**CHRONIC KIDNEY DISEASE PREDICTION USING MACHINE LEARNING**

Ashif Ahmed(22MCS0032)

Pratik Kumar(22MCS0035)

Anand Mohan(22MCS0048)

Department of computer science and engineering School of Computer science and engineering

Vellore Institute of Technology, Vellore – 632014, Tamil Nadu, India.

[b] Dr . ABDUL GAFFAR H

Vellore Institute of Technology

Vellore – 632014, Tamil Nadu, India

E-mail: abdulgaffar@vit.ac.in

E-mail: [ashif.ahmed2022@vitstudent.ac.in](mailto:ashif.ahmed2022@vitstudent.ac.in)

E-mail : [pratik.kumar2022@vitstudent.ac.in](mailto:pratik.kumar2022@vitstudent.ac.in)

E-mail: [anand.mohan2022@vitstudent.ac.in](mailto:anand.mohan2022@vitstudent.ac.in)

**ABSTRACT:**

The value and fragility of human life should be respected. Death is unavoidable, but thanks to advances in science and technology, there are many ways to improve everyone's quality of life through research and analysis of the different aspects of a healthy body. As everyone is aware, in today's world. Every fifth person has a chronic illness of some kind. Chronic conditions impede daily life activities, necessitate ongoing medical care, or both. The sooner a person's sickness is predicted, the simpler it will be for him or her to change his or her lifestyle. In our study, we will attempt to predict the numerous aspects of the chronic ailment "kidney disease" using machine learning techniques. Chronic kidney disease, also referred to as chronic kidney failure, is the continuous decline in kidney function. Our kidneys filter out wastes and surplus fluid from our blood, which is ultimately excreted as urine. When chronic kidney disease reaches an advanced level, our bodies may begin to amass dangerous quantities of fluid, electrolytes, and waste products. In the early phases of chronic renal illness, there may not be many symptoms or warning signs. The signs of chronic kidney disease could not appear until our renal function is severely impaired.

**I. INTRODUCTION**

Chronic kidney disease, also referred to as chronic kidney failure, is the continuous decline in kidney function. Wastes and extra fluid are removed from the circulation by the kidneys and expelled in the urine. Dangerous concentrations of fluid, electrolytes, and waste products can accumulate in the body when chronic kidney disease reaches an advanced stage your body. There may not be many symptoms or warning indications kidney disease in its early stages. It's possible that symptoms of chronic kidney disease won't appear until our renal function has been severely damaged. By addressing the underlying cause, chronic kidney disease is typically treated with the intention of delaying the progression of kidney damage. End-stage renal failure brought on by chronic kidney disease is fatal without mechanical filtering (dialysis) or a kidney transplant.

**II. LITERATURE REVIEW**

1.The present diagnostic procedure makes use of serum creatinine levels in conjunction with urine analyses. For this goal, a variety of medical techniques, including screening, ultrasonography procedure. During the screening process, patients are checked for hypertension, cardiovascular disease history, current sickness, and family history of renal disease. This technique involves predicting GFR from the blood creatinine level and measuring the albumin to creatinine ratio (ACR) in a first-morning urine sample. In order to increase prediction accuracy, this research focuses on machine learning approaches like ACO and SVM by decreasing the features and choosing the best features

2.In order to predict CKD status using clinical data, the current work offers a technique that involves data preparation, a mechanism for handling missing values, data aggregation, and feature extraction. Three distinct models were trained in this study for precise prediction using a range of physiological data using various techniques such logistic regression (LR), decision tree (dt) classification, and k-nearest neighbor (KNN). With a precision of 96.25 percent, the decision tree method and logistic regression can be utilized to predict chronic renal disease more precisely.

3. A condition called chronic kidney disease (CKD), which is defined by a gradual decline in renal function over months to years without any overt symptoms, carries a significant risk of death. According to the intensity, there are six levels. It is divided into stages based on the glomerular filtration rate (GFR), which depends on elements like age, sex, race, and serum creatinine. Because it enables the identification of all phases of CKD, the chronic renal disease epidemiology collaboration (CKD-epi), a linear model that is one of many models available for estimating GFR value, has been shown to be incredibly effective. A comparison of the results revealed that j48 predicted CKD at all phases more accurately than random forest, with an accuracy of 85.5%. Additionally, the study showed that J48 performs better than random forest.

4.A data preprocessing step that makes use of the Synthetic Minority Oversampling Technique (SMOTE), which is necessary to guarantee that the dataset instances are distributed evenly and, as a result, creates efficient classification models to foretell the likelihood that CKD would develop.A comparative evaluation of the performance of various models is presented taking into account the most popular metrics, such as Precision, Recall, F-Measure, Accuracy, and AUC. A features analysis with the following three distinct steps as sub steps: I statistical description of numerical attributes, II. measurement of order of importance using three different approaches, and III. recording of nominal features frequency of occurrence in tabular form.  
The performance of multiple models is compared and evaluated using the most used metrics, including Precision, Recall, F-Measure, Accuracy, and AUC.  
A performance evaluation is shown, with Rotation Forest attaining the top results across all metrics, serving as the major recommendation of this analysis. All models showed extraordinarily high results.

5.Machine learning techniques are useful for predicting CKD. In this work, a workflow for predicting CKD status based on clinical data is proposed. This workflow includes data prepossessing, a mechanism for handling missing values, collaborative filtering, and attribute selection. The additional tree classifier and random forest classifier are demonstrated to produce the highest accuracy and least amount of bias to the characteristics out of the 11 machine learning techniques taken into consideration. The study emphasizes the value of combining domain expertise when applying machine learning for CKD status prediction, as well as practical issues of data collecting.

6.Proposed a system that uses machine learning algorithms like ACO and SVM. The CKD Dataset from the UCI repository is the dataset that we are using in this instance. 400 samples from two separate classes are included. There are a total of 25 qualities: 11 are numerical, 13 are nominal, and 1 is a class trait. The dataset is used to test the prediction made by their model. SVM is used for classification in order to predict how the disease would perform. In this situation, ACO is a feature selection wrapper method. Use ACO and SVM to select the best features and reduce the remaining features to improve prediction accuracy. To test our model, it has 400 samples from two separate classes, which will greatly aid in predicting results with greater accuracy. To employ an ant colony algorithm, the optimization issue must be transformed into the problem of finding the shortest path on a weighted network. Each ant stochastically develops a solution, or the right order for following the edges of the graph. Because each ant must walk every path before delivering a result, the process will take some time.

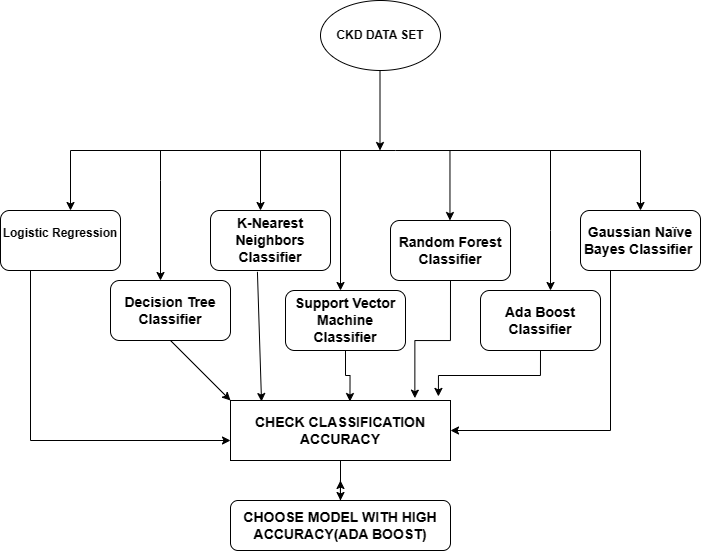
7.Long-term kidney disease The UCI Machine Learning repository's illness dataset, which has 25 characteristics, was examined using classifiers for logistic regression (LR), decision tree (DT), and support vector machines (SVM). The bagging ensemble approach was then used to enhance the results of the produced model. The clusters from the chronic renal illness dataset were used to train the machine learning classifiers. The clusters from the chronic renal illness dataset were used to train the machine learning classifiers. We get the greatest outcomes while using an accurate decision tree. Finally, we use the bagging ensemble method to attain the best accuracy. We improved the effectiveness of the fundamental classifiers using the bagging ensemble method, and we found that the decision tree had the highest accuracy (97.23%). This can help with the early detection of chronic renal illness, saving both patients' and medical professionals' lives.Their study was built upon the data extracted from UCI Machine learning repository with 25 features and only focusing GFR (glomerular filtration rate) but actually we have to focusing on every single attribute when we train the dataset.

8.Permanent dialysis or kidney transplants are common outcomes of CKD. High likelihood of CKD is also increased by a family history of renal disease.  
According to published research, CKD affects roughly one in three individuals with a diabetes diagnosis. The literature also offers proof that treating and diagnosing CKD early can enhance the patient's quality of life. Permanent dialysis or kidney transplants are common outcomes of CKD. The thorough study of the literature demonstrates how several machine learning algorithms are used to forecast CKD. . This study suggests the optimal prediction model for CKD and attempts to predict it using classifiers like Decision Tree, Random Forest, and Support Vector Machine. Researchers in engineering and medicine are working to create machine learning algorithms and models that can detect chronic kidney disease at an early stage. The issue is that the amount of data produced by the health sector is significant and complicated, which makes data analysis challenging. However, by applying data mining technology, we can convert this data into a format that can later be used by machine learning algorithms. To make clinical judgments concerning diagnosis and treatment, a kidney biopsy is necessary.

9.Chronic kidney disease with 25 characteristics, the disease dataset from the UCI Machine Learning repository was analyzed using logistic regression (LR), decision tree (DT), and support vector machine (SVM) classifiers. The findings of the generated model were then improved using the bagging ensemble approach. The machine learning classifiers were trained using the chronic renal disease dataset's clusters. In the case of a decision tree that is accurate, we achieve the best results. Finally, we achieve the highest accuracy after using the bagging ensemble method.

10.Early and accurate detection of the stages of CKD is thought to be crucial in order to reduce the effects of patient health complications like hypertension, anaemia (low blood count), mineral bone disorder, poor nutritional health, acid base abnormalities, and neurological complications with prompt intervention through appropriate medications. Using machine learning techniques, several research on the early detection of CKD have been undertaken. Predicting the precise stages was not their main focus. This work has performed both binary and multiple classifications for stage prediction. The prediction models used include Random Forest (RF), Support Vector Machine (SVM), and Decision Tree (DT). Analysis of variance and recursive feature removal with cross validation have both been used for feature selection. Early CKD identification is essential to minimize cost burden and optimize treatment effectiveness. Early CKD detection using predictive analysis using machine learning techniques can be beneficial for quick and effective interventions.

**III.METHODOLOGY**

It's possible for chronic renal disease to go undiagnosed until kidney function is severely compromised. Without mechanical filtering (dialysis) or a kidney transplant, end-stage renal failure brought on by chronic kidney disease is fatal. A chronic disease currently affects one in five people worldwide. Chronic illnesses limit everyday activities, demand continuing medical attention, or both. Changing one's lifestyle can make life easier for someone whose illness is predicted to appear sooner rather than later.

**Fig 1:** Flow diagram of CKD prediction using machine Learning Models

The most current advances in machine learning technology can be utilized to find hidden patterns that might help in the early diagnosis of chronic renal disease .This project offers a way for predicting chronic kidney disease utilizing unfavorable machine learning techniques.We implemented the following Classification Algorithms on our dataset:

1. Decision Tree Classifier

2. K-Nearest Neighbours Classifier

3. Support Vector Machine Classifier

4. Random Forest Classifier

5. Logistic Regression

6. Ada Boost Classifier

7. Gaussian Naïve Bayes Classifier

**IV.CHRONIC KIDNEY PREDICTION USING MACHINE LEARNING MODEL**

**DECISION TREE METHOD**

The most effective and well-liked categorization and prediction method is the decision tree. In a decision tree, each internal node stands for an attribute test, each branch for the test result, and each leaf node (terminal node) for the class label. Decision trees are a sort of tree structure that mimics flowcharts.

**K-NEAREST NEIGHBOUR METHOD**

K-Nearest Neighbour is one of the most fundamental supervised learning-based machine learning algorithms. The K-NN algorithm groups the new instance, assuming that it is equivalent to the prior instances, into the category that most closely resembles the present categories. A new data point is categorised using the K-NN algorithm based on similarity after storing all the prior data. This shows that the K-NN technique can quickly and accurately categorise new data. Although the K-NN technique is most frequently employed to solve classification problems, it can also be used to solve regression difficulties. K-NN uses no assumptions about the underlying data and is a non-parametric approach.

The method is also known as a lazy learner since it saves the training dataset rather than learning from it right away. As an alternative, it does a task while classifying data using the dataset. The KNN method simply saves the dataset during the training phase and then classifies new data into a category that is extremely similar to the new data.

**SUPPORT VECTOR MACHINES**

One of the most popular supervised learning algorithms is called the Support Vector Machine, or SVM, and it is used to solve Classification and Regression problems. However, it is primarily used in Machine Learning Classification issues. The goal of the SVM approach is to develop the optimal decision boundary or line that can swiftly classify fresh data points in n-dimensional space in the future. To assist in creating the hyperplane, SVM selects the extreme vectors and points. Support vectors, which are used to represent these extreme scenarios, are the basis of the SVM approach. The SVM method can be used for a variety of tasks, including face recognition, picture classification, text categorization, etc.

**RANDOM FOREST**

Unsupervised machine learning techniques like random forest are frequently employed in classification and regression issues. On a variety of samples, it builds decision trees and uses the average for classification and the majority vote for regression. The Random Forest Algorithm's ability to handle data sets comprising both continuous variables, as in regression, and categorical variables, as in classification, is one of its most important qualities. Regarding issues involving classification, it yields better outcomes. Random Forest functions according to the Bagging principle.

**LOGISTIC REGRESSION**

Another potent supervised machine learning approach utilized for binary classification issues is logistic regression (when target is categorical). Logistic regression is best understood as a linear regression that is used to solve classification problems. Using the logistic function shown below, logistic regression essentially models a binary output variable. Logistic regression's range is restricted to values between 0 and 1, which is the primary contrast between it and linear regression. Additionally, logistic regression does not require a linear relationship between the input and output variables, in contrast to linear regression. This is because the odds ratio underwent a nonlinear log change.

**NAÏVE BAYES**

The Bayesian probability theorem serves as the foundation of the Naive Bayes classification algorithm. The classifier functions under the fundamental Naive Bayes assumptions, which assume that each feature contributes equally and independently to the outcome, and that the presence or absence of a feature has no bearing on the presence or absence of any other feature. An instance is classified using Naive Bayes by calculating the likelihood that it belongs to each target class; the instance is then assumed to belong to the target class with the highest probability.

**ADA BOOST**

The question of whether a series of weak classifiers may be transformed into a strong classifier gave rise to the concept of boosting in machine learning. A learner who is weaker than random guessing is called a classifier or weak learner. This will hold true even if a huge number of weak classifiers are used, each of which is better than random. A straightforward threshold on a single feature is typically employed as a weak classifier. The feature belongs to the positive category if it exceeds the projected threshold; otherwise, it belongs to the negative category. AdaBoost stands for "Adaptive Boosting," a technique used to solve classification issues by turning weak learners or predictors into strong forecasters.

**Confusion Matrix**

A table known as a confusion matrix is frequently used to describe how a classification model, also known as a "classifier," performed on a set of test data for which the true values were known. It is a way to gauge how well a machine learning classification algorithm performs when the output can include two or more classes. The table contains four possible combinations of predicted and actual values. It is very helpful for determining accuracy, precision, recall, specificity, and, most crucially, AUC-ROC curves.

The table contains four possible combinations of predicted and actual values:

1. TN / True Negative: the case was negative and predicted negative
2. TP / True Positive: the case was positive and predicted positive
3. FN / False Negative: the case was positive but predicted negative
4. FP / False Positive: the case was negative but predicted positive

**Classification Metrices**

A classification algorithm's predictions are evaluated using classification metrics. How many of the forecasts came true, and how many were wrong? To be more precise, the metrics of a categorization report are predicted using True Positives, False Positives, True Negatives, and False Negatives.

The model's precision, recall, F1 and support scores are shown in the classification report visualizer.

* **Recall**

Recall is the ability of a classifier to find each successful occurrence. The ratio of true positives to the sum of true positives and false negatives for each class is how it is defined. Recall: Accurately identified positives as a percentage

**Recall = TP/(TP+FN)**

* **Precision**
* Precision is the ability of a classifier to prevent classifying anything that is actually negative as positive. The ratio of true positives to the sum of true positives and false positives is how it is defined for each class accuracy of confident predictions that are accurate.

**Precision = TP/(TP + FP)**

* **F1-score**

The F1 score, which ranges from 0.0 to 1.0 depending on recall and precision, is a weighted harmonic mean. F1 scores are lower than accuracy measurements because the 38 computation includes precision and recall. The weighted average of F1 is frequently suggested when comparing classifier models as opposed to overall accuracy.

**F1 Score = 2\*(Recall \* Precision) / (Recall + Precision)**

* **Support**

Support is the proportion of actual occurrences of the class in the dataset. Unbalanced support in the training data, which may imply structural issues with the classifier's reported scores, may suggest the need for stratified sampling or rebalancing. Support is the same across models, but diagnosis is on the evaluation process.

* **Macro-Average**

The macro-average is the mean overall class average for precision/recall/f1.

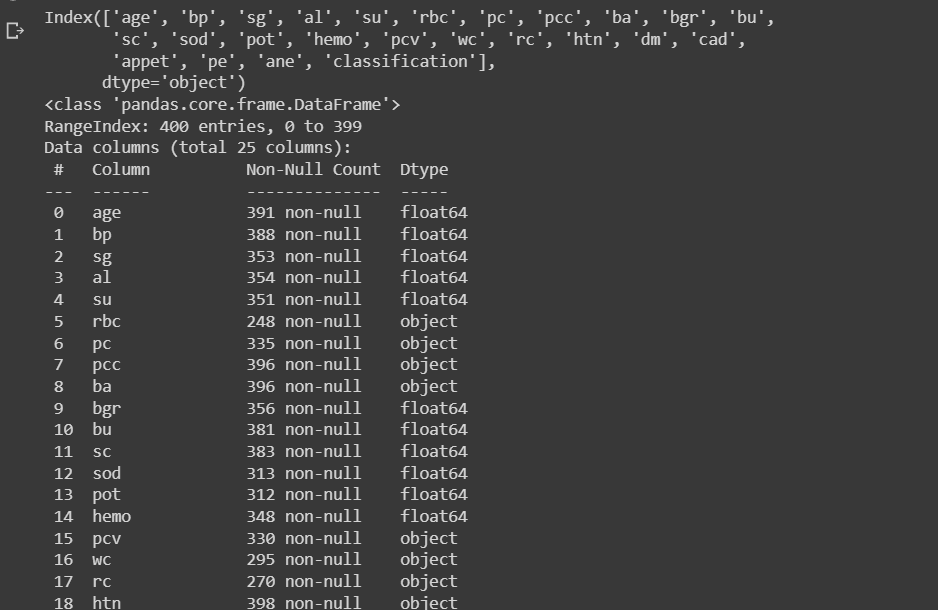
* **Weighted Average**

The macro-average is the mean overall class average for precision/recall/f1.

**V. DATA CLEANSING AND DATA VIRTUALIZTION/ANALYSIS**

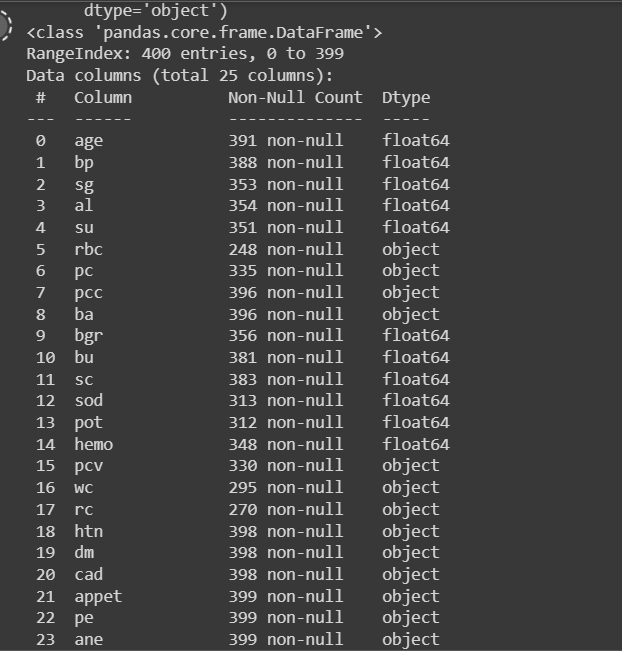
Below is the snapshot of the output which displays the dataset.

* **Deleting Unnecessary Column(s)**



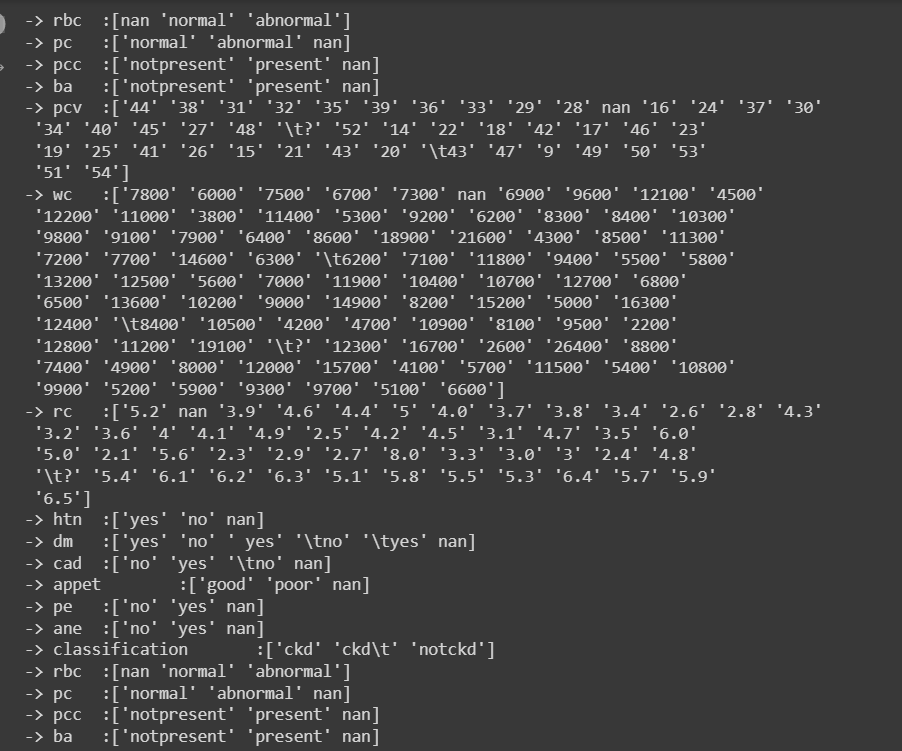
**Fig 2: Unnecessary Column Deleting**

* **Data Transformation**

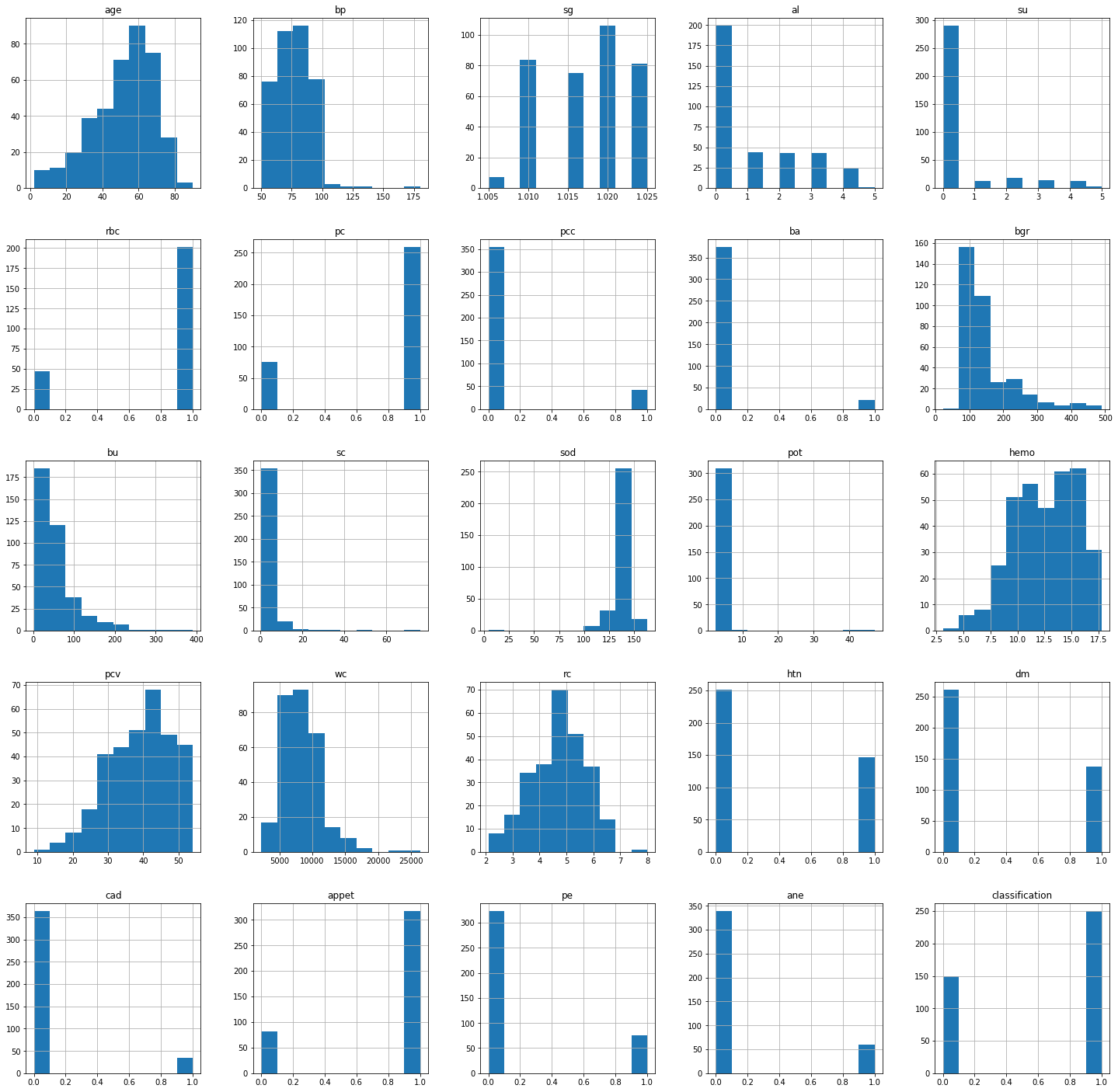
****

**Fig 3: Data Transformation**

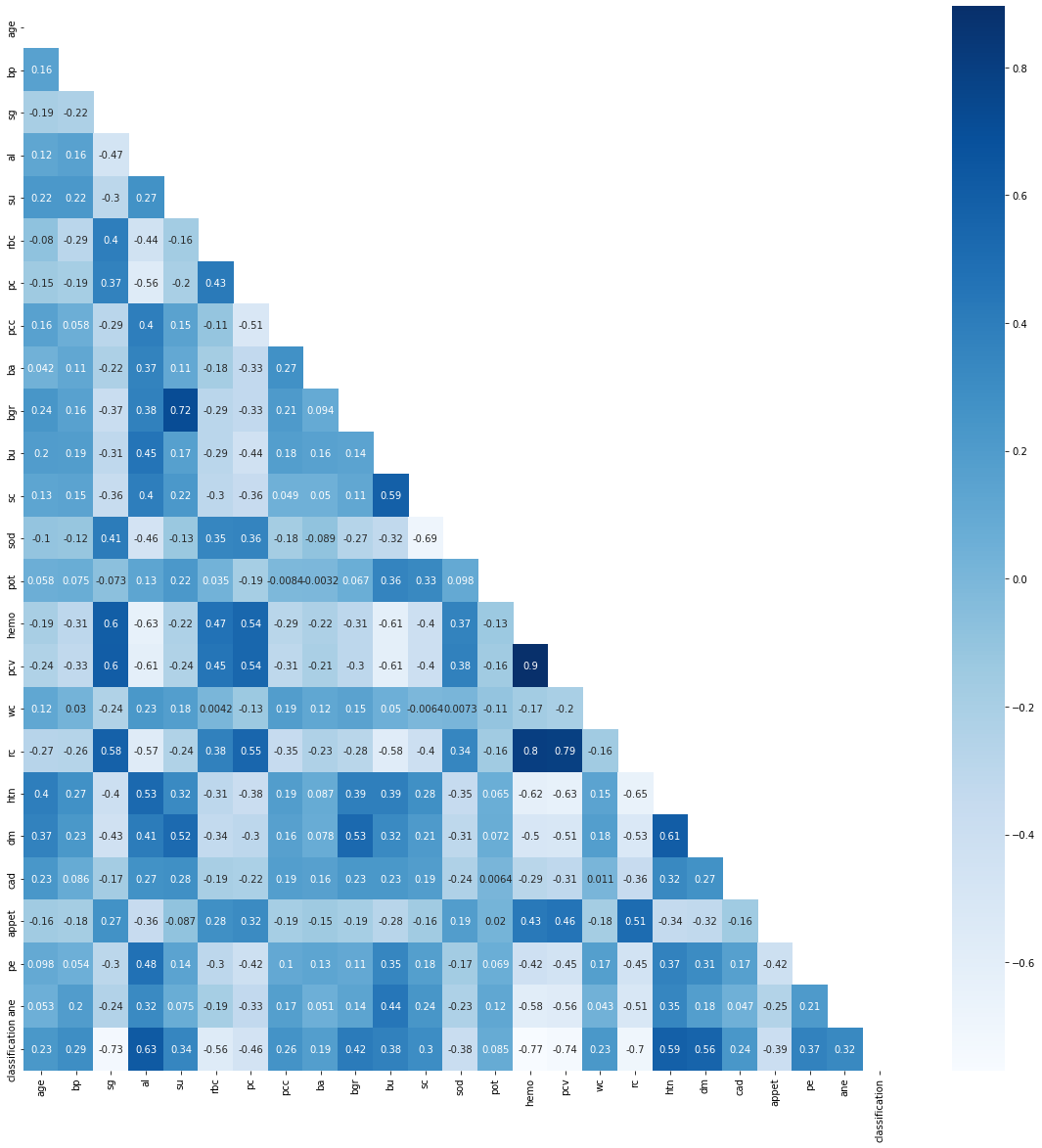
* **Transforming attributes of object type**

****

**Fig 3:Atrributes of object type**

* **Univariate Plot and Analysis**
* **Plotting Histogram**

**Fig 4: Plotting Histogram**

* **Multivariate Plot and Analysis**
* **Plotting Heatmap using Pearson’s Correlation**

**Fig 5: Heatmap as graphical representation of data using colours**

**VI. OBSERVATIONS AND RESULTS**

The 400 cases and 24 attributes with 2 classes that make up the Indians Chronic Kidney Disease (CKD) dataset were taken from the UCI machine learning repository. This dataset's attributes are split into two categories, numeric and nominal, with 11 numeric attributes and 14 nominal attributes. The patients at Apollo Hospitals for Indians provided the data for this dataset. The dataset is split into two groups: a training group and a testing group. Data used for testing and training are split 70/30, accordingly.

**Accuracy of Decision Tree:** Confusion Matrix was produced using a decision tree model for the test data (120 instances), with the target variable's class (values: CKD, NON-CKD) as shown in Table 1. According to the confusion matrix, 114 examples have been correctly identified whereas 6 occurrences are definitely misclassified. The accuracy of this classifier model is 95%.

**Accuracy of K-Nearest Neighbour Classifier:** Confusion Matrix was produced using a K-Nearest Neighbour Classifier model for the test data (120 instances), with the target variable's class (values: CKD, NON-CKD) as shown in Table 1. According to the confusion matrix, 83 examples have been correctly identified whereas 37 occurrences are definitely misclassified. The accuracy of this classifier model is 69%.

**Accuracy of Random Forest:** Confusion Matrix was produced using Random forest model for the test data (120 instances), with the target variable's class (values: CKD, NON-CKD) as shown in Table 1. According to the confusion matrix, 115 examples have been correctly identified whereas 5 occurrences are definitely misclassified. The accuracy of this classifier model is 95.8%.

**Accuracy of SVM:** Confusion Matrix was produced using SVM model for the test data (120 instances), with the target variable's class (values: CKD, NON-CKD) as shown in Table 1. According to the confusion matrix, 79 examples have been correctly identified whereas 41 occurrences are definitely misclassified. The accuracy of this classifier model is 66%.

**Accuracy of Ada Boot:** Confusion Matrix was produced using SVM model for the test data (120 instances), with the target variable's class (values: CKD, NON-CKD) as shown in Table 1. According to the confusion matrix, 118 examples have been correctly identified whereas 2 occurrences are definitely misclassified. The accuracy of this classifier model is 98.3%.

**Accuracy of Gaussian Naïve Bayes:** Confusion Matrix was produced using Gaussian Naïve Bayes model for the test data (120 instances), with the target variable's class (values: CKD, NON-CKD) as shown in Table 1. According to the confusion matrix, 114 examples have been correctly identified whereas 6 occurrences are definitely misclassified. The accuracy of this classifier model is 95%.

**Accuracy of Logistic Regression**: Confusion Matrix was produced using Logistic Regression model for the test data (120 instances), with the target variable's class (values: CKD, NON-CKD) as shown in Table 1. According to the confusion matrix, 117 examples have been correctly identified whereas 3 occurrences are definitely misclassified. The accuracy of this classifier model is 97.5%.

**VII. CONCLUSION**

Our project, "Prediction of Chronic Kidney Disease using Machine Learning Methods," has demonstrated good classification performance using machine learning algorithms. The usage of machine learning techniques is therefore not as widespread as one might anticipate. Inconsistencies in the diagnosis of chronic kidney disease, we contend, can at least largely be blamed for our model. If a person has chronic kidney disease, experimental results can help them control it so that it doesn't progress to other stages, which will in turn help them reduce the number of kidney failures caused by CKD. Early detection of chronic kidney disease can help healthcare professionals protect a person from it. This project's major objective was to Create and implement a model that will use machine learning techniques on a pertinent dataset to predict whether a patient has chronic kidney disease or not. **Feature selection** is a machine learning technique that we used to determine the top 10 features that have the biggest impact on our decision out of the total of 24 attributes:

1. Count of White Blood Cells

2. Random Blood Glucose

3. Third Blood Urea

4.Serum Creatinine

5. Volume of a Packed Cell

6. Albumin

7. Blood oxygen

8. Age

9. Sugar

10.Hypertension

According to the dataset and our research, these 10 attributes are the most crucial variables among the 24 features for the prediction of chronic kidney disease. We implemented seven classifications on our website because our project was classification-based. So, using our dataset, we applied seven classifications. The Decision Tree Classifier, K-Nearest Neighbour, Support Vector, Ada Boost, Gaussian Naive Bayes, Random Forest, and Logistics Regression Classifier algorithms were all implemented. The maximum accuracy of these seven algorithms, 97% or thereabouts, is achieved by the Random Forest Classifier, Ada Boost, and Decision Tree. With accuracy of 78%, 74%, 97%, 93%, 97%, 97%, and 95%, respectively, the KNN, SVM, Ada Boost, GNB, Decision Tree, Random Forest, and Logistics Regression are the most accurate. Accuracy, F1-score, Precision, and Recall are the many performance metrics being compared.

**VIII. FUTURE ENHANCEMENT**

The aim of this project was to build a model which would predict whether a patient would develop Chronic Kidney Disease or not based on his/her medical data. We did the same by implementing machine learning algorithms on the dataset. We implemented more than one algorithm on the dataset to find the best suitable algorithm based on the accuracy score.

**Acknowledgements:** For this project We would like to thank VIT for providing us the opportunity to implement this project. It was an honour to do this project under the guidance of **Dr. ABDUL GAFFAR H Sir Department of Computer Science, VIT, VELLORE** .He has helped us in each and every aspect to complete the project successfully. we will always be grateful for getting this opportunity as it has helped me to boost my knowledge and gather in depth analysis of real-world technology.

**IX. REFERENCES**

1. Almasoud, Marwa & Ward, Tomas. (2019). Detection of Chronic Kidney Disease using Machine Learning Algorithms with Least Number of Predictors. International Journal of Advanced Computer Science and Applications. 10. 10.14569/IJACSA.2019.0100813.
2. Ifraz, G. M., Rashid, M. H., Tazin, T., Bourouis, S., & Khan, M. M. (2021, December 3). *Comparative analysis for prediction of kidney disease using intelligent machine learning methods*. Computational and Mathematical Methods in Medicine. Retrieved January 18, 2023, from https://www.hindawi.com/journals/cmmm/2021/6141470/
3. Ilyas, H., Ali, S., Ponum, M., Hasan, O., Mahmood, M. T., Iftikhar, M., & Malik, M. H. (2021, August 9). *Chronic kidney disease diagnosis using decision tree algorithms - BMC Nephrology*. BioMed Central. Retrieved January 18, 2023, from https://bmcnephrol.biomedcentral.com/articles/10.1186/s12882-021-02474-z
4. Dritsas, E., & Trigka, M. (2022, September 14). *Machine learning techniques for chronic kidney disease risk prediction*. MDPI. Retrieved January 20, 2023, from https://www.mdpi.com/2504-2289/6/3/98
5. Ekanayake, I. U., & Herath, D. (2020, September 20). *Chronic kidney disease prediction using machine learning methods*. Retrieved January 20, 2023, from https://www.researchgate.net/profile/Imesh-Ekanayake/publication/344319206\_Chronic\_Kidney\_Disease\_Prediction\_Using\_Machine\_Learning\_Methods/links/5f672571458515b7cf418d5b/Chronic-Kidney-Disease-Prediction-Using-Machine-Learning-Methods.pdf
6. S, R., Shaji, S., Ajina, S. R., R, V. P. S., & A, J. (2020, July 9). *Chronic kidney disease prediction using machine learning*. International Journal of Engineering Research & Technology. Retrieved January 20, 2023, from https://www.ijert.org/chronic-kidney-disease-prediction-using-machine-learning
7. Chaurasia, V., Pal, S., & Tiwari, B. B. (2019, January 7). *Chronic kidney disease: A predictive model using decision tree*. SSRN. Retrieved January 20, 2023, from https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=3298343
8. Ramesh, Revathy. (2020). Chronic Kidney Disease Prediction using Machine Learning Models. International Journal of Engineering and Advanced Technology. 9. 6364. 10.35940/ijeat.A2213.109119.
9. Chaurasia, V., Pal, S., & Tiwari, B. B. (2019, January 7). *Chronic kidney disease: A predictive model using decision tree*. SSRN. Retrieved January 20, 2023, from https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=3298343
10. Debal, D. A., & Sitote, T. M. (2022, November 20). *Chronic kidney disease prediction using machine learning techniques - journal of big data*. SpringerOpen. Retrieved January 20, 2023, from https://journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00657-5