International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

# CHRONIC KIDNEY DISEASE PREDICTIONS USING MACHINE LEARNING MODELS

[1]Dr.R.Sasthees Kuamr, [2]Shaikafrin, [3]Vidiyala Ls Charishma Devi, [4]Battula Mythri Kavya Lakshmi, [5]Donthi Bala Amrutha Susmitha, [6]Papasani Siva
[1,2,3,4,,5,6]Assistant professors, Department of CSE in Narasaraopet Institute Of Technology

## ABSTRACT

Chronic kidney disease (CKD) refers to the many clinical abnormalities that progressively worsen as kidney function declines.CKD results from a large number of systemic diseases that damage the kidney or from disorders that are intrinsic to the kidney. Chronic kidney disease (CKD) is a global public health problem with a rising prevalence. (GFR) is considered the best overall index of kidney function, and low GFR is associated with higher risk of kidney failure requiring dialysis and cardiovascular disease, hypertension, anaemia, and other metabolic complications. Logistic regression, support vector machines, random forest, and gradient boosting algorithms have been trained and tested using 10-fold cross-validation. We achieve an accuracy of 99.1 according to F1measure from Gradient Boosting classifier. Also, we found that haemoglobin has higher importance for both random forest and gradient boosting in detecting CKD.

## INTRODUCTION

Chronic kidney disease (CKD) is a significant public health problem worldwide, especially for low and medium income countries. Chronic kidney disease (CKD) means that the kidney does not work as expected and cannot correctly filter blood. About 10% of the population worldwide suffers from (CKD), and millions die each year because they cannot get affordable treatment, with the number increasing in the elderly. According to the Global Burden Disease 2010 study conducted by the International Society of Nephrology, chronic kidney disease (CKD) has been raised as an important cause of mortality worldwide with the number of deaths increasing by 82.3% in the last two decades [1, 2]. Also, the number of patients reaching end-stage renal disease (ESRD) is increasing, which requires kidney transplantation or dialysis to save patients' lives. Aimed in their work to detect chronic kidney disease for diabetic patients using machine learning methods. In their research, they used 600 clinical records collected from a leading Chennai-based diabetes research centre. The authors have tested the dataset using the decision tree and Naïve Bayes methods for classification using the machine learning tool. They concluded that the decision tree algorithm outweighs the Naïve Bayes with an accuracy of 91%. Predict the patient status of CKD. By considering 14 attributes out of 25, they compared four different algorithms, which were Multiclass Decision Forest, Multiclass Decision Jungle, Multiclass Decision Regression, and Multiclass Neural Network. After comparison, they found that Multiclass Decision Forest performed the best with 99.1% accuracy.

**LITERATURE SURVEY**

1. M. P. N. M. Wickramasinghe, D. M. Perera and K. A. D. C. P. Kahandawaarachchi, "Dietary prediction for patients with Chronic Kidney Disease (CKD) by considering blood potassium level using machine learning algorithms," 2017 IEEE Life Sciences Conference (LSC), Sydney, NSW, 2017, pp. 300-303.

Kidney damage and diminished function that lasts longer than three months is known as Chronic Kidney Disease (CKD). The primary goal of this research study is to identify the suitable diet plan for a CKD patient by applying the classification algorithms on the test result obtained from patients' medical records. The aim of this work is to control the disease using the suitable diet plan and to identify that suitable diet plan using classification algorithms. The suggested work pacts with the recommendation of various diet plans by using predicted potassium zone for CKD patients according to their blood potassium level. The experiment is performed on different algorithms like Multiclass Decision Jungle, Multiclass Decision Forest, Multiclass Neural Network and Multiclass Logistic Regression. The experimental results show that Multiclass Decision Forest algorithm gives a better result than the other classification algorithms and produces 99.17% accuracy

2. H. A. Wibawa, I. Malik and N. Bahtiar, "Evaluation of Kernel-Based Extreme Learning Machine Performance for Prediction of Chronic Kidney Disease," 2018 2nd International Conference on Informatics and Computational Sciences (ICICoS), Semarang, Indonesia, 2018, pp. 1-4 The field of biosciences have advanced to a larger extent and have generated large amounts of information from Electronic Health Records. This have given rise to the acute need of knowledge generation from this enormous amount of data. Data mining methods and machine learning play a major role in this aspect of biosciences. Chronic Kidney Disease(CKD) is a condition in which the kidneys are damaged and cannot filter blood as they always do. A family history of kidney diseases or failure, high blood pressure, type 2 diabetes may lead to CKD. This is a lasting damage to the kidney and chances of getting worser by time is high. The very common complications that results due to a kidney failure are heart diseases, anemia, bone diseases, high potasium and calcium.An early detection of CKD can improve the quality of life to a greater extent. This calls for good prediction algorithm to predict CKD at an earlier stage. Literature shows a wide range of machine learning algorithms employed for the prediction of CKD. This paper uses data preprocessing,data transformation and various classifiers to predict CKD and also proposes best Prediction framework for CKD. The results of the framework show promising results of better prediction at an early stage of CKD.

**Existing System**

We can also evaluate and compare the performance of the used classifiers with other existing classifiers. CKD early

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

detection helps in timely treatment of the patients suffering from the disease and also to avoid the disease from getting worse. Early prediction of the disease and timely treatment are the need for medical sector. New classifiers can be used and their performance can be evaluated to find better solutions of the objective function in future work.

**3.2 Disadvantages:**

 Less efficient.
 No accuracy score
 Unble to identify the disease

**3.3 Proposed System**

The work proposed here uses three classification techniques to predict the presence of chronic kidney disease in humans. The classifiers used are Support vector machine and KNN classifier. The data set for chronic kidney disease was gathered and applied on each classifier to predict the disease and the performance of the classifiers evaluated based on accuracy, precision and F measure.

**3.4 Advantages**

 Performance More efficient.
 Good accuracy.
 Predict the disease

## 4. SYSTEM STUDY

### 4.1 FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are

 ECONOMICAL FEASIBILITY
 TECHNICAL FEASIBILITY
 SOCIAL FEASIBILITY

#### 4.1.1 ECONOMICAL FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

#### 4.1.2 TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have amodest requirement, as only minimal or null changes are required for implementing this system.

#### 4.1.3 SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

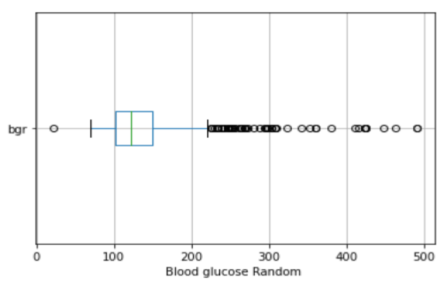
## 5.Result:

### Data description:

| | id | age | bp | sg | al | su | bgr | bu | sc | sod | pot | hemo | pcv | wc | rc |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| count | 400.000000 | 391.000000 | 388.000000 | 353.000000 | 354.000000 | 351.000000 | 356.000000 | 381.000000 | 383.000000 | 313.000000 | 312.000000 | 348.000000 | 330.000000 | 295.000000 | 270.000000 |
| mean | 199.500000 | 51.483376 | 76.469072 | 1.017408 | 1.016949 | 0.450142 | 148.036517 | 57.425722 | 3.072454 | 137.528754 | 4.627244 | 12.526437 | 38.766667 | 8377.630508 | 4.700370 |
| std | 115.614301 | 17.169714 | 13.683637 | 0.005717 | 1.352679 | 1.099191 | 79.281714 | 50.503006 | 5.741126 | 10.408752 | 3.193904 | 2.912587 | 9.228118 | 2979.918951 | 1.029978 |
| min | 0.000000 | 2.000000 | 50.000000 | 1.005000 | 0.000000 | 0.000000 | 22.000000 | 1.500000 | 0.400000 | 4.500000 | 2.500000 | 3.100000 | 0.000000 | 1.000000 | 2.100000 |
| 25% | 99.750000 | 42.000000 | 70.000000 | 1.010000 | 0.000000 | 0.000000 | 99.000000 | 27.000000 | 0.900000 | 135.000000 | 3.800000 | 10.300000 | 32.000000 | 6500.000000 | 3.900000 |
| 50% | 199.500000 | 55.000000 | 80.000000 | 1.020000 | 0.000000 | 0.000000 | 121.000000 | 42.000000 | 1.300000 | 138.000000 | 4.400000 | 12.650000 | 40.000000 | 8000.000000 | 4.800000 |
| 75% | 299.250000 | 64.500000 | 80.000000 | 1.020000 | 2.000000 | 0.000000 | 163.000000 | 66.000000 | 2.800000 | 142.000000 | 4.900000 | 15.000000 | 45.000000 | 9800.000000 | 5.400000 |
| max | 399.000000 | 90.000000 | 180.000000 | 1.025000 | 5.000000 | 5.000000 | 490.000000 | 391.000000 | 76.000000 | 163.000000 | 47.000000 | 17.800000 | 54.000000 | 26400.000000 | 8.000000 |

Fig 10.6: Data description

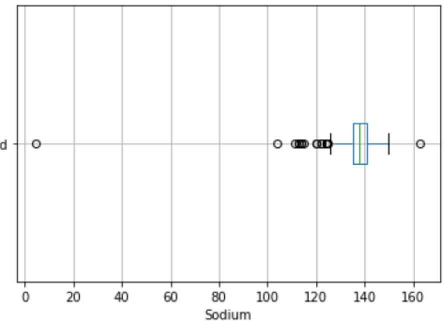
### Blood Glucose Radom:




**Blood Glucose Radom**

*Pottasium:*

**Sodium:**



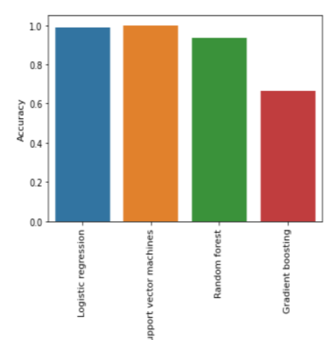

**Sodium**

### Algorithms accuracy:

| | Classifier | Accuracy | F1 | Precision | Sensitivity | Specificity |
|---|---|---|---|---|---|---|
| 0 | Logistic regression | 0.9875 | 98.75 | 98.79 | 98.11 | 1.0 |
| 1 | Support vector machines | 1.0000 | 52.80 | 43.89 | 1.00 | 0.0 |
| 2 | Random forest | 0.9375 | 1.00 | 1.00 | 1.00 | 1.0 |
| 3 | Gradient boosting | 0.6625 | 1.00 | 1.00 | 1.00 | 1.0 |

**Algorithms accuracy**

```
(array([0, 1, 2, 3]),
 [Text(0, 0, 'Logistic regression'),
  Text(1, 0, 'Support vector machines'),
  Text(2, 0, 'Random forest'),
  Text(3, 0, 'Gradient boosting')])
```




**Accuracy result**

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

## CONCLUSION:

Work examines the ability to detect CKD using machine learning algorithms while considering the least number of tests or features. We approach this aim by applying four machine learning classifiers: logistic regression, SVM, random forest, and gradient boosting on a small dataset of 400 records. In order to reduce the number of features and remove redundancy, the association between variables have been studied. A filter feature selection method has been applied to the remaining attributes and found that there are haemoglobin, albumin, and specific gravity have the most impact to predict the CKD.

## REFERENCES

**1. M. P. N. M. Wickramasinghe, D. M. Perera and K. A. D. C. P. Kahandawaarachchi, "Dietary prediction for patients with Chronic Kidney Disease (CKD) by considering blood potassium level using machine learning algorithms," 2017 IEEE Life Sciences Conference (LSC), Sydney, NSW, 2017, pp. 300-303.**

**2. 2. H. A. Wibawa, I. Malik and N. Bahtiar, "Evaluation of Kernel-Based Extreme Learning Machine Performance for Prediction of Chronic Kidney Disease," 2018 2nd International Conference on Informatics and Computational Sciences (ICICoS), Semarang, Indonesia, 2018, pp. 1-4**

**3. 3. U. N. Dulhare and M. Ayesha, "Extraction of action rules for chronic kidney disease using Naïve bayes classifier," 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Chennai, 2016, pp. 1-5.**

**4. 4. H. Zhang, C. Hung, W. C. Chu, P. Chiu and C. Y. Tang, "Chronic Kidney Disease Survival Prediction with Artificial Neural Networks," 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, 2018, pp. 1351-1356**

**5. 5. J. Aljaaf et al., "Early Prediction of Chronic Kidney Disease Using Machine Learning Supported by Predictive Analytics," 2018 IEEE Congress on Evolutionary Computation (CEC), Rio de Janeiro, 2018, pp. 1-9.**

**6. 6. Arif-Ul-Islam and S. H. Ripon, "Rule Induction and Prediction of Chronic Kidney Disease Using Boosting Classifiers, Ant-Miner and J48 Decision Tree," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox'sBazar, Bangladesh, 2019, pp. 1-6.**

**7. G. Kaur and A. Sharma, "Predict chronic kidney disease using data mining algorithms in hadoop," 2017 International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, 2017, pp. 973-979.**

**8. N. Tazin, S. A. Sabab and M. T. Chowdhury, "Diagnosis of Chronic Kidney Disease using effective classification and feature selection technique," 2016 International Conference on Medical Engineering, Health Informatics and Technology (MediTec), Dhaka, 2016, pp. 1-6.**