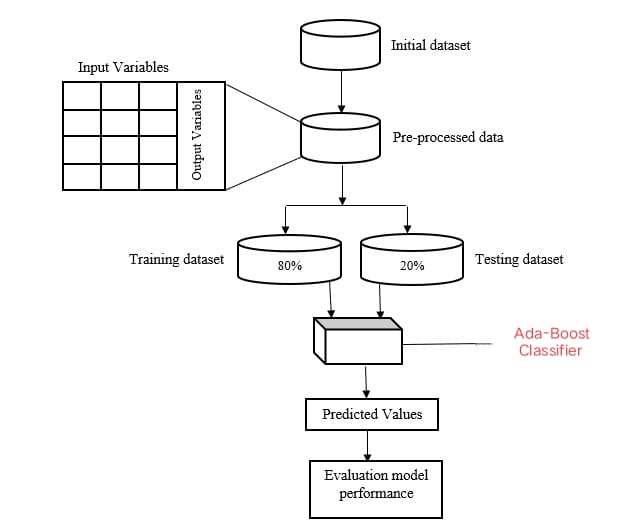
The research compares the performance of three ML

algorithms- Random forest, Adaboost and Decision tree

classifier

# PROPOSED SYSTEM

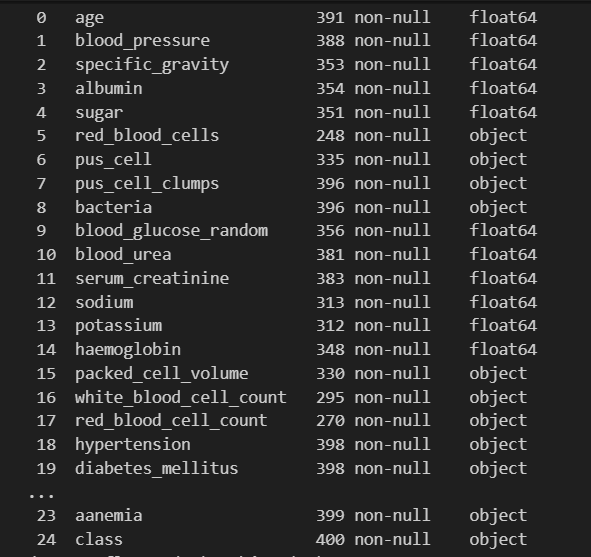
Prediction is a statement about future events. Chronic Kidney Disease Prediction has become the need of the patients and Physician. Although future events are uncertain, so accurate prediction is not possible. The proposed system aims to develop a model which predicts Chronic Kidney Disease(CKD) by implementing an AdaBoost Classifier trained on preprocessed medical data that can be helpful for doctors to provide better medication and also for patients.

Fig 1.1 Architecture of Proposed system

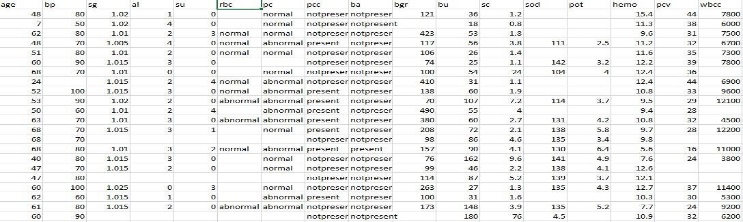
DATASET

This dataset has been collected from Kaggle Repository.

It contains 26 attributes and 401 instances.



There are 25 features and 1 class label for every chronic kidney disease record, and the features include basic age, bp, sugar, serum creatine, sodium, haemoglobin etc.



DATA PREPROCESSING

The data preprocessing steps involve dropping the 'id' column, renaming columns, converting specific columns to numeric type, handling categorical data by replacing values and mapping labels, and encoding categorical variables.

The 'class' column, representing the target variable, is cleaned by correcting values and mapping them to numeric labels (0 for 'ckd' and 1 for 'not ckd'). Remaining categorical columns are encoded using Label Encoder, converting them into numerical labels. Missing numerical values are filled using random value imputation, while missing categorical values are imputed with the mode. Overall, these steps ensure data consistency and completeness for further analysis.

SPLITTING DATA

Splits the data into independent (X) and dependent (y) variables. Splits the data into training and testing sets using a 80-20 split ratio.

MODEL TRAINING

Imports AdaBoost Classifier from sklearn.ensemble and

initializes an AdaBoost Classifier and fits it to the training data.

EVALUATION

Calculates training and test accuracies of the AdaBoost Classifier. Prints the confusion matrix and classification report for the test data.

ADVANTAGES OF PROPOSED SYSTEM

* + Useful to health department to predict the CKD.
  + Useful for the patients to take better recovery.
  + We use data science techniques for accurate results.
  + On click of button output will be generated, no too much time required for CKD prediction. ▪ No need to analyze manually.

APPLICATION

* + Proposed system can be used in medical department for the prediction of CKD .
  + Proposed system can be used by patients to know the if the CKD is present or not by inputting data such as “age”, “blood pressure”, “serumcreatine”, “sugar” and “bacteria” etc.

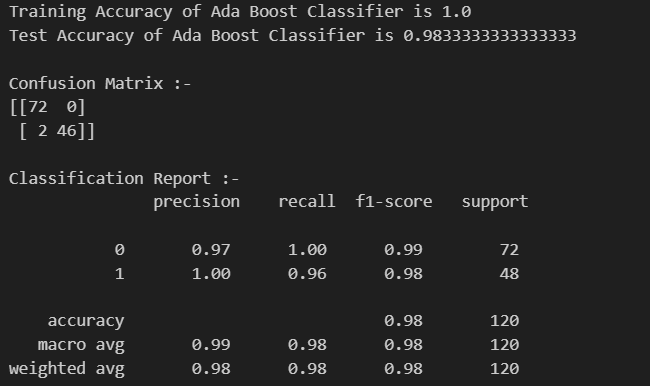
# RESULTS

Finally concluding that Adaboost Algorithm achieved

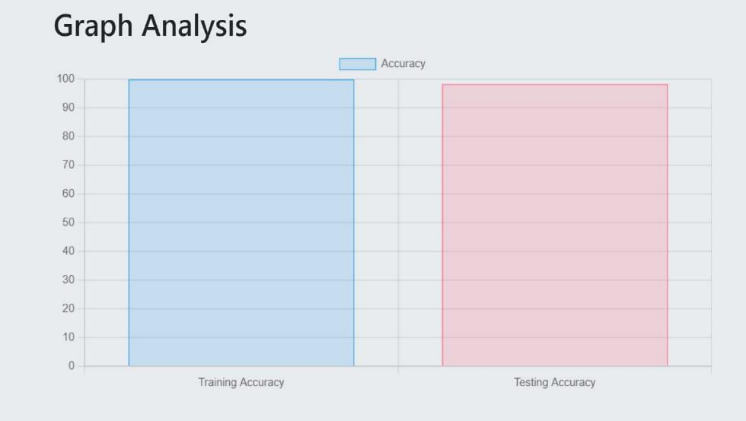
98% Accuracy here some of the metrics for the

Adaboost

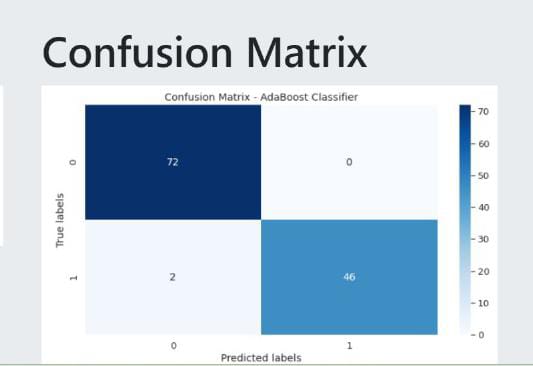
Classification Report:



Training and Testing Accuracy Bar Graph:



Confusion Matrix:



# Conclusion

In this project, CKD prediction has been accomplished using the ensemble model from the CKD dataset as they can

reduce the risk factors and improve the outcome in

terms of efficiency and accuracy. We collected

diagnostic data set with 26 CKD attributes of 401

patients for the study. Based on these attributes we

applied basic machine learning algorithms

ensembled method AdaBoost .After the comparative

analysis among all the models, it is evident that the

ensembled model Ada Boost accuracy supersedes

over other models with the accuracy of 98%.

##### ACKNOWLEDGEMENT

The authors would like to than Ass.Prof. Mr M M

Meera Durga sir,Bapatla Engineering College,Bapatla

for guiding throughout the work and the authors

would also thank research paper writers for proving

base to our work

##### REFERENCE

1. Bikbov B, Purcell CA, Levey AS, Smith M, Abdoli A, Abebe M, Adebayo OM, Afarideh M, Agarwal SK, Agudelo-Botero M, Ahmadian E. Global, regional, and national burden of chronic kidney disease, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet. 2020 Oct 24;396(10258):1135-49. DOI: 10.1016/S0140-6736(20)30888-9
2. Hill NR, Fatoba ST, Oke JL, Hirst JA, O'Callaghan CA, Lasserson DS, Hobbs FD. Global prevalence of chronic kidney disease–a systematic review and meta-analysis. PloS one. 2016 Jul 6;11(7):e0158765. DOI: 10.1371/journal.pone.0158765
3. Gansevoort RT, Correa-Rotter R, Hemmelgarn BR, Jafar TH, Heerspink HJ, Mann JF, Matsushita K, Wen CP. Chronic kidney disease and cardiovascular risk: epidemiology, mechanisms, and prevention. The Lancet. 2013 Jul 13;382(9889):339-52. DOI: 10.1016/S0140-6736(13)60595-4
4. Webster AC, Nagler EV, Morton RL, Masson P. Chronic kidney disease. The Lancet. 2017 Mar 11;389(10075):1238-52. DOI: 10.1016/S0140-6736(16)32064-5
5. Matsushita K, Ballew SH, Astor BC, Jong PE, Gansevoort RT, Hemmelgarn BR, Levey AS, Levin A, Wen CP, Woodward M, Coresh J. Cohort profile: the chronic kidney disease prognosis consortium. International journal of epidemiology. 2013 Apr 1;42(6):1660-8. DOI: 10.1093/ije/dys173
6. Grams ME, Sang Y, Ballew SH, Gansevoort RT, Kimm H, Kovesdy CP, Naimark D, Oien C, Smith DH, Coresh J, Sarnak MJ. A meta-analysis of the association of estimated GFR, albuminuria, diabetes mellitus, and hypertension with acute kidney injury. American Journal of Kidney Diseases. 2015 Mar 1;66(3):602-12. DOI: 10.1053/j.ajkd.2015.03.017
7. Matsushita K, van der Velde M, Astor BC, Woodward M, Levey AS, de Jong PE, Coresh J, Gansevoort RT, Consortium CKDP. Association of estimated glomerular filtration rate and albuminuria with all-cause and cardiovascular mortality in general population cohorts: a collaborative meta-analysis. The Lancet. 2010 Jun 12;375(9731):2073-81. DOI: 10.1016/S0140-6736(10)60674-5
8. Schaeffner ES, Kurth T, Curhan GC, Glynn RJ, Rexrode KM, Baigent C, Buring JE, Gaziano JM. Cholesterol and the risk of renal dysfunction in apparently healthy men. Journal of the American Society of Nephrology. 2003 Nov 1;14(11):2084-91. DOI: 10.1097/01.ASN.0000083905.84248.61
9. Thomas B, Matsushita K, Abate KH, Al-Aly Z, Ärnlöv J, Asayama K, Atkins R, Badawi A, Ballew SH, Banerjee A, Barregard L. Global cardiovascular and renal outcomes of reduced GFR. Journal of the American Society of Nephrology. 2017 Jan 1;28(7):2167-79. DOI: 10.1681/ASN.2016050562
10. Levin A, Stevens PE, Bilous RW, Coresh J, De Francisco AL, De Jong PE, Griffith KE, Hemmelgarn BR, Iseki K, Lamb EJ, Levey AS. Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. Kidney international supplements. 2013 Jan 1;3(1):1-150. DOI: 10.1038/kisup.2012.73