International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 9, ISSN No. 2455-2143, Pages 76-83
Published Online January 2023 in IJEAST (http://www.ijeast.com)

COMPARATIVE ANALYSIS OF KIDNEY FAILURE PREDICTION AT AN EARLY STAGE USING MACHINE LEARNING ALGORITHMS

Mr.Saurav Raj, Mr. Kartik Attri, Mr. Sahil Chawla Mr. Saurabh Rastogi
Asst. professor
Computer Science Department
Maharaja Agrasen Institute of Technology
New Delhi, India

Abstract: **Chronic kidney disease (CKD) is a medical state of a person in which the kidney can't filter the blood due to which the body fills with extra water and waste products. It can lead to stroke, heart attack, heart failure, swelling of the feet and kidney failure, which can lead to death. The global health problem is growing rapidly as more and more people are being diagnosed with CKD. With advancing technology, as well as ongoing medical research, machine learning is being used in the healthcare sector to diagnose many diseases early. ML algorithms and decoding methods have been very useful in extracting, analyzing data and making predictions when a person is positive or negative about a disease based on the given data sets. ML algorithms and in-depth reading have been proven to be very true in detecting CKD early. Machine learning algorithms Cat boost classifier, Support Vector Machine (SVM), Decision Trees (DT), Random Forests, KNN, Artificial Neural Networks have been studied and applied in this work to perform comparative analysis to create a ML model that can accurately predict if a person is positive to CKD or not. This paper uses pre-data processing, including background and above-mentioned machine learning algorithms to build the most accurate model to accurately detect this disease CKD and perform a comparative analysis of various Machine learning models for prognosis of CKD.**

Keywords: **Chronic Kidney Disease, Machine Learning, Data Mining, ANN, KNN, Decision Trees, Logistic Regression, Support Vector Machine, Data Preprocessing, Feature extraction.**

I. INTRODUCTION

When the blood filtering capability of the kidney is hindered, the disease is called Chronic Kidney disease. Chronic Diseases refers to diseases persisting for a long time. When the kidney is unable to filter out the waste products and blood, fluid starts building up in the body due to which excess fluids get deposited in the body and several consequences are seen ,like swelling in feet, irritation, and high blood pressure due to which there is an increased risk of heart related diseases (Heart Attack, Heart Failure, Difficulty in breathing) . These fluid buildup can cause deposition of some salts in the body which can cause cystic ovaries in women. Some complications of CKD are:

- High BP
- Low blood count
- Weaker bones and muscles
- Nerve damage
- Poor nutritional health
- Fluid retention
- Uremia

CKD is evaluated by measured glomerular filtration rate (GFR) and creatinine level which is calculated by blood test KFT. All such parameters like creatinine level, GFR value are calculated using KFT (kidney function test). In severe cases, CKD may lead to kidney transplants or permanent dialysis. This paper is made to detect CKD at an early stage, as in later stages, the chances of survival and recovery is very low. We have used various ML algorithms like Cat boost Classifier, KNN, Artificial neural networks, Decision Tree, SVM, and Random Forest and also indicate a more accurate prediction model by performing a comparative analysis on the basis of efficiency of predicted outcomes by different techniques.

II. LITERATURE SURVEY

Md. Rashed-Al-Mahfuz[1] introduces a reliable method of classifying CKD and selecting annotations with simplified enhancement and cost effectiveness. In this study, classifiers were built using different algorithms such as Random Forest (RF), Gradient Boosting (GB), and Extreme. Gradient Boosting (XGB), Logistic Regression (LR), Support Vector Machines (SVM). The results of this study showed that the key features identified by SHAPE are consistent with today's clinical thinking. The HF taxonomy was also found to provide the greatest accuracy on a set of pathologically partitioned attributes. Therefore, the proposed HF separator and reduced diagnostic criteria could be used to reduce diagnostic costs and enable quality control in early treatment programs. Anusorn Charleonnan [2] presented a paper in which machine learning techniques are used to predict Chronic kidney disease based on clinical data. Four machine learning techniques are tested, including K Nearest Neighbors (ANN), Support Vector Machine (SVM), Logistic Regression (LR), and Decision Tree Classifier. These speculative models are based on the CRF database and combine the validity of these models to select the best prognosis for CRF.

Gazi Mohammed Ifraz [3] presents a paper in which they have done a comparative analysis of different machine learning algorithms and When cross validation measurements are used in the prediction of chronic kidney disease, the LR method outperforms the other processes.

DM. Perera [4] presents a method for controlling CKD using proper dietary planning by applying a classification algorithm to laboratory results from patient medical records. In this work, classifiers are built using various algorithms such as multiclass decision jungle, multiclass decision forest, multiclass neural network, and multiclass logistic regression [3].

Akash Maurya[5] presents a machine learning algorithm that uses a medical test record classification algorithm to suggest suitable meal plans for CKD patients. To slow progression of CKD.H, patients are given a recommended diet according to the potassium zone calculated from blood potassium levels. Zhang et al. [6] investigated the performance of artificial neural network (ANN) models when applied to survival prediction in patients with chronic kidney disease (CKD). Early prediction of CKD and proper diet can slow the progression of the disease. J. Aljaaf et al., T. G. Kaur et al. [7] used various data mining algorithms in a Hadoop environment to predict chronic kidney disease. Classifiers such as KNN (K-Nearest Neighbor) and SVM (Support Vector Machine) are used in research. Blood levels of creatinine, sodium, and urea play important roles in determining the prognosis of survival or the need for renal transplantation in dialysis patients and worse.

The CKD diagnostic methodology proposed by Jiong Ming Qin [8] is feasible in terms of data imputation and specimen diagnosis. After unsupervised dropping of non-data values determined by ANN imputation, the combined model can achieve satisfactory accuracy.

P.Panwong.et al[9] created a classification model to predict transition intervals for stages 3-5 of renal disease, using decision trees, K-nearest neighbors, Naive Bayes, and artificial neural networks to apply knowledge. Confirm, Classification Model One Attribute Group Selected.

S. Vijayarani. at.al [10] used support vector machines (SVMs) and artificial neural networks (ANNs) to predict kidney disease. This analysis compares the behavior of his two algorithms above in terms of accuracy and execution time.

S. Krishnamurthy [12] In this study, he developed and evaluated a number of models based on artificial intelligence, taking into account small variables such as gender, age, comorbidities and medications. These models predict a patient's risk of developing chronic kidney disease over the next 6 or 12 months. Among the various models tested. The most prominent factors included diabetes, age, gout, and use of sulfonamides and angiotensin, all of which are helpful with respect to his CKD.

Qiong Bai, [13] Five ML algorithms, including logistic regression, naive Bayes, random forest, decision tree, and K nearest neighbors, were trained and tested with 5-fold validation. The likelihood of each model was compared to that of the renal failure risk equation (KFRE). Three ML models, including logistic regression, naive Bayes, and random forest, showed similarity and high sensitivity compared to KFRE. KFRE had the highest accuracy and precision. This study demonstrated the potential of ML in assessing predictors of CKD based on readily available factors. Three ML models with reasonable performance and sensitive ratings suggest potential use in patient assessment.

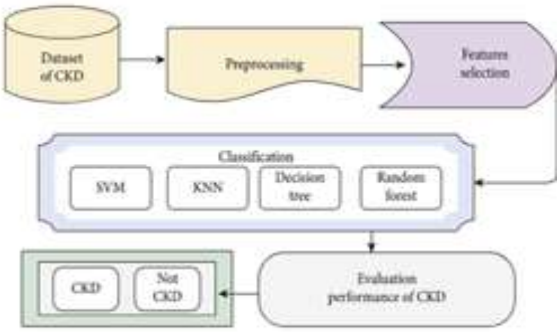
Marwa Almasoud [14] This paper aims to test the ability of machine learning algorithms to predict chronic kidney disease using small feature sets. Several mathematical tests have been performed to remove outdated functions such as the ANOVA test, Pearson's correlation, and Cramer's His V test. The logistic regression, support vector machine, random forest, and gradient boosting algorithms have been trained and tested with 10x validation. More accurate results are obtained from gradient boosting classifiers.

Arsh k Jain,In this paperperformed a structured assessment of the effects of neutral pH and low GDP compared to standard PD solutions. In this analysis, neutral pH and low GDP solutions led to higher urine output over 6 months and he improved RRF retention throughout all treatment periods. There were no significant differences in peritoneal UF and peritoneal small solute transport when solution was used [15] Giorgina Barbara Piccoli, This study focuses specifically on stage 1, pregnancy-related outcomes in women with associated CKD. We aimed to identify factors associated with the risk of adverse events in CKD. We aimed to determine whether poor pregnancy outcomes in women with stage 1 CKD were due to hypertension, proteinuria, systemic disease, or other clinically irrelevant factors associated with CKD [16].

Kunihiro Matsushita, e GFR and albuminuria, two key steps in CKD, improve prognosis for CVD risk in addition to the usual risk factors, especially CVD mortality and heart failure. Overall, albuminuria shows a much more marked improvement than serum creatinine-based eGFR. It seems reasonable to consider these his CKD measurements to predict CVD, especially when the data are already restricted to other clinical indicators (such as CKD patients). Other biological manifestations may be helpful if the risk profile is to be continuously improved, and measurements indicative of the pathophysiological processes of CVD, such as coronary artery calcium and sensitive troponin T cardiac, are particularly promising. Future clinical guidelines will address how to integrate CKD levels and other biological indicators into CVD predictors, depending on the CVD outcome of interest, the target population, and the availability of such tools and biomarkers. Updates may be required [17]. Hamida Ilyas [18] Early detection and treatment of CKD is highly desirable as it helps prevent unwanted side effects. Machine learning mechanisms are widely recommended for early symptom detection and diagnosis of various diseases. For the same reason, the aim of this study was to predict different stages of CKD using machine learning algorithms from datasets available from the medical records of affected individuals. Specifically, this study used random forests and the J48 algorithm to obtain a robust and workable model for determining the various stages of her CKD with perfect medical accuracy. A comparative analysis of the results revealed that J48 predicted her CKD better than random forest at all stages. Studies also show that J48 performs better than random forests.

Manal AAbdel-Fattah [19] In this paper, we proposea combination of machine learning techniques including feature selection and machine learning classification methods based on a big data platform (Apache Spark) used for diagnosis of chronic kidney disease (CKD). I'm here. feature selection techniques, d. H. Relief-F and chi-square were used to select key features. The following six classification algorithms were used in this study:Decision Trees (DT), Logistic Regression (LR), Naive Bayes (NB), Random Forests (RF), Support Vector Machines (SVM), and Gradient Boosted Trees (GBT Classifier) are available as built-in learning algorithms. Alternatively, for each algorithm, cross-validation and test results are computer-aided based on full features, relief F, and chi-square selection features. The results showed that his SVM, DT and GBT classifiers with selected features achieved good performance and high accuracy. Overall, the relief F-features selected outperform the full-featured features and the features selected by chi-square.

III. CKD PREDICTION USING MACHINE LEARNING MODELS



1) Decision Tree

It is a form of supervised learning. By using this we can solve both retrieve and edit problems.This separator acts as a tree data structure, where internal nodes display attributes or datasets, branches showing decision rules and leaf nodes indicating the result. It includes two types of nodes, name Decision and Leaf. The decision tree is based on the CART (Classification And Regression Tree) algorithm. It simply asks the question and is based on the result and divides the tree into a few small trees.

2) K-nearest neighbors (KNN)

Used for stage prediction problems. It is a type of supervised learning strategy. It does not think about the data provided, It is also known as a non-parameter learning algorithm. It is also known the The learning algorithm is lazy because it does not learn anything during training from the training set,instead it performs a database function as required. It keeps data in training and separates data continuously when it receives new data.

3) Random Forest

The informal forest forms many decision-making trees and based on their results further advances the diversification and reversal of processes. Some decision trees are constructed using random subsets of the training database. A large collection of decision trees provides highly accurate results. The algorithm runs faster and takes into account missing data. Random Forest creates a random algorithm. A decision phase is a type of class created by chopping down trees.

4) Support Vector Machine

Linear model of separation and reversing Vector Support Machine (SVM) that can be used to solve linear and indirect problems. it uses hyperplanes to separate the data. By doing so, each data object is sorted as a point in n-d space, with the value of each element being the sum of something.

5) Logistic Regression

Indentation is a mathematical method of predicting a binary outcome, such as yes or no, based on a predetermined data set. The systematic regression model predicts the variance of the dependent data by analyzing the relationship between one or more independent variables present.

6) Neural Networks

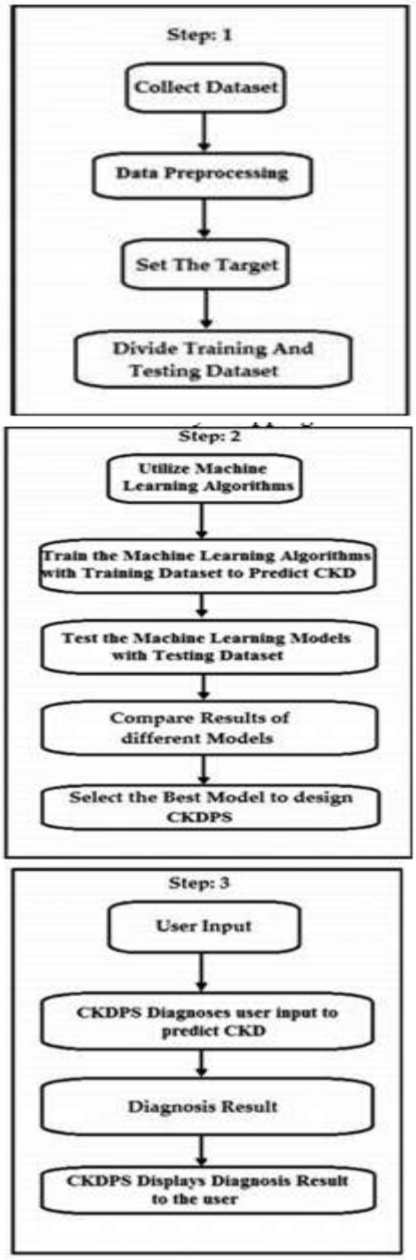
It's a set of algorithms that try to find fundamental relationships in a set of data through a process that mimics how the human brain works. In this sense, neural networks refer to neural systems, which can be organic or synthetic environments.

7) Artificial Neural Network

ANN models are extremely flexible for human emotional systems. ANN incorporates computational units similar to those of neurons of the biological nervous system known as synthetic neurons.

ANN is capable of learning and modeling indirect and complex relationships as most relationships between input and output are not interconnected. After training, ANN can determine non-binding relationships from intangible data, and as a result is generally made.

IV. DATA FLOW DIAGRAM



V. DATASET

We did some research on a set of chronic kidney disease data downloaded from kaggel. This database contains 25 attributes (including10 categories, 14 numbers and a class attribute). Specific 'category', possibly 'ckd' or 'not ckd' - ckd = chronic kidney disease. There are 400 lines. Data needs to be cleaned up: because it has NaNs and numerical features need to be forced to float. Basically, we were instructed to delete ALL LINES with Nans, without limitation - that is, any line with one NN, is deleted.

Chronic Kidney Disease Dataset

	age	bp	sg	al	su	rbc	pc	pcc	ba	bgr	...	pcv	wbcc	rbcc	htn	dm	cad	appet	pe	ane	class
0	48.0	80.0	1.020	1.0	0.0	NaN	normal	notpresent	notpresent	121.0	...	44.0	7800.0	5.2	yes	yes	no	good	no	no	ckd
1	7.0	50.0	1.020	4.0	0.0	NaN	normal	notpresent	notpresent	NaN	...	38.0	6000.0	NaN	no	no	no	good	no	no	ckd
2	62.0	80.0	1.010	2.0	3.0	normal	normal	notpresent	notpresent	423.0	...	31.0	7500.0	NaN	no	yes	no	poor	no	yes	ckd
3	48.0	70.0	1.005	4.0	0.0	normal	abnormal	present	notpresent	117.0	...	32.0	6700.0	3.9	yes	no	no	poor	yes	yes	ckd
4	51.0	80.0	1.010	2.0	0.0	normal	normal	notpresent	notpresent	106.0	...	35.0	7300.0	4.6	no	no	no	good	no	no	ckd

Dataset Data Set Information:

We use the following presentations to collect data
years - years
bp - blood pressure
sg - certain gravitational force
albumin
su - sugar
rbc - red blood cells
pc - pus cell
pcc - pus cell clumps
bacteria
bgr - random blood sugar
blood urea
sc - serum creatinine
salt - sodium
pot - potassium
hemoglobin - hemoglobin
pcv - full cell volume
wc - the number of white blood cells
rc - the number of red blood cells
htn - high blood pressure
dm - diabetes
Cad - a disease of the coronary artery
hunger - appetite
pedal edema
anemia
class - class

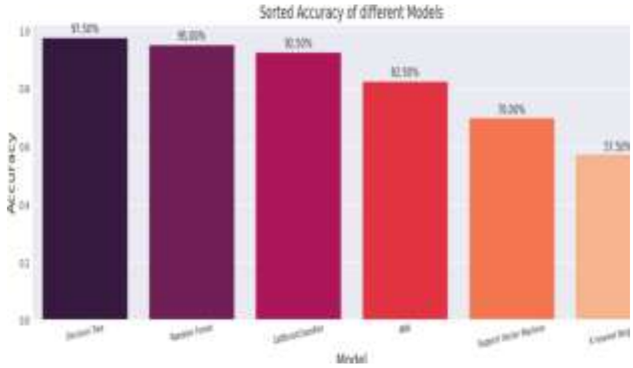
VI. RESULT ANALYSIS

In this paper, we presented a comparative analysis of prediction algorithms to find the most efficient algorithms for predicting CKD at an early stage. The dataset shows the input parameters collected from CKD patients and the model was trained and validated on the specified input parameters. Build decision trees, random forests, K-nearest neighbors, artificial neural networks, logistic regression, and cat-boost classifier learning models to perform CKD diagnosis. Model performance is evaluated based on the

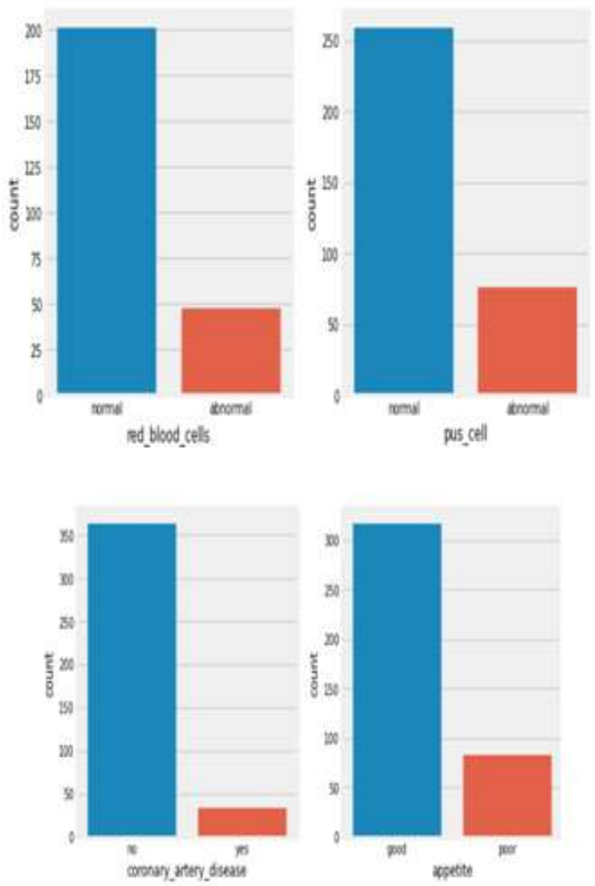
International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 9, ISSN No. 2455-2143, Pages 76-83
Published Online January 2023 in IJEAST (http://www.ijeast.com)

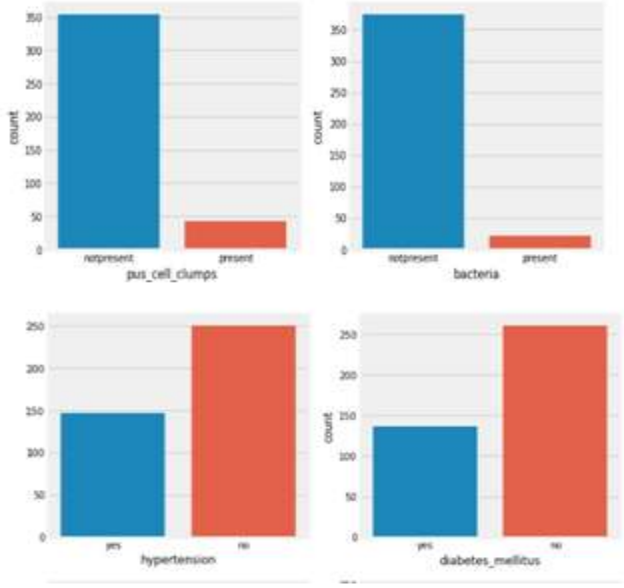
accuracy of predictions. Research has shown that Decision tree model and random forest classifiers predict CKD better compared to other algorithms. Comparisons can also be made based on run time, feature set selection, as an improvisation of this study.

VII. FUTURE SCOPE



Comparative analysis of various attributes of dataset



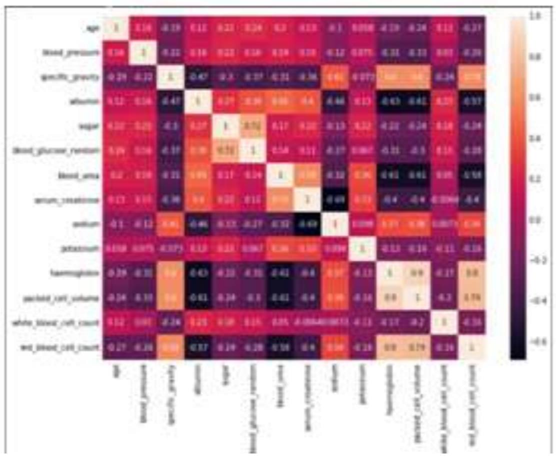


Heat map

A heat map is a way of representing a particular kind of matrix structure. To use data in heatmap, it must be in matrix format. Matrix means that the index names and column names must be the same in order for the data you enter in the cells to be related.

Violin plot of red blood cell

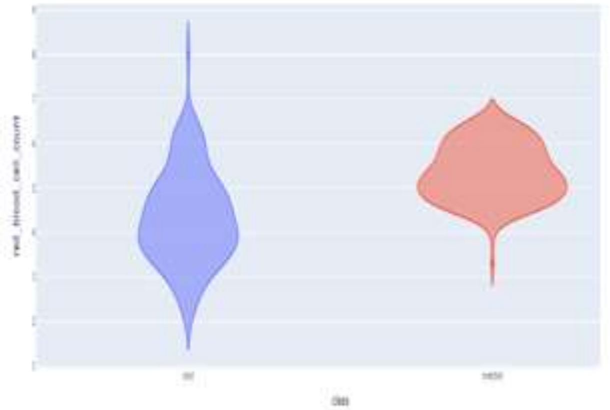
A violin plot is a combination of a box structure and a kernel density plot that reflects the height of the data. Used to visualize the distribution of numerical data. Unlike the box structure, which can only show summary statistics, the violin section shows summary statistics and the density of each variation.



International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 9, ISSN No. 2455-2143, Pages 76-83
Published Online January 2023 in IJEAST (http://www.ijeast.com)

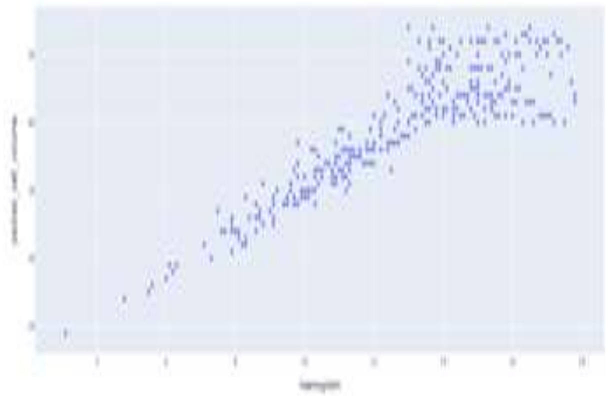
Scatter plot

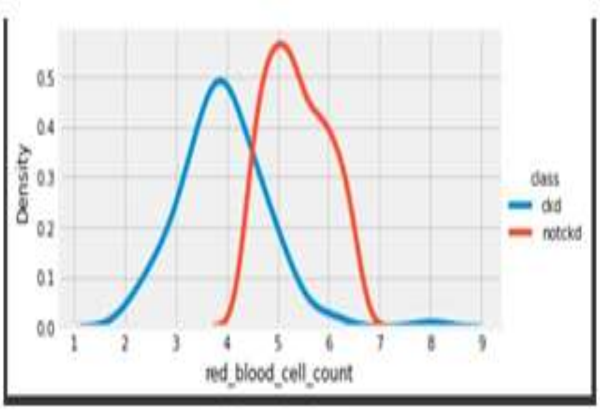
A scatter structure is a type of structure or diagram that uses Cartesian links to display values, usually with two dynamic data sets.Additional variants can be displayed if the points are coded.



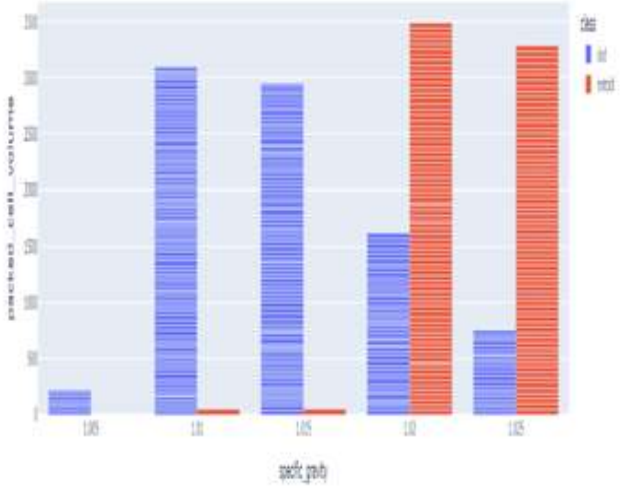
KDE plot

The Kernel Density Estimate (KDE) structure is a way to see the distribution of visuals in your database, similar to a histogram. KDE represents data using continuous curves in one or more directions.





Bar graph



Preprocessing

Pre-data processing is the process of preparing raw data and fitting it to a machine learning model. This is the first and most important step in creating a machine learning model.

Getting clean and formatted data is not always a problem for us when creating machine learning projects. And while any data processing is done, it is compulsory to clean it and format it in a format. So in this case, we are using a data processing function.

	age	blood_pressure	specific_gravity	albumin	sugar	red_blood_cells	pus_cell
0	48.0	80.0	1.020	1.0	0.0	1	1
1	7.0	50.0	1.020	4.0	0.0	1	1
2	62.0	80.0	1.010	2.0	3.0	1	1
3	48.0	70.0	1.005	4.0	0.0	1	0
4	51.0	80.0	1.010	2.0	0.0	1	1

pus_cell_clumps	bacteria	blood_glucose_random	blood_urea	serum_creatinine	sodiu
0	0	121.0	36.0	1.2	137.
0	0	352.0	18.0	0.8	137.
0	0	423.0	53.0	1.8	141.
1	0	117.0	56.0	3.8	111.
0	0	106.0	26.0	1.4	131.

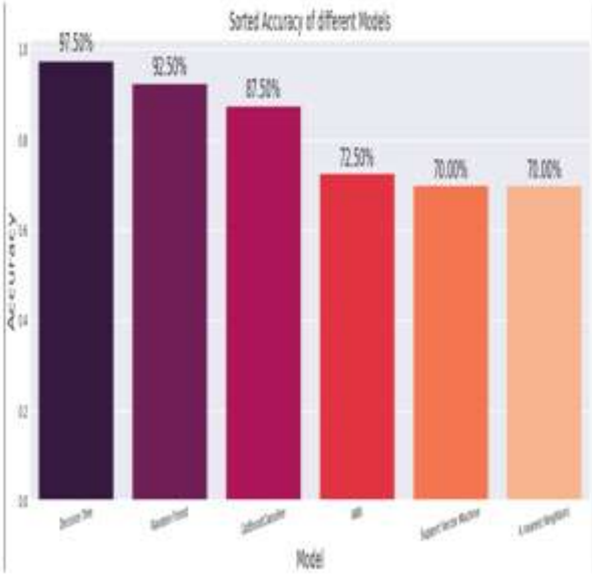
Feature extraction
Feature extraction refers to the process of transforming raw data into numerical features that can be processed while preserving the information in the original data. better results are generated when raw data is used directly on machine learning.

	white_blood_cell_count	blood_glucose_random	blood_urea
0	7800.0	121.0	36.0
1	6000.0	352.0	18.0
2	7500.0	423.0	53.0
3	6700.0	117.0	56.0
4	7300.0	106.0	26.0
...	...	...	...
395	6700.0	140.0	49.0
396	7800.0	75.0	31.0
397	6600.0	100.0	26.0
398	7200.0	114.0	50.0
399	6800.0	131.0	18.0

packed_cell_volume	albumin	haemoglobin	age	sugar	hypertension
44.0	1.0	15.4	48.0	0.0	1
38.0	4.0	11.3	7.0	0.0	0
31.0	2.0	9.6	62.0	3.0	0
32.0	4.0	11.2	48.0	0.0	1
35.0	2.0	11.6	51.0	0.0	0
...	...	...	...	...	...
47.0	0.0	15.7	55.0	0.0	0
54.0	0.0	16.5	42.0	0.0	0
49.0	0.0	15.8	12.0	0.0	0
51.0	0.0	14.2	17.0	0.0	0
53.0	0.0	15.8	58.0	0.0	0

Finalresult

```
Confusion matrix of  CatBoostClassifier
[[25  3]
 [ 0 12]]
Accuracy score is  0.925
--------------------------------------------------------
Confusion matrix of  Support Vector Machine
[[28  0]
 [12  0]]
Accuracy score is  0.7
--------------------------------------------------------
Confusion matrix of  Decision Tree
[[26  2]
 [ 0 12]]
Accuracy score is  0.95
--------------------------------------------------------
Confusion matrix of  ANN
[[27  1]
 [ 6  6]]
Accuracy score is  0.825
--------------------------------------------------------
Confusion matrix of  Random Forest
[[26  2]
 [ 1 11]]
Accuracy score is  0.925
```



VIII. CONCLUSION:

Chronic kidney disease has a significant impact on patient morbidity and mortality. A conservative treatment regimen is important to slow the progression of renal dysfunction and reduce the incidence of complications, which positively impacts the prognosis of the affected population.
The accuracy of the Decision tree model is highest as compared to the other machine learning models to predict the Chronic Kidney Disease.

IX. REFERENCES:

[1]. Rashed-Al-Mahfuz, M., Haque, A., Azad, A., Alyami, S. A., Quinn, J. M., & Moni, M. A. (2021). Clinically applicable machine learning approaches to identify attributes of Chronic Kidney Disease (CKD) for use in low-cost diagnostic screening. IEEE Journal of Translational

Engineering in Health and Medicine, 9, 1-11.

[2]. Ifraz, G. M., Rashid, M. H., Tazin, T., Bourouis, S., & Khan, M. M. (2021). Comparative Analysis for Prediction of Kidney Disease Using Intelligent Machine Learning Methods. Computational and Mathematical Methods in Medicine, 2021.

[3]. Charleonnan, A., Fufaung, T., Niyom Wong, T., ChokchaiPattanakit, W., Suwannawach, S., &Ninchawee, N. (2016, October). Predictive analytics for chronic kidney disease using machine learning techniques. In 2016 management and innovation technology international conference (MITicon) (pp. MIT-80). IEEE.

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

[5]. Maurya, A., Wable, R., Shinde, R., John, S., Jadhav, R., &Dakshayani, R. (2019, January).

[6]. Chronic kidney disease prediction and recommendation of a suitable diet plan by using machine learning. In 2019 International Conference on NascentTechnologiesin Engineering (ICNTE) (pp. 14). IEEE.

[7]. 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

[8]. 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.

[9]. Qin, J., Chen, L., Liu, Y., Liu, C., Feng, C., & Chen, B. (2019). A machine learning methodology for diagnosing chronic kidney disease. IEEE Access, 8, 20991-21002.

[10]. P. Panwong and N. Iam-On, "Predicting transitional interval of kidney disease stages 3 to 5 using data mining method," 2016 Second Asian Conference on Defense Technology (ACDT), Chiang Mai, 2016, pp. 145-150.

[11]. S. Vijayarani, S. Dhayanand, "KIDNEY DISEASE PREDICTION USING SVM AND ANN ALGORITHMS", International Journal of Computing and Business Research (IJCBR), vol. 6, no. 2, 2015.

[12]. Krishnamurthy, S.; KS, K.; Dovgan, E.; Luštrek, M.; GradišekPileti˘c, B.; Srinivasan, K.; Li, Y.-C.; Gradišek, A.; Syed-Abdul, S. Machine Learning Prediction Models for Chronic Kidney Disease Using National Health Insurance Claim Data in Taiwan. Healthcare2021,9,546.https://doi.org/10.3390/healthcare9050546

[13]. Bai, Qiong, et al. "Machine learning to predict end stage kidney disease in chronic kidney disease." Scientific reports 12.1 (2022): 1-8.

[14]. Almasoud, Marwa, and Tomas E. Ward. "Detection of chronic kidney disease using machine learning algorithms with the least number of predictors." International Journal of Soft Computing and Its Applications 10.8 (2019).

[15]. Yohanna, S., Alkatheeri, A. M., Brimble, S. K., McCormick, B., Iansavitchous, A., Blake, P. G., & Jain, A. K. (2015). Effect of Neutral-pH,Low–Glucose-Degradat ion Product Peritoneal Dialysis Solutions on Residual Renal Function, Urine Volume, and Ultrafiltration: A Systematic Review and Meta-Analysis. Clinical Journal of the American Society of Nephrology, 10(8), 1380-1388.

[16]. Piccoli, G. B., Cabiddu, G., Attini, R., Vigotti, F. N., Maxia, S., Lepori, N., ... &Todros, T. (2015). Risk of adverse pregnancy outcomes in women with CKD. Journal of the American Society of Nephrology, 26(8), 2011-2022.

[17]. Matsushita, K., Ballew, S. H., & Coresh, J. (2016). Cardiovascular risk prediction in people with CKD. Current opinion in nephrology and hypertension, 25(6), 518.

[18]. Ilyas, H., Ali, S., Ponum, M., Hasan, O., Mahmood, M.T., Iftikhar, M. and Malik, M.H., 2021. Chronic kidney disease diagnosis using decision tree algorithms. BMC nephrology, 22(1), pp.1-11.

[19]. Abdel-Fattah, Manal A., NerminAbdelhakim Othman, and NagwaGoher. "Predicting Chronic Kidney Disease Using Hybrid Machine Learning Based on Apache Spark."

[20]. Agarwala, Abhimanyu, AsfarSharief, and Faaiz Ahmed. Analysis And Prediction Of Chronic Kidney Disease Using Machine Learning Classification Approaches. Diss. CMR Institute of Technology. Bangalore, 2020.