[1] **Varsha Bansa ***

[2] **Dr. Ritu Sindhu**

A Comparative Study on Chronic Kidney Disease using Different Machine Learning Techniques

Journal of Electrical Systems

Abstract: - CKD (Chronic Kidney Disease) has become increasingly prevalent. Those with chronic kidney disease can suffer from the disease for the rest of their lives. There are two causes of this condition: kidney malignancy and reduced kidney function. It is possible to prevent patients from progressing to an end-stage of this disease that requires dialysis or surgery in the early stages. The likelihood of this happening can be increased if the disease is detected early and treated appropriately. Several machine learning approaches have been evaluated in this research for early detection of CKD. The topic of this research has received a great deal of attention. While this is the case, we are utilizing predictive modeling to enhance our approach. Through machine learning and predictive analytics, we can develop a collection of prediction models based on better measures of attributes. A supervised learning environment has been used to test different machine learning-based classifiers. As a result of the research, we can conclude that recent advances in machine learning, coupled with predictive analytics, could yield new treatments for kidney disease and other conditions.

Keywords: Chronic Kidney Disease (CKD), Data Pre-processing, Machine Learning Technique, Classification algorithms, Random Forest Classifier.

I. INTRODUCTION

It plays an important role in maintaining the body's balance by eliminating metabolic waste products from the bloodstream. Every aspect of a disease should be approached pragmatically, including willingness to undergo clinical therapy. When there are no symptoms present, it can lead to a wide variety of behavioral patterns. Among its many functions, the kidney produces red blood cells, vitamin D in an active form for the body, and more.

In addition to high blood pressure, diabetes, and cardiovascular disease (CVD), chronic kidney disease (CKD) is also associated with high blood sugar levels. Due to the high mortality rate associated with CKD, it has received a lot of attention.[1] .

A kidney disease occurs when the kidneys cannot filter blood and perform other essential functions. Over time, kidney cells degenerate steadily, which is known as chronic kidney disease. Cardiovascular disease can develop as a result of CKD, which is irreversible.[2]. A permanent dialysis or kidney transplant is often required to treat this condition. CKD can be treated and diagnosed early, enhancing the quality of life of the patient. To slow the progression of CKD, early detection and diagnosis are crucial so that treatment can begin as soon as possible.

In the past few decades, machine learning has been used to predict and classify a wide variety of diseases, including heart disease, breast cancer, kidney disease, and stroke. In addition to healthcare, this technology is used in other sectors such as sustainable energy. Computer algorithms are used to find patterns in enormous datasets, and EMRs, which use big data, are increasingly using machine learning. [3]. ML prediction algorithms provide early treatment for many diseases through the use of intelligent algorithms. In other words, CKD diagnosis may be possible using this method. Using clinical datasets of 400 patients, we apply nine machine learning algorithms and compare their corresponding results to detect kidney disease. With the use of ML techniques, this study preprocesses and predicts Chronic Kidney Disease, contributing to the identification of relevant features in raw data. [4]. By conducting this study, we will be able to identify and treat the risk factors of these kidney diseases in a timely and accurate manner. The results presented in this paper are also compared to those presented in previous publications.

An organ found in vertebrates is the kidney, which is shaped like a bean and is rusty red in color. There are about 12 cm (4 1/2 inches) long tubes found in the left and right retroperitoneal areas of an adult human body.[5]. The

[1] Ph.D. Scholar, Department of Computer Science and Engineering, Lingaya's Vidyapeeth, Faridabad, Haryana, India
bvarsha52@gmail.com

[2] Professor ,Department of Computer Science and Engineering, Lingaya's Vidyapeeth, Faridabad, Haryana, India,
ritu.sindhu2628@gmail.com

* Corresponding Author Email: bvarsha52@gmail.com

kidney contains almost one million nephrons, which carry out essential functions. The kidney's nephrons filter urine as a byproduct of the filtration process. Each 24 hours, the kidneys filter and return about 200 quarts of fluid to the body. It is estimated that around 198 pints are lost perspiration, while two points are gained. It takes between one and eight hours for urine to be released from the bladder. [6].

The Global Kidney Disease Association estimates that one in four people suffer from this disease, which results in a slow decline in kidney function. The number of people who need dialysis or kidney transplants around the world is around two million, but only about ten percent of these people get them. Cancer of the breast and prostate are not the only causes of death among individuals with chronic kidney disease every year. [7].

There are five phases of chronic kidney disease based on glomerular filtration rate (eGFR) and creatinine levels. In addition to serum creatinine levels, glomerular filtration rates (eGFR) can also be calculated based on equations tailored to the patient. By using serum creatinine as a measurement, it is possible to express GFR physiologically. Glomerular function loss of 50% is necessary before serum creatinine levels become elevated; the serum creatinine level is insensitive to glomerular function. Creatinine-derived eGFR has increased relevance when compared with serum creatinine.[8]. It is possible to divide kidney function into five distinct levels. Having an eGFR greater than 90 when chronic kidney disease is in stage 1 will cause minimal damage to the kidneys. With an eGFR between 60 and 89, Stage 2 Chronic Kidney Disease is characterized by moderate renal function loss. A mild to moderate loss of renal function is indicated by an eGFR (estimated glomerular filtration rate) of 30-59 in Chronic Kidney Disease and 3b. [9]. Stage 4 Chronic Kidney Disease is characterized by extreme kidney damage with an estimated glomerular filtration rate of 15-19. Symptoms of Stage 5 Chronic Kidney Disease include kidney failure or kidney damage to the point of failure (Fig-1).

	Low Risk			ACR Categories(mg/mmol), Description and Range)		
	Moderately Increased Risk			A1	A2	A3
	High Risk			Normal to Mildly Increased	Moderately Increased	Severely Increased
	Very High Risk			<30mg/g <3 mg/mmol	30-299 mg/g 3-29 mg/mmol	>=300mg/g >=30 mg/mmol
	Highest Risk					
GFR Stages	G1	Normal or High	>=90			
	G2	Mildly Decreased	60-90			
	G3a	Mildly to Moderately Decreased	45-59			
	G3b	Moderately to Severely Decreased	30-44			
	G4	Severely Decreased	15-29			
	G5	Kidney Failure	<15			

Fig 1: Heat Map of Severity

The phases and prevalence of Chronic Kidney Disease may be predicted using machine learning. Machine learning predictive models require high-quality training data. A computer algorithm can extract characteristics relevant to a research goal from this data in order to uncover significant relationships. Machine learning methods have been investigated for the prediction of Chronic Kidney Disease using several data sets. [10].There are a number of these datasets, including the UCI repository dataset (hereafter referred to as the UCI dataset). In the vast majority of similar studies, the benchmark dataset aforementioned has been used. Chronic Kidney Disease clinical data analysis should consider how predictable the loss of missing characteristics is before selecting a

missing values handling approach. A further advantage of the UCI dataset is the fact that it includes 400 examples with 25 properties, which are smaller sample sizes. Possibly, some of the characteristics in the data set are redundant (closely related) and others are missing.
Based on serum creatinine levels, glomerular filtration rate (eGFR) is estimated using a patient-specific algorithm. Physiological explanations of GFR estimates can be provided by using this method. In order for serum creatinine levels to increase, glomerular function must decline, which is why serum creatinine levels are considered insensitive indicators of glomerular function.[11]. However, serum creatinine is less relevant than eGFR based on creatinine.

Decision Tree:

In addition to its generality, Decision Tree Analysis can be applied to a variety of fields. Data subsets are commonly determined using an algorithm when constructing a decision tree. In terms of supervised learning, it is one of the most popular and useful approaches. In classification and regression applications, decision trees can be used as nonparametric supervised learning approaches. To predict the value of the target variable, basic decision rules will be used to build models.[12]. If-then statements are commonly used in the decision-making process. When the tree structure is more complex, the rules are more complex and the model is better fitted.

Naive Bayes (NB):

Classifiers can be developed based on Naive Bayes in an easy-to-follow manner. In this probabilistic classifier, Bayes' theorem is used as the basis. A Naive Bayes classifier treats each feature value independently based on the class variable.
The Bayes theorem can be represented as follows:

P(C|X)=P(X|C)*P(C)/P(X)

C is a class where P(X) is always the same for X and X is a tuple of data. Even when attribute values are assumed to be conditionally independent, this method produces accurate results.

Random Forest:

As a kind of nearest neighbor prediction, Random Forests are useful in regression and classification. The final product is a merged decision tree derived from many decision trees, which is merged to generate more reliable predictions. This program creates a forest of decision trees as part of its training process. Trees produce this mode of classification, and it is this mode of classification that is called a class. Averaging minimizes both extremes of variation and bias by locating a sweet spot between them. Both Python and R have libraries that support this approach.

K-Nearest Neighbour:

Because of its simplicity and effectiveness, the K-Nearest Neighbor method is widely used in classification. Lazy learning algorithms include K-Nearest Neighbor. We predict the classification of a new point based on the classification of points in an existing dataset. K-Nearest Neighbor classifies new data based on similarities with existing data in order to assign labels. According to popular vote, it is categorized among its neighbors. The closest neighbors are taken into account when determining which category should receive the data. By increasing the number of nearest neighbors, or k, precision can be improved. It is not necessary to know the distribution of the data in order to analyze it using non-parametric methods. It is possible to infer a model's structure based on its data. The fact that the majority of "real life" facts don't match the standard theoretical assumptions makes this a very important tool. To complete a classification study, K-Nearest Neighbors should be used to examine additional background information.

Logistic Regression:

A supervised learning algorithm called logistic regression forms part of the category of Machine Learning. This method is capable of predicting categorical dependent variables from a collection of independent variables. An outcome may be predicted by Logistic Regression using a categorical dependent variable. In other words, the result must be categorical or discrete in nature. It might be true or false, 0 or 1, 0 or 1, etc, depending on the probability values between 0 & 1. Linear Regression differs from Logistic Regression primarily in its application.

In Regression problems, Linear Regression is the most efficient method, whereas Logistic Regression is most efficient in Classification problems.

By applying nine machine learning algorithms, we are able to determine kidney disease in 400 patients by analyzing their corresponding clinical datasets. Using machine learning techniques, the study identified relevant features from the raw dataset and predicted Chronic Kidney Disease using those features. As a result of this study, any appropriate and safe diagnosis of CKD could be promptly and accurately treated for the risks identified. A comparison is also made between the results presented in this paper and those previously published.

II. Literature Review

Researchers have already used algorithms to compute and analyze data to improve people's lives through meaningful conclusions. There are many industries that benefit from experiments utilizing machine learning algorithms, including those in business, medicine, science, and mathematics.

Using K/DOQI guidelines, the authors discussed how chronic kidney disease patients and dialysis patients can reduce illness and death rates. Besides being helpful for identifying and treating a variety of anemias, renal osteodystrophy, uremic malnutrition, hyperlipidemia, & cardiovascular disease, these Guidelines can be helpful for treating anemias, renal osteodystrophy, uremic malnutrition, & hyperlipidemia.[13]. The guidelines are not the only way doctors and nephrologists improve the lives of patients suffering from chronic renal disease.

According to S.K. Agarwal et al. A significant portion of the Indian population suffers from chronic kidney disease. [14] In the future, the Indian government and Ministry of Health and Family Welfare will benefit from investing in prevention and health care programs for Chronic Kidney Disease.

According to Teresa K. Chen et al., chronic kidney disease incidence rates, risk factors, and preventative measures are discussed. [15] It is essential for Chronic Kidney Disease patients to receive primary care from doctors who struggle with keeping up with the staggering demands of the condition, in addition to being closely monitored for complications like hyperkalaemia, metabolic acidosis, anemia, and other metabolic abnormalities. The disease is also responsible for 8-16% of all deaths worldwide due to the fact that it affects 8-16% of the global population. According to WaadAllah S. Mula-Abed et al., CKD is more effectively treated when detected early by eGFR. Understanding the effects of Chronic Kidney Disease on kidney function so quickly.

According to a study published by Jing Xiao et al., there are nine models (LR, Lasso, EN, XGBoost, RIDGE, SVM, RF, NN, and K-Nearest Neighbor) that can be used to predict Chronic Kidney Disease based on readily accessible clinical signs and symptoms. [16] In addition, online tools were developed for outpatients to monitor their proteinuria progress.

According to Almasoud, M et al., the forecast of CKD was 99.0% accurate. A hemoglobin level, albumin level, and specific gravity are considered to be the three most important indicators of Chronic Kidney Disease, according to the researchers. [17].

Murale C et al. predicted Chronic Kidney Disease with 100% accuracy using Logistic regression, K-NN classifiers, and Random Forest algorithms. [18] The severity of the disease was determined based on their proposed work. In order to calculate eGFR, the gender and race of a person are necessary.

According to Imesh Udara Ekanayake et al., Chronic Kidney Disease can be predicted with 100% accuracy. [19] The features that were considered as 100% accurate were hemoglobin, albumin, serum creatinine, hypertension & diabetes mellitus. A total of 11 models were used to provide 100% accuracy, including decision trees, random forest classifiers, extra trees, and ADA boost classifiers.

Aruna, O., and Sk Sameerunnisa et al. found that random forest classifiers provided the greatest accuracy for predicting Chronic Kidney Disease by using real-time data and UCT datasets. A neural network and support vector machines are used to predict chronic kidney disease.[20]. Two suggested methods have been developed based on optimally acquired constraints and characteristics. An Artificial Neural Network had a higher accuracy of 99.75% than a Support Vector Machine, with an accuracy of 99% compared to 99.75% for an SVM.

An analysis of clusters has been developed for predicting CKD by S.Gopika and co-workers. By using clustering techniques, the study aims to identify kidney function failure. Based on the results of the trial, fuzzy C performed better and was 89% accurate. It was Almasoud and Ward who analyzed the CKD dataset. Four hundred and twenty-five instances were analyzed. In the CKD dataset, the feature attributes are hemoglobin, albumin, and specific gravity, which were selected using the filter feature selection approach. The dataset was analyzed and cross-validated ten times after the features were selected. Gradient boosting was the algorithm that achieved the highest accuracy, 99.1%.

Medical laboratories, research facilities, and hospitals provided Vijayarani and Dhayanand with kidney function test (KFT) datasets. Support vector machines (SVMs) and artificial neural networks (ANNs) were used to classify the dataset. The dataset consisted of 584 occurrences with six attributes. The highest level of accuracy was reached by ANN, with 87.7%.

Using data mining and machine learning techniques, Sujata Drall, et al. were able to accurately predict sick people's CKD states. [21]. The five available CKD features, along with the KNN and Nave Bayes algorithms, were used in this study to predict CKD in patients. It had a 100% accuracy rate in predicting CKD with KNN, while 96.25% accuracy with Nave Bayes Classifier.

An algorithm for predicting chronic kidney disease was developed using an aging dataset of chronic kidney disease. In the dataset, 24 attributes were present and one target variable was present. supervised machine learning algorithms such as KNN and Nave Bayes were used in the development of the model. The accuracy levels for KNN and Naive Bayes were 91% and 97%, respectively.

A cognitive model for categorizing CKD using neural networks was created by Farjana, Afia, Fatema Tabassum Liza et al. [22]. An early detection of CKD was achieved with the help of a generalized feedforward neural network (GRNN), a backpropagation neural network (BPN), and a modular neural network (MNN). The authors combined the GA with the above-mentioned models in their hybrid models.

It has been proposed that real-time datasets from Khulna City Medical College can be used to predict chronic kidney disease risk in an article published in the UCI Machine Learning Repository. We used a 10-fold cross-validation procedure for training and testing the RF and ANN. With 97.12% and 94.5% accuracy, the RF and ANN are both effective.

Ensemble learning was performed using AdaBoost ensembles, and feature selection was performed using correlation-based feature selections (CFSs). Furthermore, the researchers' system had the best accuracy, at 98.1%, among KNN, CFS, and AdaBoost. Moreover, UCI's machine learning repository was used 400 times by the researchers.

III. METHODOLOGY

In order to predict kidney-related disorders, medical professionals must develop a program that takes factors such as age, gender, blood pressure, and more into consideration. For kidney disease to be effectively treated, doctors should use a decision system to help them make accurate diagnoses and provide patients with information about their kidney health. It improves accuracy as well as speed in predicting renal illness using the suggested technology.

Decision trees, random forests, Naive Bayes, and K-Nearest Neighbors are just some of the machine learning algorithms that can make reliable predictions. A research workflow takes place in which data is collected, text is replaced, attributes with more than 20% missing values are removed, and missing values are filled in with statistical methods. In selecting features, statistical methods and medical importance were taken into account first. The next step involved training the models and tuning their hyperparameters, and finally, selecting the most accurate model. (Fig 2)

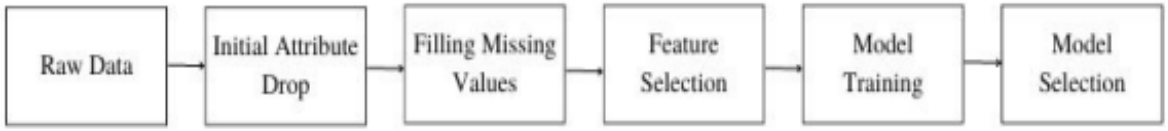

Fig 2: Proposed Workflow

In the design phase of machine learning, datasets are preprocessed, divided into training and test datasets, and missing values are filled in using methods. After analyzing the test database and choosing the best method, the final step is to implement a test system. In the following diagram, you can see the many stages involved in the design process. In the first step, 400 records were divided into training and test datasets, then the algorithm was applied to them.

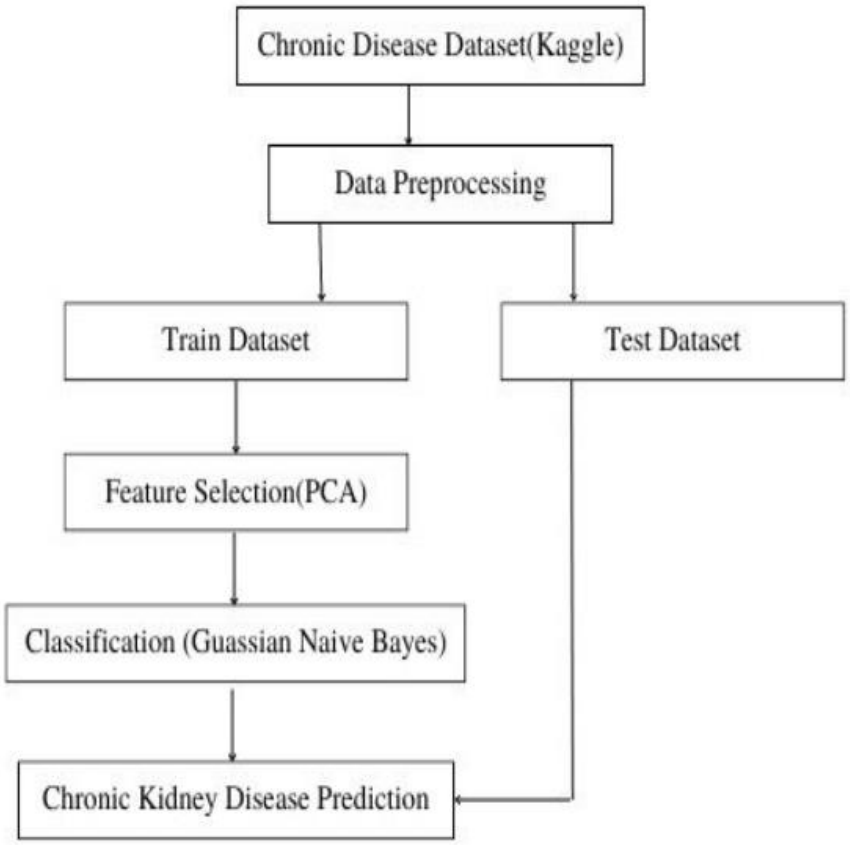

Fig 3: Modeling the System

A number of strategies have been assessed using the UCI dataset in this study for dealing with missing values. A dataset from UCI's repository was used to train and test our system for predicting Chronic Kidney Disease. (Fig 4)

0 age
1 blood_pressure
2 specific_gravity
3 albumin
4 sugar
5 red_blood_cells
6 pus_cell
7 pus_cell_clumps
8 bacteria
9 blood_glucose_random
10 blood_urea
11 serum_creatinine
12 sodium
13 potassium
14 hemoglobin
15 packed_cell_volume
16 white_blood_cell_count
17 red_blood_cell_count
18 hypertension
19 diabetes_mellitus
20 coronary_artery_disease
21 appetite
22 peda_edema
23 anemia
24 class

	id	age	bp	sg	al	su	rbc	pc	pcc	ba	...	pcv	wc	rc	htn	dm	cad	appet	pe	ane	cl
0	0	48.0	80.0	1.020	1.0	0.0	NaN	normal	notpresent	notpresent	...	44	7800	5.2	yes	yes	no	good	no	no	
1	1	7.0	50.0	1.020	4.0	0.0	NaN	normal	notpresent	notpresent	...	38	6000	NaN	no	no	no	good	no	no	
2	2	62.0	80.0	1.010	2.0	3.0	normal	normal	notpresent	notpresent	...	31	7500	NaN	no	yes	no	poor	no	yes	
3	3	48.0	70.0	1.005	4.0	0.0	normal	abnormal	present	notpresent	...	32	6700	3.9	yes	no	no	poor	yes	yes	
4	4	51.0	80.0	1.010	2.0	0.0	normal	normal	notpresent	notpresent	...	35	7300	4.6	no	no	no	good	no	no	

Fig 4: Chronic Kidney Disease(CKD) Dataset

The suggested methodology involves three primary phases: data preparation, data modeling, and model selection (Fig- 2).

IV. DATA PRE-PROCESSING

Missing Value Handling:

In this study, the data were prepared in two ways. The first step was to eliminate characteristics with missing data greater than 20%, as shown in Fig 3. Due to this, we only consider a small number of factors, such as sodium, potassium, WBC count, and red blood cell count. Afterwards, the remaining blanks were filled in during the second phase of preprocessing. The preprocessing phase should manage missing values based on their distributions to achieve adequate precision. A mean statistical method is used to replace missing values, which refers to averaging out the feature columns.

Filling in missing data is one of the pre-processing methods. Because '? ' is a non-null value, it must be transformed into NaN for inclusion in the dataset. Data that is missing will be identified by Pandas. In Fig-1, the missing data is depicted as a heatmap. These graphs do not display any characteristics. Filling in blanks was done using the median of characteristics. This is what our heat map looked like after we included the missing data.

As shown below, each characteristic has a proportion of missing values.

id 0.00
age 2.25
bp 3.00
sg 11.75
al 11.50
su 12.25
rbc 38.00
pc 16.25
pcc 1.00
ba 1.00
bgr 11.00
bu 4.75
sc 4.25
sod 21.75
pot 22.00
hemo 13.00
pcv 17.50
wc 26.25
rc 32.50
htn 0.50
dm 0.50
cad 0.50
appet 0.25
pe 0.25

ane 0.25

classification 0.00

Feature Selection :

Based on the characteristics of heat map to the class label (Fig- 6), albumin, sugar, blood pressure, hemoglobin, pus cell, and age had the highest absolute correlations. More than 0.3 correlation is observed between packed cell volume and secondary attributes, including blood glucose random and serum creatinine. On a comparison of absolute values of different tree classifications for class label selection, specific gravity, hypertension, packed cell volume, diabetes mellitus, hemoglobin, and albumin all score over 0.5.

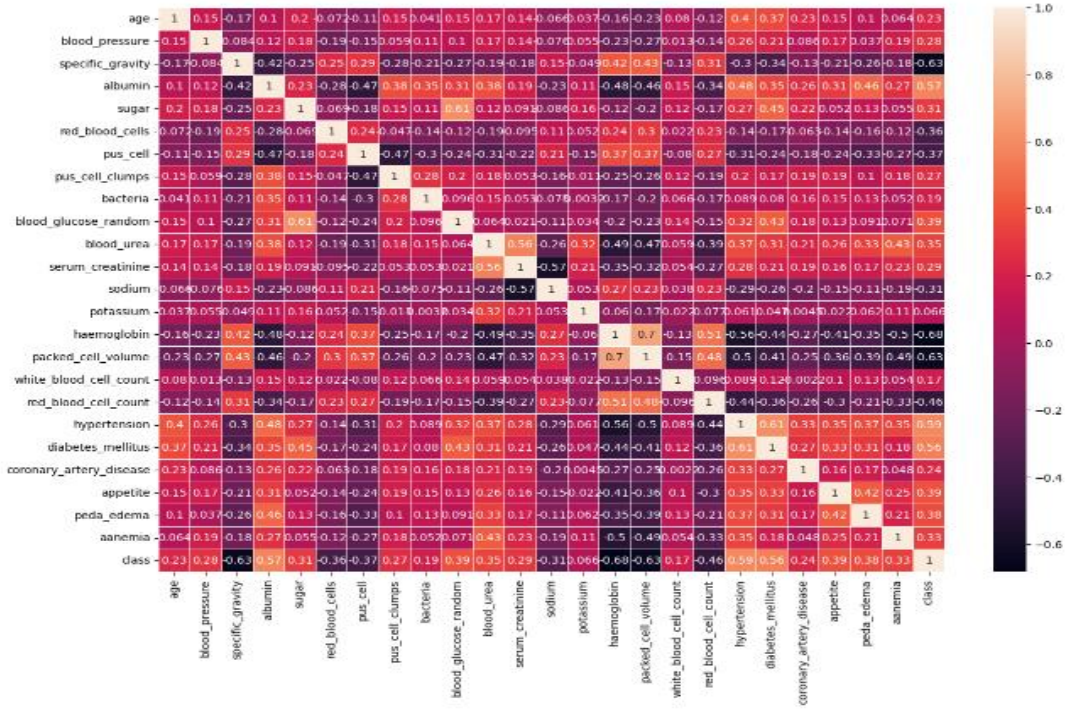

Fig 5: Heat Map of co-relation of attributing with class variable

Figure 5 shows the heat Map of co-relation of attributing with class variable

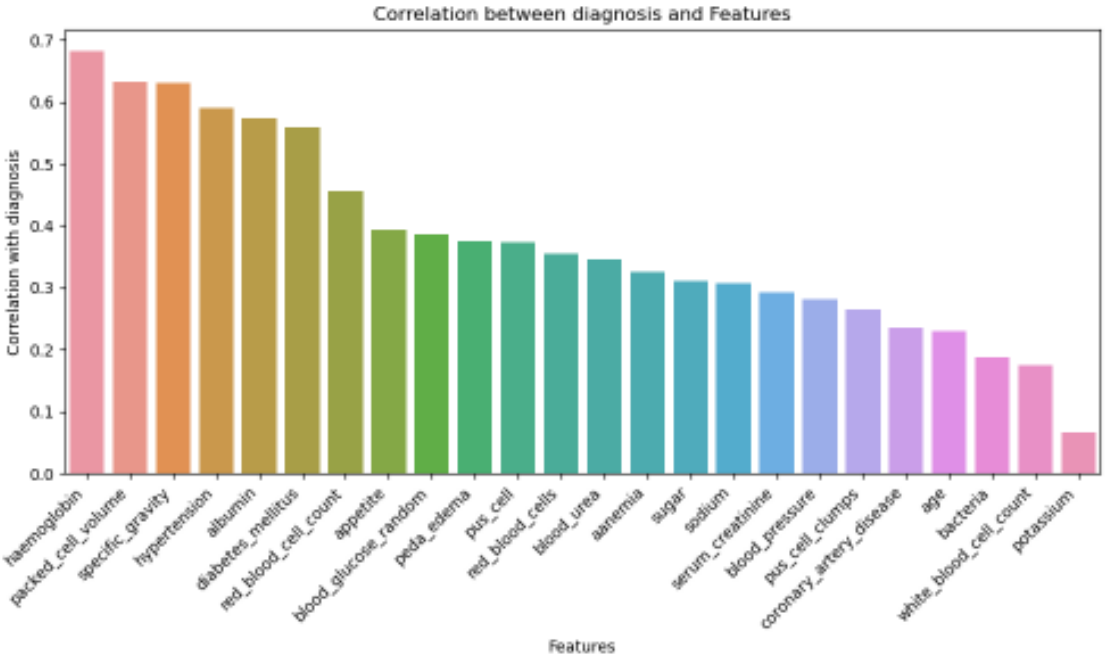

Fig 6: Extra tree classifier for feature selection to the class variable

The Figure 6 shows the Extra tree classifier for feature selection to the class variable **.** Using the distribution of characteristics values and the medical viewpoint of characteristics, we can predict chronic kidney disease using specific severity, hemoglobin, hypotension, type 2 diabetes, packed cell volume albumin, creatinine levels, and random blood glucose levels.

Model Training:

The learning process was investigated through nine different categorization models. K-Nearest Neighbor regressions, linear kernel SVCs, Gaussian NBs, decision tree classifiers, random forest classifiers, ADAA boost & gradient boost classifiers, etc., are among the techniques used. We randomly divided the data set into three halves, with 70% used for training and 30% for testing.

A total of nine algorithms were considered, but only four performed well on both testing and training. We implemented and assessed the algorithms using the Python Sci-kit & Keras libraries.

Table 1: Accuracies , Precision, Recall and F1-Score of each Algorithm

Algorithm	Training Accuracy	Testing Accuracy	Precision	Recall	F1-Score
K-Nearest Neighbour	77.14%	71.66%	0.81	0.69	0.75
Decision Tree Classifier	98.21%	95%	0.93	0.99	0.96
Random Forest Classifier	100%	97.50%	0.96	1.00	0.98
Ada Boost Classifier	100%	97.50%	0.96	1.00	0.98
Gradient Boosting Classifier	100%	98.30%	0.97	1.00	0.99
Guassian NB	100%	96.66%	0.95	1.00	0.97
XgBoost	99.99%	97.50%	0.96	1.00	0.98
Logistic Regression	99.92%	99.06%	1.00	0.97	0.98
SVC Linear	88.33%	88.33%	0.99	0.95	0.97

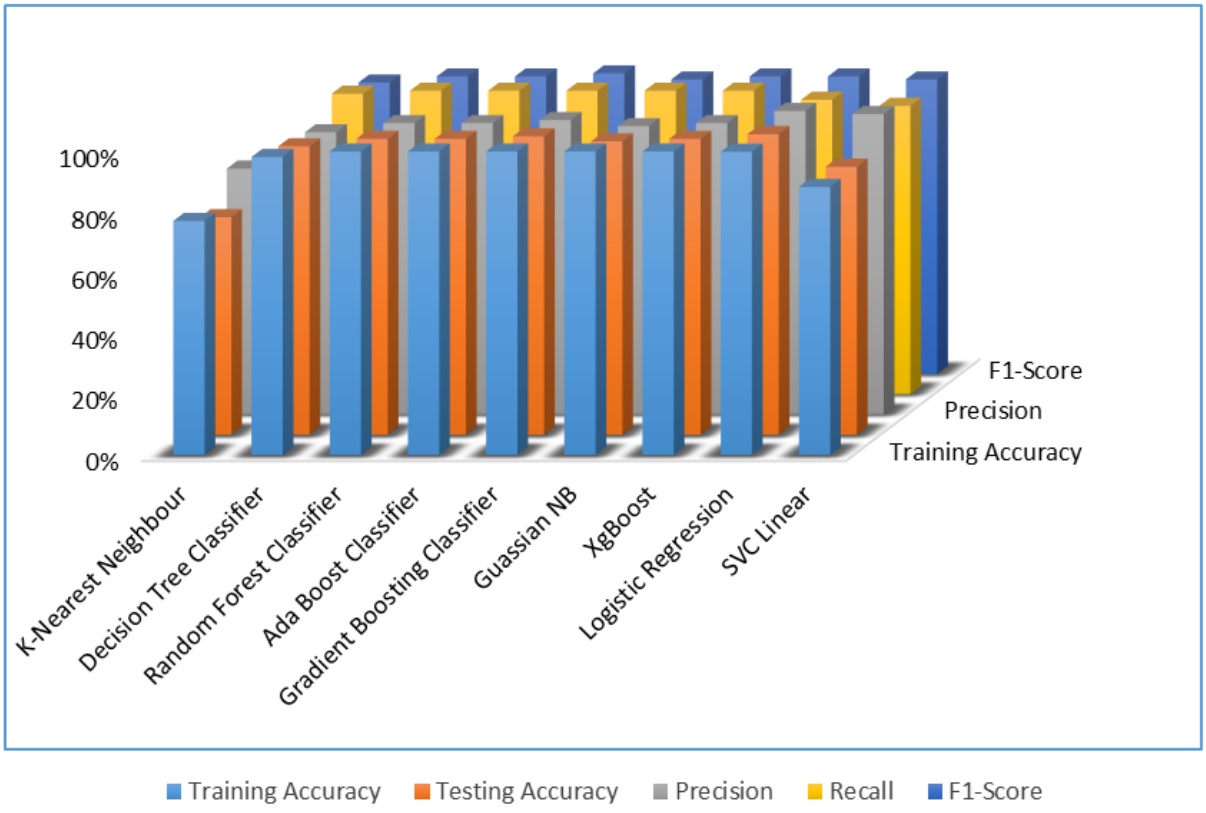

Fig 7: Accuracies, precision, recall & f1 score of each algorithm

The Figure 7 shows the Accuracies, precision, recall & f1 score of each algorithm

Model Evaluation and Selection:

Based on outcomes (Table -1), techniques with the highest accuracy were selected across all three datasets. In addition, there are three types of classifiers: random forest classifiers, ada boost classifiers, and gradient boosting classifiers. These models achieved 100% accuracy, which contributed significantly to the prediction of chronic kidney disease. These are ada boost classifiers, gradient boosting classifiers, and random forest classifiers. It was a perfect study in terms of accuracy. A number of factors, including hunger, pedal edema, anemia, etc., are not considered because they are insignificant in determining CLD prognosis. A collaborative method is more precise than an imputer, because missing values are omitted completely at random. It is important for a patient's health that these features first appear at a particular stage in their development, but this stage affects their degree of interconnectedness. This information can therefore be used to improve accuracy by training models. In contrast to alternative approaches, tree topologies provided greater accuracy with the exception of serum creatinine. Combining medical importance and statistical methods, we can narrow down the qualities that are most likely to predict Chronic Kidney Disease with a random forest classifier.

V. Conclusion And Future Work

It will be possible for patients to learn about chronic kidney disease early, when it can be treated most effectively from a cost and risk perspective, if chronic kidney disease is predicted with 100% accuracy. An efficient prediction algorithm may also be able to reduce the number of diagnostic tests required with careful feature engineering.

Classifiers that use random forests are best for predicting Chronic Kidney Disease. Data preparation, managing missing values, and selecting features are crucial to the prediction of Chronic Kidney Disease. Feature selection was also based on domain expertise when analyzing Chronic Kidney Disease-related clinical data.

Considering that many diseases have missing values, it might be useful to explore using K-Nearest Neighbor-imput grounded strategies in the future. Families, water consumption habits, and dietary preferences may also play a role in furthering our understanding of Chronic Kidney Disease.

References

[1] Adam, Ryan J., Adaysha C. Williams, and Alison J. Kriegel. "Comparison of the surgical resection and infarct 5/6 nephrectomy rat models of chronic kidney disease." *American Journal of Physiology-Renal Physiology* 322, no. 6 (2022): F639-F654.

[2] Mahaffey, Kenneth W., Meg J. Jardine, Severine Bompoint, Christopher P. Cannon, Bruce Neal, Hiddo JL Heerspink, David M. Charytan et al. "Canagliflozin and cardiovascular and renal outcomes in type 2 diabetes mellitus and chronic kidney disease in primary and secondary cardiovascular prevention groups: results from the randomized CREDENCE trial." Circulation 140, no. 9 (2019): 739-750.

[3] Ahsan, Md Manjurul, Shahana Akter Luna, and Zahed Siddique. "Machine-learning-based disease diagnosis: A comprehensive review." In *Healthcare*, vol. 10, no. 3, p. 541. MDPI, 2022.

[4] Banegas-Luna, Antonio Jesús, Jorge Peña-García, Adrian Iftene, Fiorella Guadagni, Patrizia Ferroni, Noemi Scarpato, Fabio Massimo Zanzotto, Andrés Bueno-Crespo, and Horacio Pérez-Sánchez. "Towards the interpretability of machine learning predictions for medical applications targeting personalised therapies: A cancer case survey." *International Journal of Molecular Sciences* 22, no. 9 (2021): 4394.

[5] Evans, Roger G. "Evolution of the glomerulus in a marine environment and its implications for renal function in terrestrial vertebrates." *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 324, no. 2 (2023): R143-R151.

[6] Jones-Freeman, Bernadette, Michelle Chonwerawong, Vanessa R. Marcelino, Aniruddh V. Deshpande, Samuel C. Forster, and Malcolm R. Starkey. "The microbiome and host mucosal interactions in urinary tract diseases." *Mucosal Immunology* 14, no. 4 (2021): 779-792.

[7] Akwaboah, Patrick Kwame, Akosua Animwah Somuah, and Stephen T. Odonkor. "Prevalence And Distribution Of Non-Communicable Diseases In Sub-Saharan Africa: The Case Of Hypertension, Diabetes, And Chronic Kidney Disease/Acute Kidney Injury." *Central European Management Journal* 30, no. 4 (2022): 1310-1326.

[8] Lowe, Kevin M., Jan Buenacosa Cruz, and Katerina M. Jones. "Complications in Patients with Chronic Kidney Disease." *Critical Care Nursing Clinics* 34, no. 4 (2022): 395-407.

[9] Garcia Sanchez, Juan Jose, Juliette Thompson, David A. Scott, Rachel Evans, Naveen Rao, Elisabeth Sörstadius, Glen James et al. "Treatments for chronic kidney disease: a systematic literature review of randomized controlled trials." *Advances in Therapy* 39, no. 1 (2022): 193-220.

[10] Javaid, Mohd, Abid Haleem, Ravi Pratap Singh, Rajiv Suman, and Shanay Rab. "Significance of machine learning in healthcare: Features, pillars and applications." *International Journal of Intelligent Networks* 3 (2022): 58-73.

[11] Zhong, Jianyong, Hai-Chun Yang, and Agnes B. Fogo. "A perspective on chronic kidney disease progression." *American Journal of Physiology-Renal Physiology* 312, no. 3 (2017): F375-F384.

[12] Jia, Hanxi, Guofu Qiao, and Peng Han. "Machine learning algorithms in the environmental corrosion evaluation of reinforced concrete structures-A review." *Cement and Concrete Composites* (2022): 104725.

[13] Solomon, Richard J., Madhu K. Natarajan, Serge Doucet, Samin K. Sharma, Cezar S. Staniloae, Richard E. Katholi, Joseph L. Gelormini, Marino Labinaz, Abel E. Moreyra, and Investigators of the CARE Study. "Cardiac Angiography in Renally Impaired Patients (CARE) study: a randomized double-blind trial of contrast-induced nephropathy in patients with chronic kidney disease." *Circulation* 115, no. 25 (2007): 3189-3196.

[14] Rajapurkar, Mohan M., George T. John, Ashok L. Kirpalani, Georgi Abraham, Sanjay K. Agarwal, Alan F. Almeida, Sishir Gang et al. "What do we know about chronic kidney disease in India: first report of the Indian CKD registry." *BMC nephrology* 13 (2012): 1-8.

[15] Ishigami, Junichi, Jonathan T. Taliercio, Harold I. Feldman, Anand Srivastava, Raymond R. Townsend, Debbie L. Cohen, Edward J. Horwitz et al. "Fibroblast growth factor 23 and risk of hospitalization with infection in chronic kidney disease: the chronic renal insufficiency cohort (CRIC) study." *Journal of the American Society of Nephrology: JASN* 31, no. 8 (2020): 1836.

[16] Xiao, Jing, Ruifeng Ding, Xiulin Xu, Haochen Guan, Xinhui Feng, Tao Sun, Sibo Zhu, and Zhibin Ye. "Comparison and development of machine learning tools in the prediction of chronic kidney disease progression." *Journal of translational medicine* 17, no. 1 (2019): 1-13.

[17] Almasoud, Marwa, and Tomas E. Ward. "Detection of chronic kidney disease using machine learning algorithms with least number of predictors." *International Journal of Soft Computing and Its Applications* 10, no. 8 (2019).

[18] Murale, C., M. Sundarambal, and R. Nedunchezhian. "Analysis on Ensemble Methods for the Prediction of Cardiovascular Disease." *Journal of Medical Imaging and Health Informatics* 11, no. 10 (2021): 2529-2537.

[19] Ekanayake, Imesh Udara, and Damayanthi Herath. "Chronic kidney disease prediction using machine learning methods." In *2020 Moratuwa Engineering Research Conference (MERCon)*, pp. 260-265. IEEE, 2020.

[20] Aruna, O., and Sk Sameerunnisa. "Chronic Kidney Disease Prediction using Data Pre-Processing Techniques." In 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), pp. 1665-1670. IEEE, 2023.

[21] Yashfi, Shanila Yunus, Md Ashikul Islam, Nazmus Sakib, Tanzila Islam, Mohammad Shahbaaz, and Sadaf Salman Pantho. "Risk prediction of chronic kidney disease using machine learning algorithms." In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pp. 1-5. IEEE, 2020.

[22] Farjana, Afia, Fatema Tabassum Liza, Parth Pratim Pandit, Madhab Chandra Das, Mahadi Hasan, Fariha Tabassum, and Md Helal Hossen. "Predicting Chronic Kidney Disease Using Machine Learning Algorithms." In *2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 1267-1271. IEEE, 2023.