

RISK PREDICTION OF RENAL FAILURE USING MACHINE LEARNING

Submitted in partial fulfillment of the requirements for the award of
Bachelor of Technology degree in
Information Technology
by

I. HEMANTH KUMAR (Reg. No – 39120039)

B. LAKSHMI VIGNESH (Reg. No - 39120059)



**DEPARTMENT OF INFORMATION TECHNOLOGY
SCHOOL OF COMPUTING**

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

Accredited with Grade "A" by NAAC | 12B Status by UGC | Approved by AICTE

JEPPIAAR NAGAR, RAJIV GANDHI SALAI,

CHENNAI - 600119

MAY- 2023



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

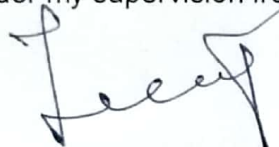
Accredited with Grade "A" by NAAC | 12B Status by UGC | Approved by AICTE

www.sathyabama.ac.in

DEPARTMENT OF INFORMATION TECHNOLOGY

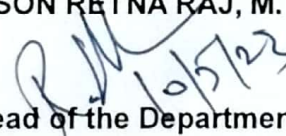
BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **I. HEMANTH KUMAR (39120039)** and **B. LAKSHMI VIGNESH (39120059)** who carried out the Project Phase-2 entitled "**RISK PREDICTION OF RENAL FAILURE USING MACHINE LEARNING**" under my supervision from January 2023 to May 2023.



Internal Guide

Dr. R. JEBERSON RETNA RAJ, M. Tech., Ph. D,

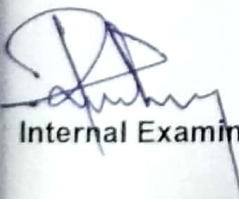


Head of the Department

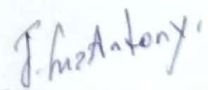
Dr. R. SUBHASHINI M.E., Ph.D.

Submitted for Viva voce Examination held on

10-05-2023



Internal Examiner



External Examiner

DECLARATION

I, I. HEMANTH KUMAR (39120039), hereby declare that the Project Phase-2 Report entitled "RISK PREDICTION OF RENAL FAILURE USING MACHINE LEARNING" done by me under the guidance of DR. R. JEBERSON RETNA RAJ, M.E., Ph. D is submitted in partial fulfilment of the requirements for the award of Bachelor of Technology degree in Information Technology.

DATE: 10-05-2023

PLACE: Chennai

I. Hemanthkumar.

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management of SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. SASIKALA M.E., Ph. D, Dean, School of Computing, Dr. R. SUBHASHINI, M.E., Ph.D.,** Head of the Department of Information Technology and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **DR. JEBERSON RETNA RAJ, M.E., Ph. D.,** for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of my phase- 2 project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Information Technology** who were helpful in many ways for the completion of the project.

ABSTRACT

Chronic kidney disease (CKD) is a worldwide health issue with an excessive morbidity and mortality rate, and it is associated with different diseases. Since there aren't any conspicuous facial outcomes throughout the early levels of CKD, sufferers often overlook the illness. Early discovery of CKD empowers sufferers to get opportune remedies to avoid the spread of this infection. Machine learning technologies correctly assist clinicians in accomplishing this goal due to their instant and accurate execution. In India, nearly one million cases were registered each year, with nearly 13 percent of those converting to a mortality rate as the patients died. In this project, the proposed system consists of different suitable prediction models like the KNN (K Nearest Neighbour), Decision Tree (DT), Random Forest (RF), and SVM (super vector machine or supervised machine learning). These are machine learning prediction models for finding the better performing prediction models and also for diagnosing chronic kidney disease. The proposed system uses the CKD clinical data taken from the University of California, Irvine (UCI) AI store. It is one of the best machine learning repositories and has a brilliant quantity of lacking characteristics. And with all this clinical data and prediction models, we can test clinical data one by one, and in different machine learning environments, we can find a better system with the performance parameters.

Chapter No	TITLE		Page No.
	ABSTRACT		v
	LIST OF FIGURES		viii
	LIST OF ABBREVIATIONS		ix
1	INTRODUCTION		1
	1.1 CHRONIC KIDNEY DISEASE		1
	1.2 CKD EFFECTS		1
	1.3 MACHINE LEARNING		1
2	LITERATURE SURVEY		3
	2.1 INFERENCES FROM LITERATURE SURVEY		16
	2.2 EXISTING SYSTEM		16
3	REQUIREMENTS ANALYSIS		18
	3.1	FEASIBILITY STUDIES/RISK ANALYSIS OF THE PROJECT	18
	3.2	SOFTWARE REQUIREMENTS SPECIFICATION DOCUMENT	19
4	DESCRIPTION OF PROPOSED SYSTEM		20
	4.1	SELECTED METHODOLOGY OR PROCESS MODEL	20
	4.2	ARCHITECTURE / OVERALL DESIGN OF PROPOSED SYSTEM	20
	4.3	DATA FLOW DIAGRAM	21
	4.4	USECASE DIAGRAM	22
	4.5	CLASS DIAGRAM	23
	4.6	ACTIVITY DIAGRAM	24
	4.7	SEQUENCE DIAGRAM	25

	4.8	DESCRIPTION OF SOFTWARE FOR IMPLEMENTATION AND TESTING PLAN OF THE PROPOSED MODEL/SYSTEM	26
5	IMPLEMENTATION DETAILS		30
	5.1	DEVELOPMENT AND DEPLOYMENT SETUP	30
	5.2	ALGORITHMS	31
	5.3	INPUT DESIGN	32
	5.4	OUTPUT DESIGN	32
	5.5	UNIT TESTING	33
6	RESULTS AND DISCUSSION		34
	6.1	RESULT	34
	6.2	COMPARISION	35
7	CONCLUSION		36
	7.1	CONCLUSION	36
	7.2	FUTURE WORK	36
	REFERENCES		38
	APPENDIX		41
	A. SOURCE CODE		41
	B. SCREENSHOTS		46
	C. RESEARCH PAPER		48

LIST OF FIGURES

Figure No	Title	Page No
4.2	ARCHITECTURE DIAGRAM	20
4.3	DATA FLOW DIAGRAM	22
4.4	USECASE DIAGRAM	23
4.5	CLASS DIAGRAM	24
4.6	ACTIVITY DIAGRAM	25
4.7	SEQUENCE DIAGRAM	26
5.3	INPUT DATASET	32
5.4	OUTPUT DESIGN	33
5.5	UNIT TEST RESULT	33

ABBREVIATIONS

ESKD: End-stage kidney disease

CKD: Chronic kidney disease

KFRE: Kidney Failure Risk Equation

eGFR: Estimated glomerular filtration rate

ACR: Urine albumin-to-creatinine ratio

EHR: Electronical health record

UA: Uric acid AST: Aspartate aminotransferase

ALT: Alanine transaminase

TBIL: Total bilirubin

CREA: Serum creatine

LR: Logistic regression

RF: Random forest

DT: Decision tree

SVM: Support vector machine

AUC: Area under the receiver operating characteristic curve

NPV: Negative predictive value

PPV: Positive predictive value

HDL-C: high-density lipoprotein cholesterol

LDL-C: Low-density lipoprotein cholesterol

ROC: Receiver operating characteristic

CHAPTER 1

INTRODUCTION

1.1 CHRONIC KIDNEY DISEASE

Chronic kidney disease, also known as chronic renal failure, is associated with a gradual loss of renal function. The kidneys filter waste products and excess water from the blood and excrete it in the urine. Advanced chronic kidney disease can accumulate dangerous levels of water, electrolytes, and waste products in the body. In the early stages of chronic kidney disease, there may be few signs or symptoms. You may not be aware of your kidney disease until it has progressed. Treatment of chronic kidney disease usually focuses on slowing the progression of kidney damage by controlling the cause. But even managing the cause cannot stop the progression of kidney damage.

1.2 CKD EFFECTS

Chronic kidney disease can progress to end-stage renal disease, which is fatal without artificial filtration (dialysis) or kidney transplantation. Kidney disorder additionally will increase your chance of getting coronary heart and blood vessel disorder. These issues can also additionally manifest slowly over an extended time. Early detection and remedy can regularly maintain persistent kidney disorder from getting worse. When kidney disorder progresses, it is able to finally result in kidney failure, which calls for dialysis or a kidney transplant to keep life. So as earlier we have discussed we need a better prediction system to avoid risk of dying as well as controlling the morbidity rate.

1.3 MACHINE LEARNING

Machine learning (ML) is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy. As the name says, is all about machines learning automatically without being explicitly programmed or learning without any direct human intervention. This machine learning process starts with feeding them good quality data and then training the machines by building various machine learning models using the data and different algorithms. The choice of algorithms depends on what type of data we have and what kind of task we are trying to automate. It is fair to start with this point, because ML is very good at

diagnosis; actually, this is one of the most effective areas. There are plenty of types of cancer and genetic diseases that are hard to detect; however, ML could handle many of them in the initial stages. It's no secret that clinical trials could take years to complete, with significant investments required. ML can offer predictive analytics to spot the best candidates for clinical trials, based on factors like one's history of doctor visits or social media activity. The technology will also lower the number of data-based errors and could suggest the best sample sizes to be tested.

1.4 AIM OF THE PROJECT

The main aim of this project is to find the better prediction system that could predict chronic kidney disease (CKD) with the clinical data and Machine learning algorithms like K Nearest Neighbor (KNN), Decision Tree (DT), Random Forest (RF), Support vector machine (SVM) and the system must have higher accuracy in predicting the chances of CKD among the individuals as well as it must have better performance parameters.

1.5 SCOPE OF THE PROJECT

The scope of this project is finding better a pathway that acts as a prediction system, gets medical data of patients and we call that a clinical data which consist of patient medical information like Specific gravity, Hemoglobin, age etc. with the help of machine learning algorithm it predicts the patients who have chronic kidney disease instantaneously and with good accuracy as well in an economical way.

CHAPTER 2

LITERATURE SURVEY

In [1], chronic kidney disease (CKD) has grown to be a serious global health issue. Between 1990 and 2017, a study by the GBD Chronic Kidney Disease Collaboration sought to determine the burden of CKD on a global, regional, and country scale. The study discovered that, with a sharp rise in fatalities and disability-adjusted life years (DALYs) between 1990 and 2017, CKD is now the world's leading cause of death and disability. 35.8 million DALYs and 1.2 million fatalities were attributed to CKD in 2017. The research also emphasized that CKD affects low- and middle-income countries more than high-income countries. To lessen the effects of CKD, risk factors like obesity, diabetes, and hypertension must be addressed. Additional research from the US and China. Other research from the US and China has also shown how common CKD is and how it affects healthcare costs and quality of life, highlighting the need for a worldwide strategy to handle this problem.

In [2], End-stage renal disease (ESRD) is a significant public health issue in the United States with considerable effects on patients and healthcare systems. Somnath Pal reviewed the literature and looked at the main causes of ESRD in the US. According to the study, diabetes and hypertension, which account for approximately two-thirds of cases, are the two main causes of ESRD. About 44% of instances of ESRD are caused by diabetes, which is also the most common cause. Diabetes-related high blood sugar levels can gradually harm the kidneys' filtration function, resulting in kidney failure. With 28% of cases, hypertension is the second most typical cause of ESRD. The kidneys' ability to filter waste and fluids can be hampered by high blood pressure because it can harm the blood vessels in the kidneys. Glomerulonephritis, polycystic kidney disease, and autoimmune illnesses are some additional less frequent causes of ESRD. Compared to other racial and ethnic groups, African Americans, Hispanics, and Native Americans have a higher chance of having ESRD. The study emphasizes how controlling diabetes and hypertension is crucial for delaying the onset and progression of ESRD. Early

kidney disease detection and treatment can lessen the burden on patients and healthcare systems by helping to delay or avoid the need for dialysis or kidney transplantation.

In [3], A national registry with the mission of tracking the incidence, treatment, and prognosis of kidney disease in China is called the China Kidney Disease Network (CK-NET). L. Zhang et al.'s CK-NET 2015 annual data report offers a thorough examination of the network's data. According to the study, China has a high prevalence of chronic kidney disease (CKD), with an estimated 119.5 million cases in 2015. Hypertension, followed by diabetes and glomerulonephritis, was the most frequent cause of CKD. The study also showed differences in CKD incidence and prevalence by geography, with a greater burden in rural and western sections of the nation. The research emphasizes the necessity of better CKD care and prevention in China. The prevalence of CKD in the nation could be decreased with the aid of measures including early detection and treatment of hypertension and diabetes, lifestyle modifications, and public awareness campaigns. The management and results of patients with end-stage renal disease (ESRD) in China are also revealed by the CK-NET registry. According to the study, hemodialysis was the most popular form of treatment for those with ESRD, with a high prevalence of catheter use and insufficient dialysis dosage. The emphasizes the need of raising the availability and caliber of dialysis services for ESRD patients in China.

In [4], Chronic kidney disease (CKD) is a serious public health issue, and prevention and therapy of CKD depend on identifying those who are at risk of acquiring it and tracking its progression. A comprehensive study was done by J.B. Echouffo Tcheugui and A.P. Kengne to assess how well risk models predicted CKD and its progression. The study found 12 risk models for CKD progression and 22 risk models for CKD. The models' performance varied greatly; some had strong calibration and discriminatory power, while others did not. According to the study, the majority of risk models contained well-known risk factors such age, sex, race, diabetes, and hypertension. The scientists came to the conclusion that better risk models are required, ones that take into account novel risk variables and unusual

markers including albuminuria, glomerular filtration rate, and genetic markers. They also emphasized how crucial it is to test how well risk models operate across various populations and contexts. The study emphasizes the current shortcomings and inadequacies in risk prediction models for the development of CKD. In order to identify those who are at a high risk of developing CKD and to enable focused interventions and prevention measures, risk models should be improved. The study emphasizes the necessity of ongoing work in this field to enhance CKD care and prevention.

In [5], Early detection and therapy of chronic kidney disease (CKD), a serious public health issue, are crucial to halting its progression. To help identify those who are at risk of developing CKD, H. Bang and colleagues created the Screening for Occult Renal Disease (SCORED) model, a straightforward CKD prediction tool. The Third National Health and Nutrition Examination Survey (NHANES III) provided data on more than 14,000 people, and this information was used to create the SCORED model. A different cohort of more than 5,000 people was used for validation. Age, sex, race, and a history of hypertension are four readily available characteristics that the model uses to predict the likelihood of getting CKD. The study discovered that the SCORED model performed better in identifying people at risk than other prediction models because it had strong discriminatory power and calibration. The study discovered that the SCORED model beat previous prediction models in identifying those who were at risk of developing CKD and had good discriminatory power and calibration. The SCORED model is a helpful tool for primary care doctors to identify people at risk of CKD because of its simplicity and use of widely available information, according to the authors. The creation and validation of the SCORED model show that it is possible to identify people who are at risk of developing CKD by utilizing straightforward prediction models. By using such models in primary care settings, early interventions could be made to stop the course of CKD and assist in identifying people at risk. The study emphasizes how crucial it is to carry out further research in order to create and validate prediction models for CKD.

In [6], Kshirsagar and colleagues set out to develop an easy-to-use algorithm that could predict the incidence of renal illness in a certain neighborhood. The screening for Occult Renal Disease (SCORED) project data were used to develop the approach, and the Atherosclerosis Risk in Communities (ARIC) research data were used to validate it. Age, sex, race, blood pressure, diabetes status, smoking status, and history of cardiovascular disease are just a few of the seven readily available clinical factors that were incorporated to the algorithm. It demonstrated strong prediction accuracy with a C-statistic of 0.81 in the SCORED dataset and 0.79 in the ARIC dataset. Finding those at a high risk of developing renal disease is crucial for early intervention and risk factor control. This may contribute to lessening this strain.

In [7], C.S. Fox et al.'s 2010 study looked into the use of a multi-marker approach in determining the likelihood of developing chronic kidney disease (CKD) and microalbuminuria. The Framingham Offspring Study, which tracked a cohort of people for 8 years, provided the researchers with the data they used. The analysis comprised 2,345 participants in total. The multi-marker method utilized in the study involves evaluating a combination of more recent biomarkers, such as homocysteine and fibrinogen, as well as more established risk factors for CKD, such as blood pressure and glucose levels. In comparison to employing only traditional risk factors, the researchers discovered that this strategy was able to greatly enhance the prediction of incident CKD and microalbuminuria. The study also discovered certain indicators, such as cystatin C and urinary albumin-to-creatinine ratio, that were highly predictive of CKD and microalbuminuria. These markers might be helpful in identifying those who are highly susceptible to developing CKD or microalbuminuria, enabling earlier intervention and management.

In [8], Hippisley-Cox and Coupland set out to develop and verify prediction models for men and women in England and Wales with chronic kidney disease (CKD). They made use of the QResearch database, which holds the medical records of more than 7 million patients who are registered with general practitioners. The models, which were externally evaluated by utilizing a different cohort of

patients, contained variables such as age, sex, ethnicity, smoking status, diabetes, hypertension, and past cardiovascular disease. With AUCs of 0.87 and 0.89 for men and women, respectively, and satisfactory calibration, which means that the projected hazards were congruent with the risks encountered by the validation group, the QKidney scores demonstrated good discriminatory power. In order to identify high-risk CKD patients for early therapies or screening programmes and to support collaborative decision-making, the QKidney scores could be employed. More investigation is required.

In [9], calculating the likelihood that Taiwanese people will develop chronic kidney disease (CKD), Chien et al. built a prediction model. Data from a community-based cohort study with over 12,000 people aged 35 to 64 was used by the authors. At baseline and during follow-up visits, the participants' demographic data, medical history, and laboratory test results were gathered. To build their prediction model, the authors used demographic and clinical factors such as age, sex, body mass index, blood pressure, diabetes, smoking status, and laboratory results. Several statistical techniques, such as calibration plots and the area under the receiver operating characteristic curve (AUC), were used to evaluate the model's performance. An AUC of 0.77 indicated from the findings that the prediction model had good discriminating power. The model's calibration was also accurate, which means that the anticipated and observed hazards were closely related. The scientists came to the conclusion that people in Taiwan who are at a high risk of having CKD might be identified using their prediction model. This study emphasizes the significance of early CKD risk factor identification since it enables prompt therapies to stop or delay the onset of CKD. By identifying high-risk people for focused therapies, the prediction model could also be used to optimize healthcare resource allocation and enhance patient outcomes. Further research is necessary to see whether the approach can be applied to other populations and healthcare settings.

In [10], the goal of the work by Ando et al. was to develop a straightforward model to estimate the likelihood that patients with HIV infection may develop chronic kidney disease (CKD). To create and validate their prediction model, the scientists

used data from 1,067 HIV-positive people in Japan, who were split into two groups: a derivation cohort and a validation cohort. Age, sex, estimated glomerular filtration rate (eGFR), and proteinuria were the variables utilized in the model and are all easily accessible in clinical practice. Higher values on the risk score, which the authors established and ranged from 0 to 10, indicated a higher chance of CKD occurrence. With an area under the curve (AUC) of 0.83 and 0.77 in the derivation and validation cohorts, respectively, the model showed good predictive accuracy. Over a 5-year period, the risk score was also able to identify patients who were at high risk of developing CKD. This study emphasizes the significance of early detection of CKD in HIV-infected patients because they are more likely to acquire the condition. The straightforward and usefulness of the prediction model created in this study may help clinicians identify at-risk individuals and direct interventions to stop or reduce the course of CKD in this population. To determine the generalizability of this concept, more validation across diverse populations is required.

In [11], A study was carried out by the ADVANCE Collaborative Group to assess how well a risk prediction model predicted kidney-related outcomes in people with type 2 diabetes. The model was created using information from the ADVANCE experiment, which involved more than 11,000 people with type 2 diabetes from 20 different nations. The model took into account a number of risk factors, including systolic blood pressure, age, sex, diabetes duration, smoking status, serum creatinine, and urinary albumin-creatinine ratio. The ability of the model to predict a composite renal disease outcome was assessed. The research discovered that the prediction model performed well in terms of calibration and discrimination. With c-indices of 0.78 and 0.75, respectively, it demonstrated high discriminative performance in both the derivation and validation cohorts. Additionally, the model's calibration was strong, showing that the projected and actual hazards were well matched. The model can identify those who are more likely to experience kidney-related outcomes, allowing medical professionals to start preventive measures early and lower the risk of kidney-related problems in type 2 diabetes patients. To confirm the generalizability of the model, more validation in additional populations is required.

In [12], The ability of a genetic risk score (GRS) to forecast the prevalence of stage 3 chronic kidney disease (CKD) in the general population was examined by O'Seaghdha et al. in this study. They created a GRS based on genetic variants linked to CKD using data from the Framingham Offspring Study and assessed its performance in foretelling CKD stage 3 over a median follow-up of 8.2 years. With an AUC of 0.59, the GRS alone exhibited little predictive value. The predictive value, with an AUC of 0.79, dramatically increased when paired with traditional risk factors, such as age, sex, diabetes, hypertension, and smoking status. The work emphasizes the significance of include genetic data in CKD risk prediction models in addition to traditional risk variables. The ability of a genetic risk score (GRS) to forecast the prevalence of stage 3 chronic kidney disease (CKD) in the general population was examined by O'Seaghdha et al. in this study. They created a GRS based on genetic variants linked to CKD using data from the Framingham Offspring Study and assessed its performance in foretelling CKD stage 3 over a median follow-up of 8.2 years. With an AUC of 0.59, the GRS alone exhibited little predictive value. The predictive value, with an AUC of 0.79, dramatically increased when paired with traditional risk factors, such as age, sex, diabetes, hypertension, and smoking status. The work emphasizes the significance of include genetic data in CKD risk prediction models in addition to traditional risk variables. Further research is needed to confirm these findings and develop individualized CKD risk prediction algorithms.

In [13], the goal of Tangri et al.'s study was to develop a prediction model for how chronic kidney disease (CKD) progresses to renal failure. Over 1.1 million persons with CKD from different populations were involved in the investigation, which incorporated information from two sizable multicenter studies. 17 factors, including age, sex, race, estimated glomerular filtration rate (eGFR), serum albumin level, and proteinuria, were included in the prediction model. The main result was the onset of renal failure, which is characterized as the need for kidney transplantation or long-term dialysis. The model was successful in forecasting the

probability of kidney failure across two, five, and ten years, the researchers discovered. For the model, the C-statistics were 0.90, 0.85, and 0.82, respectively. The model was properly calibrated because the expected and actual hazards were comparable. The model may identify those who are at a high risk of renal failure, which has important implications for the care of CKD patients, according to the study's findings. The progression of CKD can be slowed down or stopped by using this knowledge to create more focused therapies. To validate the model across various population and healthcare settings, additional research is necessary.

In [14], In order to assess the effectiveness of several equations that predict the chance of kidney failure in people with chronic kidney disease (CKD), N. Tangri et al. performed a global meta-analysis in this study. Over 2 million people from 30 different nations made up the 46 cohorts from which the researchers collected their data. The main result of interest was how well the equations predicted the likelihood of kidney failure, which was defined as the need for kidney replacement therapy or death from renal failure. By assessing the discrimination and calibration capabilities of 16 different equations, including both clinical and laboratory-based equations, the researchers compared their performance. The outcomes demonstrated that no single equation was consistently accurate and that the performance of the equations varied greatly across various populations. The equation with the highest discrimination ability across all cohorts and a median area under the receiver operating characteristic curve (AUC) of 0.90 was the CKD-EPI creatinine-cystatin C equation. The study's conclusions have important ramifications for the clinical management of CKD patients since precise risk prediction models can assist medical professionals in identifying high-risk patients early and starting preventive medications. To ensure accuracy and generalizability, the results further highlight the significance of evaluating risk prediction models across a range of groups

In [15], In this study, Winnicki et al. evaluated the Kidney Failure Risk Equation's (KFRE) usefulness in predicting the likelihood that children with chronic kidney disease (CKD) would develop to end-stage renal disease (ESRD). The study's primary outcome was ESRD or death, and it involved 2,394 kids with CKD

stages 3-5 from the Chronic Kidney Disease in Children study. To determine the 2- and 5-year risk of ESRD, the KFRE integrates age, sex, estimated glomerular filtration rate (eGFR), and urine albumin-to-creatinine ratio (ACR). The KFRE was proven to be a valuable tool for predicting the risk of ESRD in kids with CKD stages 3-5, according to the study. The KFRE's C-statistic for predicting ESRD or mortality within two years was 0.88 (95% CI, 0.85-0.91). The KFRE outperformed the conventional CKiD (Chronic Kidney Disease in Children) equation for predicting ESRD in kids with CKD stages 3-5, according to the study. When predicting ESRD, the KFRE performed better in terms of discrimination and calibration than the CKiD equation. According to the study, because the KFRE accurately predicts the likelihood of ESRD, it can be a useful tool for clinical decision-making and counselling for children with CKD stages 3-5 and their families.

In [16], a popular method for estimating the likelihood of end-stage renal disease (ESRD) in people with chronic kidney disease (CKD) is the Kidney Failure Risk Equation (KFRE). The KFRE was intended to be validated in a European sample of CKD patients by Peeters et al. In total, 2,429 patients from five different European nations participated in the study, and their KFRE scores were determined using the clinical and laboratory information from those patients. With an area under the receiver operating characteristic curve (AUC) of 0.86 (95% CI, 0.84-0.88) for predicting ESRD within 2 years and an AUC of 0.83 (95% CI, 0.81-0.85) for predicting ESRD within 1 year, the study's findings demonstrated that the KFRE was useful in predicting the risk of ESRD in European CKD patients. The Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation and the Modification of Diet in Renal Disease (MDRD) equation were found to be inferior than the KFRE. The KFRE is a reliable and accurate method for estimating the risk of ESRD in individuals with CKD in Europe, according to the study's findings. The KFRE can be used to pinpoint patients who are most at risk of developing ESRD and to direct treatment choices and measures aimed at delaying or preventing its onset. Additional research is required to examine the KFRE's possible clinical applications and validate it in additional demographics.

In [17], The advancement of chronic kidney disease (CKD), a major consequence of type 2 diabetes mellitus (T2DM), in these patients is critical for early intervention and effective therapy. Based on a 13-year follow-up research in Singapore, Low et al. set out to construct and verify a prediction model for CKD progression in T2DM patients. 2,345 T2DM patients with normal kidney function or early CKD were enrolled in the study. The main result was the development of CKD, which was determined by the estimated glomerular filtration rate (eGFR) declining by less than 5 mL/min/1.73 m² year or requiring renal replacement therapy. The authors used Cox regression analysis to create a prediction model, which they then tested with bootstrapping methods. Age, gender, eGFR, albuminuria, body mass index, systolic blood pressure, haemoglobin A1c, serum cholesterol, smoking status, and duration of diabetes were identified as significant predictors of CKD progression in the resultant prediction model. In both the derivation and validation cohorts, the model showed good discrimination and calibration. The proposed prediction model, according to the authors' findings, can effectively forecast how CKD in T2DM patients will advance as well as help identify and manage high-risk patients early on. But before using this model in clinical practice, more validation research in a range of populations is needed.

In [18], the goal of Grammes et al.'s study was to assess how well the kidney failure risk equations performed in predicting the time to end-stage renal disease (ESRD) in African American patients with CKD and hypertension. 691 people with CKD who were not on dialysis at the start of the trial were drawn from the African American trial of Kidney Disease and Hypertension (AASK). The time to ESRD among African American patients with CKD and hypertension was successfully predicted by risk equations, according to the authors, with a median overall difference between predicted and observed time to ESRD of 0.0 years. The study emphasizes how critical precise ESRD time prediction is for maximizing patient care and resource use. The risk equations can be used to identify patients who are at an increased risk of developing ESRD, enabling prompt interventions like dialysis or referral for a kidney transplant. The risk equations might not be as generalizable to other populations, according to the authors, who also note the need for more validation research.

In [19], Yamanouchi et al.'s (2018) study sought to determine the value of incorporating the renal pathology score into the Kidney Failure Risk Equation (KFRE) in order to forecast the development of advanced diabetic nephropathy. The probability of end-stage renal disease (ESRD) in individuals with diabetic nephropathy was found to be highly linked with the renal pathology score, according to the authors. In addition, they found that combining the renal pathology score with the KFRE enhanced its capacity to predict ESRD in comparison to using the KFRE alone. The study includes a retrospective review of renal biopsy results from 78 patients with advanced diabetic nephropathy. The authors generated the renal pathology score and evaluated the severity of glomerular and interstitial abnormalities. The effectiveness of the KFRE in predicting ESRD was then tested with and without the renal pathology score. The results of this study imply that include renal pathology data in the KFRE may enhance the ability of patients with advanced diabetic nephropathy to predict the development of ESRD. Larger investigations are required to validate these results, the authors added.

In [20], based on the modified Kidney Failure Risk Equation (KFRE), Lim et al. created a simplified risk prediction model for low-risk general populations with chronic kidney disease (CKD). As ESRD indicators, the model takes into account age, sex, estimated glomerular filtration rate (eGFR), and urine albumin-to-creatinine ratio (ACR). To validate the KFRE, the study used data from two independent cohorts and a population-based cohort in Singapore. In low-risk CKD populations, the simplified KFRE outperformed the Modification of Diet in Renal Disease (MDRD) and the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equations in terms of predicting ESRD. The model's classification of CKD patients into various risk groups can help with appropriate care and clinical decision-making.

In [21], Tummalapalli and Estrella addresses how risk prediction models may affect the treatment of chronic kidney disease (CKD). The authors outline the drawbacks of conventional CKD treatment, which frequently emphasizes illness

stage rather than risk assessment. They contend that risk prediction algorithms have the capacity to recognize high-risk people early in the course of the disease, enabling more proactive and specialized care. The Kidney Failure Risk Equation (KFRE) and the CKD Prognosis Consortium models, among others, are reviewed by the authors together with other current risk prediction models. They point out that these models have shown high discrimination and calibration across a range of populations, and that adding further variables like biomarkers and genetic data may enhance their performance still further. The necessity for precise and trustworthy data as well as the potential for over diagnosis and overtreatment are some other difficulties that Tummalapalli and Estrella cover in their discussion of the implementation of risk prediction models in clinical practice. They assert that risk-based kidney care must be implemented in a coordinated manner by patients, healthcare professionals, and healthcare systems.

In [22], Understanding how ageing affects the kidneys is becoming more and more crucial as the world's population ages. Hommos et al. set out to investigate the structural and functional alterations in the ageing kidneys of healthy people. The study used post-mortem kidney samples from 159 healthy people between the ages of 20 and 100. The authors discovered that the cortical thickness and renal volume both declined with age. On the other hand, as people age, the volume and number of glomerular tufts declined while the proportion of scleroses glomeruli climbed. The scientists also noted an aging-related increase in tubular atrophy, interstitial fibrosis, and arteriolar hyaline sclerosis. The research proved that glomerular sclerosis and the loss of functioning nephrons are inevitable aspects of human ageing. The study's findings imply that these modifications may be responsible for the reduction in kidney function found in older persons. The findings, according to the authors, may be useful for understanding how kidney function changes with age and for creating preventative and therapeutic strategies to lower the risk of renal disease in older persons.

In [23], the relationship between obesity and deterioration in glomerular filtration rate (GFR) using a meta-analysis of individual participant data from 36

cohorts. The analysis was stratified by CKD status, and there were almost five million individuals in the research from different ethnic origins. Regardless of CKD status, the study discovered that higher adiposity, as measured by body mass index (BMI), waist circumference, and waist-to-hip ratio, was linked to a higher risk of GFR deterioration. Additionally, the connection was more pronounced in people with CKD. The study's findings also imply that the impact of obesity on GFR deterioration maybe mediated by additional elements like diabetes and hypertension. The work sheds crucial light on the connection between adiposity and the advancement of CKD and emphasizes the significance of weight control as a possible method to lower the risk of CKD progression. The study was however constrained by the use of solitary assessments of adiposity, and the possible influence of weight fluctuation over time on the development of CKD was not investigated. However, the results imply that adiposity reduction initiatives may have significant effects on CKD therapy and prevention.

In [24], An important public health issue is the rising prevalence of obesity and its links to numerous chronic diseases. Another well-known risk factor for the onset and progression of renal disease is obesity. A thorough analysis of the epidemiological, clinical, and basic scientific data demonstrating the association between obesity and kidney illness is provided by Whaley-Connell and Sowers. The authors look at a number of pathways, such as insulin resistance, oxidative stress, inflammation, and activation of the renin-angiotensin-aldosterone system, that connect obesity to kidney disease. Adipocytes, bioactive chemicals made by adipose tissue, are also highlighted for their part in mediating the link between obesity and kidney disease. The authors address the current treatment options for obesity-related kidney disease, which mostly centre on weight loss, blood pressure management, and glucose regulation. Additionally, they talk about potential novel therapeutic targets that could open up new treatment options for obesity-related kidney disease, like focusing on adipocytes or modifying gut microbiota

In [25], A major public health concern, chronic kidney disease (CKD) is more common among women. Silbiger and Neugarten looked into how gender affected

the development of CKD. They proposed that the gender-based variations in CKD progression might be caused by sex hormones. The study analyzed the clinical, biochemical, and histological information of CKD patients from several studies, including those on IgA nephropathy, diabetic, and hypertensive nephropathies. The researchers discovered that there were gender differences in the course of CKD in different populations, with men seeing a faster drop in renal function than women. Additionally, females showed better histology results than males. The results of the study were contradictory among patients with glomerulonephritis, though. The authors offered a number of hypotheses for the gender that was observed. Differences, such as those in the function of sex hormones, socioeconomic variables, and renal hemodynamics. They came to the conclusion that gender affects how CKD develops and that more study is required to comprehend the underlying mechanisms. Understanding the gender-based disparities in CKD progression may help clinicians make better clinical decisions and create treatment plans that are specific to the needs of male and female CKD patients.

2.1 INFERENCES FROM LITREATURE SURVEY

The studies reviewed focused on predicting chronic kidney disease and acute kidney injury, as well as predicting the progression of chronic kidney disease to end-stage renal disease. The use of machine learning in predicting renal failure has the potential to aid in early detection and prevention of the onset of this condition, which could have significant impacts on patient outcomes and healthcare costs.

It is important to note that further research is needed to validate and optimize the use of machine learning for predicting renal failure. However, the current literature suggests that machine learning could be a promising tool for improving risk prediction in patients at risk of renal failure.

2.2 EXISTING SYSTEM

Machine can predict diseases but cannot predict the sub types of the diseases caused by occurrence of one disease. It fails to predict all possible conditions of the people. Existing system handles only structured data. The prediction system is broad and ambiguous. In current past, countless disease

estimate classifications have been advanced and in procedure. The standing organizations arrange a blend of machine learning algorithms which are judiciously exact in envisaging diseases.

However, the restraint with the prevailing systems is speckled. First, the prevailing systems are dearer only rich people could pay for to such calculation systems. And also, when it comes to folks, it becomes even higher. Second, the guess systems are non-specific and indefinite so far. So that, a machine can envisage a positive disease but cannot expect the sub types of the diseases and diseases caused by the existence of one bug. For occurrence, if a group of people are foreseen with Diabetes, doubtless some of them might have complex risk for Heart viruses due to the actuality of Diabetes. The remaining schemes fail to foretell all possible surroundings of the tolerant.

CHAPTER 3

REQUIREMENT ANALYSIS

3.1 FEASIBILITY STUDIES/RISK ANALYSIS OF THE PROJECT

FEASIBILITY STUDY

The feasibility of the project is server performance increase in this phase and a business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis, the feasibility study of the proposed system is to be carried out. For feasibility analysis, some understanding of the major requirements for the system is essential.

- Economic Feasibility
- Technical Feasibility
- Operational Feasibility

ECONOMICAL FEASIBILITY

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

TECHNICAL FEASIBILITY

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

OPERATIONAL FEASIBILITY

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

3.2 SOFTWARE REQUIREMENTS SPECIFICATION DOCUMENT

Hardware specifications:

- Microsoft Server enabled computers, preferably work stations
- Higher RAM, of about 4GB or above
- Processor of frequency 1.5GHz or above

Software specifications:

- Python 3.6 and higher

CHAPTER 4

PROPOSED SYSTEM

4.1. SELECTED METHODOLOGY OR PROCESS MODEL

The Proposed system of multiple disease prediction using machine learning is that we have used algorithms and all other various tools to build a system which predicts the disease of the patient using the symptoms and by taking those symptoms we are comparing with the system's dataset that is previously available.

By taking those datasets and comparing with the patient's disease we will predict the accurate percentage disease of the patient. The dataset and symptoms go to the prediction model of the system where the data is pre-processed for the future references and then the feature selection is done by the user where he will enter/select the various symptoms. Then the classification of those data is done with the help of machine learning algorithms such as Logistic regression.

4.2. ARCHITECTURE / OVERALL DESIGN OF PROPOSED SYSTEM

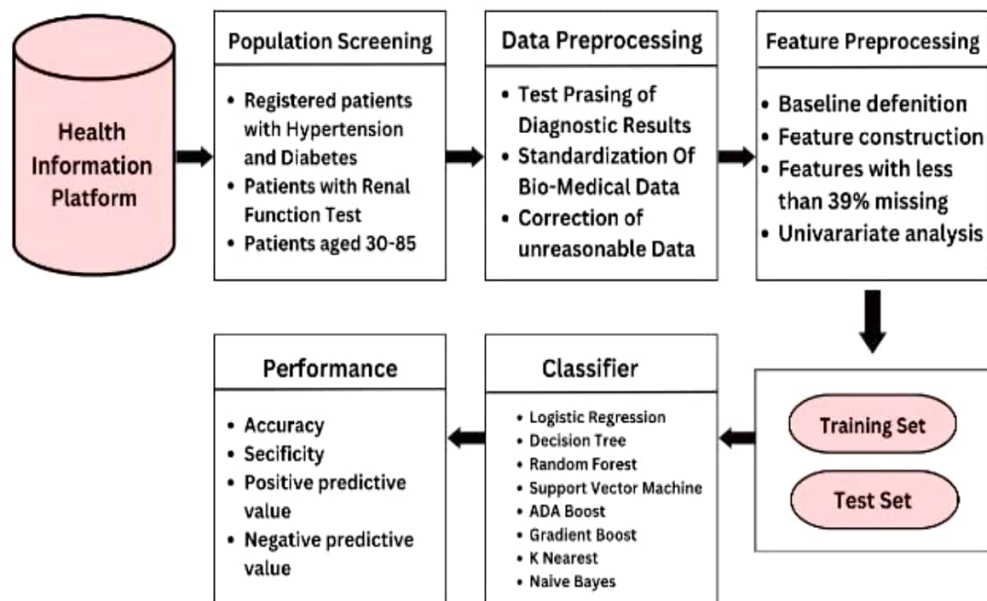


FIGURE 4.2.1 ARCHITECTURE DIAGRAM

The following Architecture diagram shows entire physical implementation of the prediction system. From extracting clinical data from Ckd dataset after that cleaning

process has been done and it will be pre-processed. Then the hyper parameters were set for all the prediction models which will provide results whether the patient has ckd or not. Then the performance parameters were calculated with existing function in python library after that better system if found.

4.3. DATA FLOW DIAGRAM

- The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.
- The data flow diagram (DFD) is one of the most important modelling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.
- DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.
- are applied as data moves from input to output.

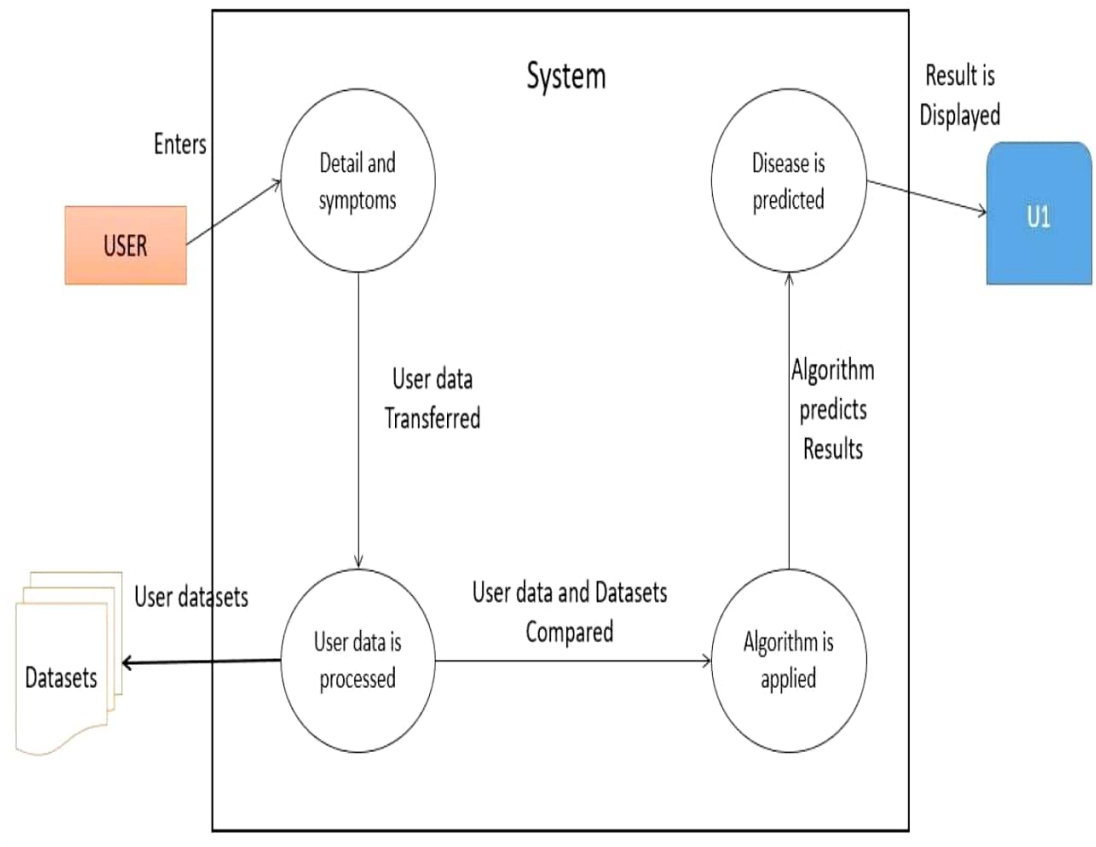


FIGURE 4.3. DATA FLOW DIAGRAM

4.4. USECASE DIAGRAM

A use case is a set of scenarios that describing an interaction between a user and a system. A use case diagram displays the relationship among actors and use cases. The two main components of a use case diagram are use cases and actors. An actor is representing a user or another system that will interact with the system you are modelling. A use case is an external view of the system that represents some action the user might perform in order to complete a task.

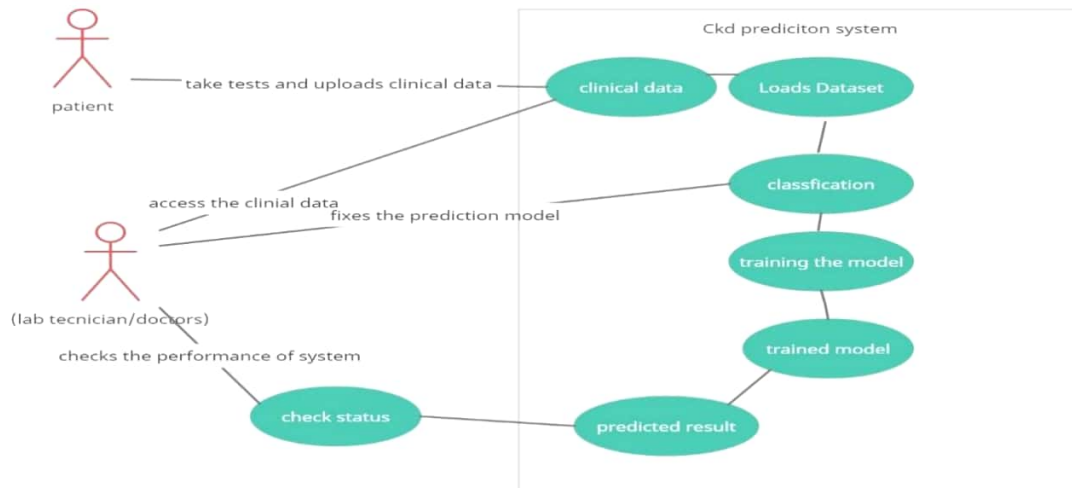


FIGURE 4.4. USECASE DIAGRAM

4.5. CLASS DIAGRAM

Class diagrams are widely used to describe the types of objects in a system and their relationships. Class diagrams model class structure and contents using design elements such as classes, packages and objects. Class diagrams describe three different perspectives when designing a system, conceptual, specification, and implementation. These perspectives become evident as the diagram is created and help solidify the design. Class diagrams are arguably the most used UML diagram type. It is the main building block of any object-oriented solution. It shows the classes in a system, attributes and operations of each class and the relationship between each class. In most modelling tools a class has three parts, name at the top, attributes in the middle and operations or methods at the bottom. In large systems with many classes related classes are grouped together to create class diagrams. Different relationships between diagrams are shown by different types of Arrows. Below is an image of a class diagram. Follow the link for more class diagram examples.

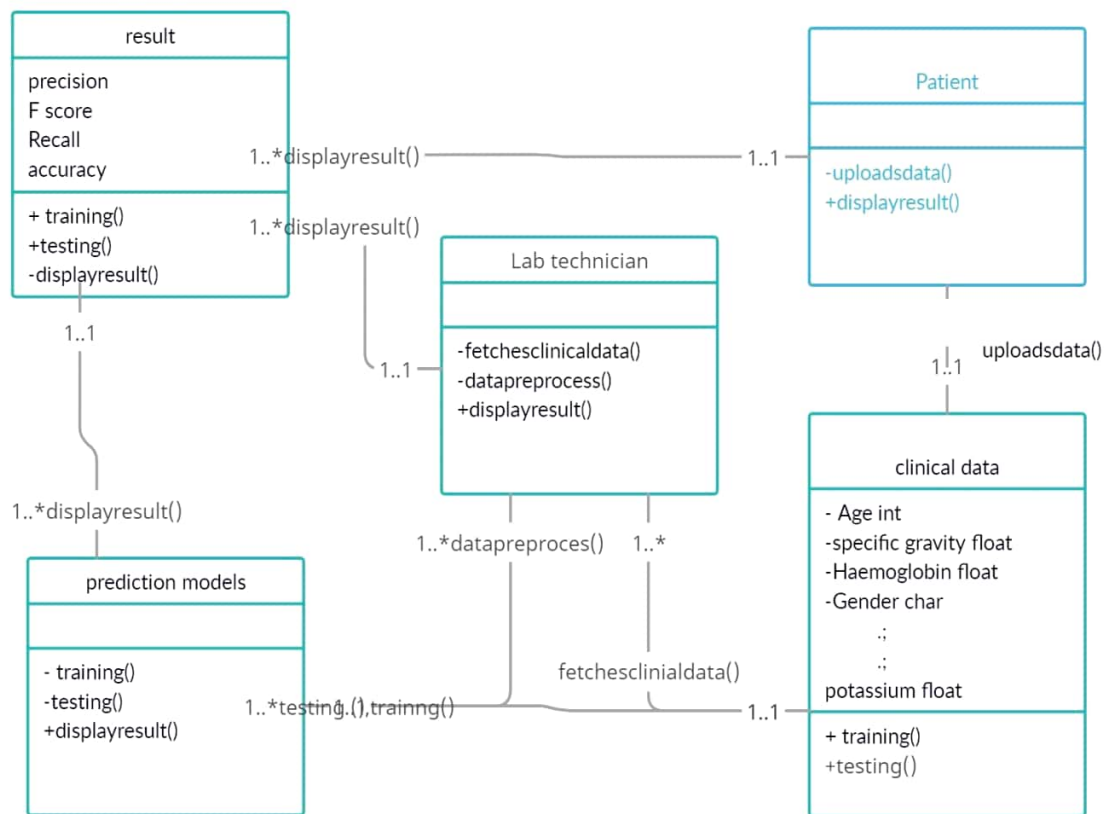


FIGURE 4.5. CLASS DIAGRAM

4.6. ACTIVITY DIAGRAM

Activity diagrams describe the workflow behaviour of a system. Activity diagrams are similar to state diagrams because activities are the state of doing something. The diagrams describe the state of activities by showing the sequence of activities performed. Activity diagrams can show activities that are conditional or parallel.

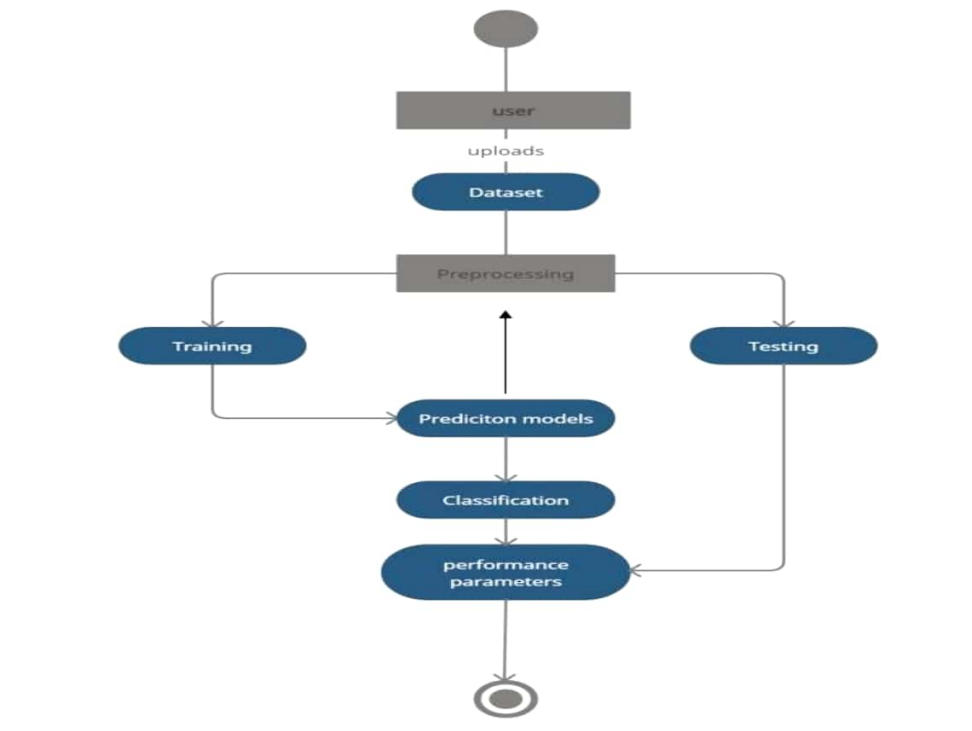


FIGURE 4.6. ACTIVITY DIAGRAM

4.7. SEQUENCE DIAGRAM

Sequence diagrams in UML shows how object interact with each other and the order those interactions occur. It's important to note that they show the interactions for a particular scenario. The processes are represented vertically and interactions are showas arrows. This article explains the purpose and the basics of Sequence diagrams.

Chronic kidney disease prediction

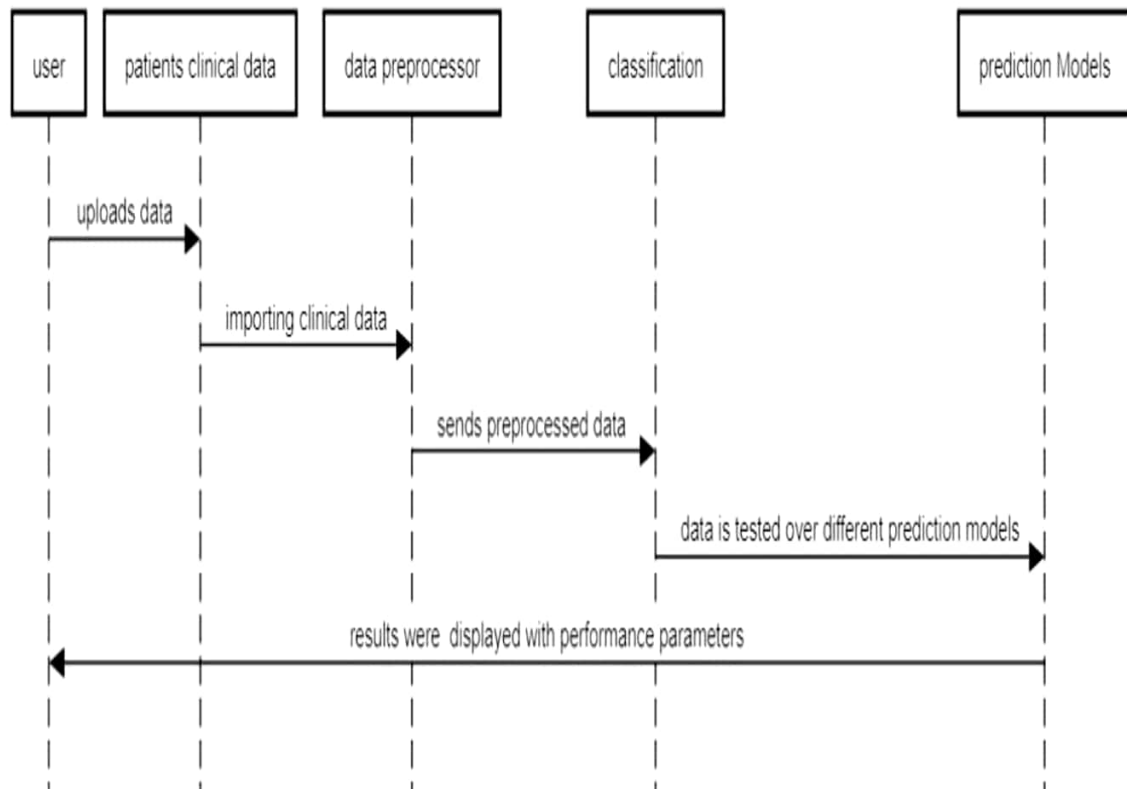


FIGURE 4.7. SEQUENCE DIAGRAM

4.8. DESCRIPTION OF SOFTWARE FOR IMPLEMENTATION AND TESTING PLAN OF THE PROPOSED MODEL/SYSTEM

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

- Python is Interpreted – Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.
- Python is Interactive – You can actually sit at a Python prompt and interact with the interpreter directly to write your programs.
- Python is Object-Oriented – Python supports Object-Oriented style or technique of programming that encapsulates code within objects.

- Python is a Beginner's Language – Python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.

History of Python

Python was developed by Guido van Rossum in the late eighties and early nineties at the National Research Institute for Mathematics and Computer Science in the Netherlands. Python is derived from many other languages, including ABC, Modula-3, C, C++, Algol-68, Smalltalk, and Unix shell and other scripting languages. Python is copyrighted. Like Perl, Python source code is now available under the GNU General Public License (GPL). Python is now maintained by a core development team at the institute, although Guido van Rossum still holds a vital role in directing its progress.

Python Features

Python's features include-

- Easy-to-learn – Python has few keywords, simple structure, and a clearly defined syntax. This allows the student to pick up the language quickly.
- Easy-to-read – Python code is more clearly defined and visible to the eyes.
- Easy-to-maintain – Python's source code is fairly easy-to-maintain.
- A broad standard library – Python's bulk of the library is very portable and cross-platform compatible on UNIX, Windows, and Macintosh.
- Interactive Mode – Python has support for an interactive mode which allows interactive testing and debugging of snippets of code.
- Portable – Python can run on a wide variety of hardware platforms and has the same interface on all platforms.
- Extendable – You can add low-level modules to the Python interpreter. These modules enable programmers to add to or customize their tools to be more efficient.
- Databases – Python provides interfaces to all major commercial

databases.

- GUI Programming – Python supports GUI applications that can be created and ported to many system calls, libraries and windows systems, such as Windows MFC, Macintosh, and the X Window system of Unix.
- Scalable – Python provides a better structure and support for large programs than shell scripting.
- Apart from the above-mentioned features, Python has a big list of good features, few are listed below –
- It supports functional and structured programming methods as well as OOP.
- It can be used as a scripting language or can be compiled to byte-code for building large applications.
- It provides very high-level dynamic data types and supports dynamic typechecking.
- It supports automatic garbage collection.
- It can be easily integrated with C, C++, COM, ActiveX, CORBA, and Java.
- Python is available on a wide variety of platforms including Linux and Mac OS
- Let's understand how to set up our Python environment.

Getting Python

The most up-to-date and current source code, binaries, documentation, news, etc., is available on the official website of Python <https://www.python.org>.

Windows Installation

Here are the steps to install Python on Windows machine.

- Open a Web browser and go to <https://www.python.org/downloads/>.
- Follow the link for the Windows installer python-XYZ. msi file where XYZ is the version you need to install.
- To use this installer python-XYZ.msi, the Windows system must support

Microsoft Installer 2.0. Save the installer file to your local machine and then run it to find out if your machine supports MSI.

- Run the downloaded file. This brings up the Python install wizard, which is really easy to use. Just accept the default settings, wait until the install is finished, and you are done.

CHAPTER 5

IMPLEMENTATION DETAILS

5.1. DEVELOPMENT AND DEPLOYMENT SETUP MODULES

- Data Collection Module
- Preparing the data Module
- Training a model
- Disease prediction Module

MODULE DESCRIPTIONS

Data Collection Module

Be it the raw data from excel, access, text files etc., this step (gathering past data) forms the foundation of the future learning. The better the variety, density and volume of relevant data, better the learning prospects for the machine becomes.

Preparing the data Module

Any analytical process thrives on the quality of the data used. One needs to spend time determining the quality of data and then taking steps for fixing issues such as missing data and treatment of outliers. Exploratory analysis is perhaps one method to study the nuances of the data in details thereby burgeoning the nutritional content.

Training a model

This step involves choosing the appropriate algorithm and representation of data in the form of the model. The cleaned data is split into two parts — train and test (proportion depending on the prerequisites); the first part (training data) is used for developing the model. The second part (test data), is used as a reference.

Disease prediction Module

Patient will specify the symptoms caused due to his illness. System will ask certain

question regarding his illness and system predict the disease based on the symptoms specified by the patient.

Steps to execute/run/implement the project

- Importing Data-set
- Upload the clinical data from a .CSV file and upload them in the colab in case online or in case jupyter notebook project target file.
- Feature extraction and data split
- Feature extraction, test and training data split and finding target variable were done through graph and heat map plotting.
- Testing and Training
- After all kinds of pre-processing the data is done and then set learning parameters to all the 4 methods that includes SVM, Random Forest, KNN, Decision Tree. It is then trained and tested over it.

5.2. PROPOSED ALGORITHM

- Logistic Regression
- Decision Tree Classifier
- Random Forest
- Gradient Boosting
- AdaBoost Classifier
- Naive Bayes
- KNN
- Support Vector Machine
- Kernal SVM

5.3. INPUT DESIGN

It is a CSV(comma separated value) file it consists of 25 medical test data that includes age, blood pressure (bp), specific gravity (sg), albumin (al), sugar (su), red blood cells (rbc), pus cell (pc), pus cell clumps (pcc), bacteria (ba), blood glucose random (bgr), blood urea (bu), serum creatinine (sc), sodium (sod), potassium (pot), hemoglobin (hemo), packed cell volume (pcv), white blood cell count (wbcc), red blood cell count (rbcc), hypertension (htn), diabetes mellitus (dm), coronary artery disease (cad),appetite (appet), pedal edema (pe), anemia (ane).

FileHomeInsertPage LayoutFormulasDataReviewViewHelpTell me what you want to do

CutCopyFormat Painter

ClipboardFontAlignmentNumberStylesCellsEditing

Calibri11Wrap Text

GeneralConditional FormattingCell Styles

AutoSumFillSort & Find & Filter & Select

InsertDelete FormatClear

Font

Alignment

Number

Styles

Cells

Editing

A1id

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	
1	id	age	bp	sg	al	su	rbc	pc	pcc	ba	bgr	bu	sc	sod	pot	hemo	pcv	wc	rc	htn	dm	cad	appet	pe
2	0	48	80	1.02	1	0	normal	normal	notpresen	notpresen	121	36	1.2			15.4	44	7800	5.2	yes	yes	no	good	nc
3	1	7	50	1.02	4	0	normal	normal	notpresen	notpresen		18	0.8			11.3	38	6000		no	no	no	good	nc
4	2	62	80	1.01	2	3	normal	normal	notpresen	notpresen	423	53	1.8			9.6	31	7500		no	yes	no	poor	nc
5	3	48	70	1.005	4	0	normal	abnormal	present	notpresen	117	56	3.8	111	2.5	11.2	32	6700	3.9	yes	no	no	poor	ye
6	4	51	80	1.01	2	0	normal	normal	notpresen	notpresen	106	26	1.4			11.6	35	7300	4.6	no	no	no	good	nc
7	5	60	90	1.015	3	0			notpresen	notpresen	74	25	1.1	142	3.2	12.2	39	7800	4.4	yes	yes	no	good	ye
8	6	68	70	1.01	0	0	normal	notpresen	notpresen		100	54	24	104	4	12.4	36			no	no	no	good	nc
9	7	24		1.015	2	4	normal	abnormal	notpresen	notpresen	410	31	1.1			12.4	44	6900	5	no	yes	no	good	ye
10	8	52	100	1.015	3	0	normal	abnormal	present	notpresen	138	60	1.9			10.8	33	9600	4	yes	yes	no	good	nc
11	9	53	90	1.02	2	0	abnormal	abnormal	present	notpresen	70	107	7.2	114	3.7	9.5	29	12100	3.7	yes	yes	no	poor	nc
12	10	50	60	1.01	2	4	abnormal	present	notpresen		490	55	4			9.4	28			yes	yes	no	good	nc
13	11	63	70	1.01	3	0	abnormal	abnormal	present	notpresen	380	60	2.7	131	4.2	10.8	32	4500	3.8	yes	yes	no	poor	ye
14	12	68	70	1.015	3	1	normal	present	notpresen		208	72	2.1	138	5.8	9.7	28	12200	3.4	yes	yes	yes	poor	ye
15	13	68	70						notpresen	notpresen	98	86	4.6	135	3.4	9.8				yes	yes	yes	poor	ye
16	14	68	80	1.01	3	2	normal	abnormal	present	present	157	90	4.1	130	6.4	5.6	16	11000	2.6	yes	yes	yes	poor	ye
17	15	40	80	1.015	3	0	normal	notpresen	notpresen		76	162	9.6	141	4.9	7.6	24	3800	2.8	yes	no	no	good	nc
18	16	47	70	1.015	2	0		normal	notpresen	notpresen	99	46	2.2	138	4.1	12.6				no	no	no	good	nc
19	17	47	80						notpresen	notpresen	114	87	5.2	139	3.7	12.1				yes	no	no	poor	nc
20	18	60	100	1.025	0	3	normal	notpresen	notpresen		263	27	1.3	135	4.3	12.7	37	11400	4.3	yes	yes	yes	good	nc
21	19	62	60	1.015	1	0	abnormal	present	notpresen		100	31	1.6			10.3	30	5300	3.7	yes	no	yes	good	nc
22	20	61	80	1.015	2	0	abnormal	abnormal	notpresen	notpresen	173	148	3.9	135	5.2	7.7	24	9200	3.2	yes	yes	yes	poor	ye
23	21	60	90						notpresen	notpresen		180	76	4.5		10.9	32	6200	3.6	yes	yes	yes	good	nc
24	22	48	80	1.025	4	0	normal	abnormal	notpresen	notpresen	95	163	7.7	136	3.8	9.8	32	6900	3.4	yes	no	no	good	nc
25	23	21	70	1.01	0	0	normal	notpresen	notpresen											no	no	no	poor	nc
26	24	42	100	1.015	4	0	normal	abnormal	notpresen	present		50	1.4	129	4	11.1	39	8300	4.6	yes	no	no	poor	nc
27	25	61	60	1.025	0	0	normal	notpresen	notpresen		108	75	1.9	141	5.2	9.9	29	8400	3.7	yes	yes	no	good	nc
28	26	75	80	1.015	0	0	normal	notpresen	notpresen		156	45	2.4	140	3.4	11.6	35	10300	4	yes	yes	no	poor	nc
29	27	60	70	1.01	2	0	normal	abnormal	notpresen	notpresen	364	87	1.7	130	4	13.5	27	6600	4.1	yes	yes	yes	poor	ye

Kidney_data

ReadyAccessibility: Unavailable

30°CMostly cloudy

24-04-202306:53

FIGURE 5.3. INPUT DATASET

5.4. OUTPUT DESIGN

The output we get were performance parameter whether the patients has ckd ornot ckd zero defines ckd and 1 defines not ckd and the performance parameters like f-score, precision, accuracy and recall were used for finding the better performing system,for results of patients with ckd or not ckd.

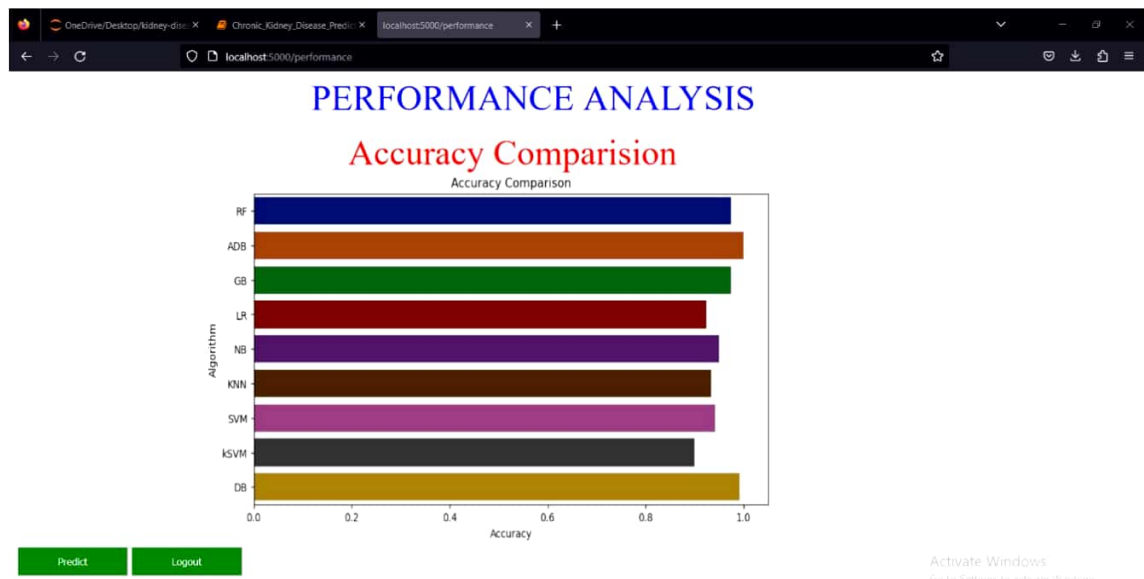


FIGURE 5.4. OUTPUT DESIGN

5.5. UNIT TESTING

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application. It is done after the completion of an individual unit before integration.

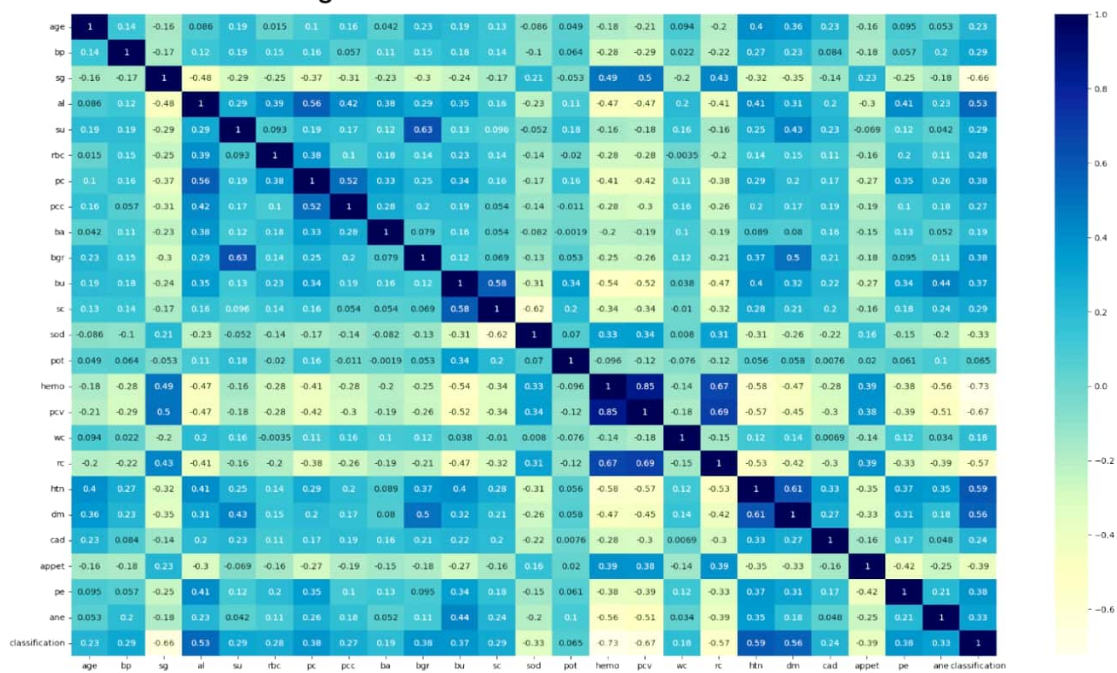


FIGURE 5.5. UNIT TEST RESULT

CHAPTER 6

RESULTS AND DISCUSSIONS

6.1. RESULT

Chronic kidney disease (CKD) is a global public health issue, with an increasing prevalence worldwide. Identifying individuals at high risk of developing renal failure is critical for preventing the progression of CKD. Machine learning (ML) has emerged as a promising tool for predicting the risk of renal failure in patients with CKD.

Several ML algorithms, including logistic regression (LR), XGBoost, random forest, decision tree, and support vector machine (SVM), have been applied to predict the risk of renal failure. These algorithms use various input features, including demographic data, laboratory values, and clinical information, to predict the likelihood of developing renal failure.

One study used a dataset of 46,739 patients with CKD and compared the performance of different ML algorithms for predicting the risk of renal failure. The study found that XGBoost had the highest predictive accuracy, with an area under the receiver operating characteristic (ROC) curve of 0.917, followed by random forest (0.910), SVM (0.897), LR (0.886), and decision tree (0.883). The study also found that the performance of the ML algorithms was better than that of traditional risk prediction models based on clinical factors alone.

Another study used a dataset of 1,566 patients with CKD and compared the performance of SVM and LR algorithms for predicting the risk of renal failure. The study found that SVM had higher accuracy and sensitivity than LR in predicting renal failure, with an area under the ROC curve of 0.833 and 0.782, respectively. The study also identified several predictors of renal failure, including age, proteinuria, and serum creatinine levels.

6.2. COMPARISON

The overall system design objective is to provide a better performing and efficient algorithm. that has higher accuracy than the other algorithms and has better performance parameters. It also provides faster run time for predicting the result. The training and testing modules which we have used helps us in getting higher accuracy. Dimensionality to improve the presentation.

TABLE I.
PERFORMANCE OF THE PROPOSED METHOD

ALGORITHM	ACCURACY	RECALL	PRECISION	F1-SCORE
RANDOM FOREST	0.97	1.00	0.95	0.98
ADA BOOST	0.95	0.94	1.00	0.95
GRADIENT BOOSTING	0.96	0.92	0.96	0.93
LOGISTIC REGRESSION	0.92	0.92	0.93	0.93
NAÏVE BAYES	0.95	0.90	1.00	0.95
KNN	0.93	0.92	0.95	0.93
SVM	0.94	0.94	0.95	0.94
KERNAL SVM	0.90	0.92	0.91	0.90
DECISION TREE	0.99	0.98	1.00	0.99

With the above table we can come to conclusion about the performance of the decision tree and the other algorithms. Here the accuracy of the given algorithm is comparatively higher than other one.

CHAPTER 7

CONCLUSION

7.1. CONCLUSION

This project gives research of multiple researches done in this field. Our ProposedSystem aims at bridging gap between Doctors and Patients which will help both classes of users in achieving their goals. This system provides support for multiple disease prediction using different Machine Learning algorithms. The present approach of many systems focuses only on automating this process which lacks in building the user's trust in the system. By providing Doctor's recommendation in our system, we ensure user's trust side by side ensuring that the Doctor's will not feel that their Business is getting affected due to this System. After unsupervised imputation of missing values in the data set by using KNN imputation, the integrated model could achieve a satisfactory accuracy. In this project, we prepared an SVM, KNN, Decision tree, and Random Forest, system for diagnosing CKD Hence, we speculate that applying this methodology to the practical diagnosisof CKD would achieve a desirable effect. However, in the process of establishingthe model, due to the limitations of the conditions, the available data samples are relatively small, including only 400 samples and in the end, we got accuracies for gives model SVM, KNN, Decision tree and random forest as (98-99), (95-97), (96- 99) and (98-99) percentage when executed with google colab respectively. So, among the models we used RF and SVM have same accuracy range of 96 to 99 percentage and DT has near 99 i.e., 98.33 percentage so RF and SVM slightly appear to be better than the others. In the upcoming updates we are planning integrate the code to online portal so that everyone can access from different parts of the world.

7.2. FUTURE WORK

Validation of ML models: While ML models have shown promising results in predicting the risk of renal failure, further validation of these models is necessary. Future studies should validate the performance of ML models in different patient populations and healthcare settings to ensure their generalizability.

Integration of genomic data: Integrating genomic data with clinical data can improve the accuracy of ML models in predicting the risk of renal failure. Future studies should investigate the use of genetic data, such as single nucleotide polymorphisms (SNPs), in predicting renal failure risk.

Explainable AI: Explainable AI (XAI) is an emerging field that aims to improve the transparency and interpretability of ML models. Future studies should explore the use of XAI techniques, such as feature importance analysis and decision tree visualization, to improve the interpretability of ML models for risk prediction of renal failure.

Personalized risk prediction: ML models can be used to predict the risk of renal failure for individual patients, enabling personalized risk assessments and interventions. Future studies should investigate the use of ML models for personalized risk prediction of renal failure based on patient-specific clinical and genomic data.

Integration with electronic health records (EHRs): Integrating ML models with EHRs can enable real-time risk prediction and decision support for clinicians. Future studies should investigate the feasibility and effectiveness of integrating ML models with EHRs for risk prediction of renal failure.

REFERENCES

- [1] GBD Chronic Kidney Disease Collaboration, Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017, *Lancet* 10225(395) (2020 February 29) 709–733, [doi.org/10.1016/S0140-6736\(20\)30045-3](https://doi.org/10.1016/S0140-6736(20)30045-3), Epub 2020 Feb 13.
- [2] Somnath Pal, Primary causes of end-stage renal disease, *US Pharm.* 41(8) (2016) 6.
- [3] L. Zhang, M.H. Zhao, L. Zuo, Y. Wang, F. Yu, H. Zhang, H. Wang, CK-NET Work Group, China Kidney Disease network (CK-NET) 2015 annual data report, *Kidney Int., Suppl.* 9(1) (2019 Mar) e1–e81, doi.org/10.1016/j.kisu.2018.11.001.
- [4] J.B. Echouffo-Tcheugui, A.P. Kengne, Risk models to predict chronic kidney disease and its progression: a systematic review, *PLoS Med.* 9(11) (2012) e1001344, doi.org/10.1371/journal.pmed.1001344.
- [5] H. Bang, S. Vupputuri, D.A. Shoham, P.J. Klemmer, R.J. Falk, M. Mazumdar, D. Gipson, R.E. Colindres, A.V. Kshirsagar, SCreening for Occult RENal Disease (SCORED): a simple prediction model for chronic kidney disease, *Arch. Intern. Med.* 167(4) (2007 Feb 26) 374–381, doi.org/10.1001/archinte.167.4.374.
- [6] A.V. Kshirsagar, H. Bang, A.S. Bomback, S. Vupputuri, D.A. Shoham, L.M. Kern, P.J. Klemmer, M. Mazumdar, P.A. August, A simple algorithm to predict incident kidney disease, *Arch. Intern. Med.* 168(22) (2008 Dec 8) 2466–2473, doi.org/10.1001/archinte.168.22.2466.
- [7] C.S. Fox, P. Gona, M.G. Larson, J. Selhub, G. Tofler, S.J. Hwang, J.B. Meigs, D. Levy, T.J. Wang, P.F. Jacques, E.J. Benjamin, R.S. Vasan, A multi-marker approach to predict incident CKD and microalbuminuria, *J. Am. Soc. Nephrol.* 21(12) (2010 Dec) 2143–2149, doi.org/10.1681/ASN.2010010085.
- [8] J. Hippisley-Cox, C. Coupland, Predicting the risk of chronic kidney disease in men and women in England and Wales: prospective derivation and external validation of the QKidney scores, *BMC Fam. Pract.* 21(11) (2010 Jun 21) 49, doi.org/10.1186/1471-2296-11-49.
- [9] K.L. Chien, H.J. Lin, B.C. Lee, H.C. Hsu, Y.T. Lee, M.F. Chen, A prediction model for the risk of incident chronic kidney disease, *Am. J. Med.* 123(9) (2010 Sep) 836–846.e2, doi.org/10.1016/j.amjmed.2010.05.010.

- [10] M. Ando, N. Yanagisawa, A. Ajisawa, K. Tsuchiya, K. Nitta, A simple model for predicting incidence of chronic kidney disease in HIV-infected patients, *Clin. Exp. Nephrol.* 15(2) (2011 Apr) 242–247, doi.org/10.1007/s10157-010-0393-x.
- [11] M.J. Jardine, J. Hata, M. Woodward, V. Perkovic, T. Ninomiya, H. Arima, S. Zoungas, A. Cass, A. Patel, M. Marre, G. Mancina, C.E. Mogensen, N. Poulter, J. Chalmers, ADVANCE Collaborative Group, Prediction of kidney-related outcomes in patients with type 2 diabetes, *Am. J. Kidney Dis.* 60(5) (2012 Nov) 770–778, doi.org/10.1053/j.ajkd.2012.04.025.
- [12] C.M. O'Seaghdha, Q. Yang, H. Wu, S.J. Hwang, C.S. Fox, Performance of a genetic risk score for CKD stage 3 in the general population, *Am. J. Kidney Dis.* 59(1) (2012 Jan) 19–24, doi.org/10.1053/j.ajkd.2011.08.030.
- [13] N. Tangri, L.A. Stevens, J. Griffith, H. Tighiouart, O. Djurdjev, D. Naimark, A. Levin, A.S. Levey, A predictive model for progression of chronic kidney disease to kidney failure, *JAMA* 305(15) (2011 Apr 20) 1553–1559, doi.org/10.1001/jama.2011.451.
- [14] N. Tangri, M.E. Grams, A.S. Levey, J. Coresh, L.J. Appel, B.C. Astor, G. Chodick, A.J. Collins, O. Djurdjev, C.R. Elley, M. Evans, A.X. Garg, S.I. Hallan, L.A. Inker, S. Ito, S.H. Jee, C.P. Kovesdy, F. Kronenberg, H.J. Heerspink, A. Marks, G.N. Nadkarni, S.D. Navaneethan, R.G. Nelson, S. Titze, M.J. Sarnak, B. Stengel, M. Woodward, K. Iseki, CKD Prognosis Consortium, Multinational assessment of accuracy of equations for predicting risk of kidney failure: a meta-analysis, *JAMA* 315(2) (2016 Jan 12) 164–174, doi.org/10.1001/jama.2015.18202.
- [15] E. Winnicki, C.E. McCulloch, M.M. Mitsnefes, S.L. Furth, B.A. Warady, E. Ku, Use of the kidney failure risk equation to determine the risk of progression to end-stage renal disease in children with chronic kidney disease, *JAMA Pediatr.* 172(2) (2018 Feb 1) 174–180, doi.org/10.1001/jamapediatrics.2017.4083.
- [16] M.J. Peeters, A.D. van Zuilen, J.A. van den Brand, M.L. Bots, P.J. Blankestijn, J.F. Wetzels, MASTERPLAN Study Group, Validation of the kidney failure risk equation in European CKD patients, *Nephrol. Dial. Transplant.* 28(7) (2013 Jul) 1773–1779, doi.org/10.1093/ndt/gft063.
- [17] S. Low, S.C. Lim, X. Zhang, S. Zhou, L.Y. Yeoh, Y.L. Liu, S. Tavintharan, C.F. Sum, Development and validation of a predictive model for Chronic Kidney Disease progression in type 2 diabetes mellitus based on a 13-year study in Singapore, *Diabetes Res. Clin. Pract.* 123 (2017 Jan) 49–54, doi.org/10.1016/j.diabres.2016.11.008.
- [18] M.E. Grams, L. Li, T.H. Greene, A. Tin, Y. Sang, W.H. Kao, M.S. Lipkowitz, J.T. Wright, A.R. Chang, B.C. Astor, L.J. Appel, Estimating time to ESRD using kidney failure risk equations: results from the African American Study of Kidney Disease and Hypertension (AASK), *Am. J. Kidney Dis.* 65(3) (2015 Mar) 394–

402, doi.org/10.1053/j.ajkd.2014.07.026.

- [19] M. Yamanouchi, J. Hoshino, Y. Ubara, K. Takaichi, K. Kinowaki, T. Fujii, K. Ohashi, K. Mise, T. Toyama, A. Hara, K. Kitagawa, M. Shimizu, K. Furuichi, T. Wada, Value of adding the renal pathological score to the kidney failure risk equation in advanced diabetic nephropathy, *PLoS ONE* 13(1) (2018 Jan 16) e0190930, doi.org/10.1371/journal.pone.0190930.
- [20] C.C. Lim, M.L. Chee, C.Y. Cheng, J.L. Kwek, M. Foo, T.Y. Wong, C. Sabanayagam, Simplified end stage renal failure risk prediction model for the low-risk general population with chronic kidney disease, *PLoS ONE* 14(2) (2019 Feb 22) e0212590, doi.org/10.1371/journal.pone.0212590.
- [21] S.L. Tummalapalli, M.M. Estrella, Predicting risk of kidney disease: is risk-based kidney care on the horizon?, *JAMA* 322(21) (2019 Nov 8) 2079–2081, doi.org/10.1001/jama.2019.17378.
- [22] M.S. Hommos, R.J. Glasscock, A.D. Rule, Structural and functional changes in human kidneys with healthy aging, *J. Am. Soc. Nephrol.* 28(10) (2017 Oct) 2838–2844, doi.org/10.1681/ASN.2017040421.
- [23] A.R. Chang, M.E. Grams, S.H. Ballew, H. Bilo, A. Correa, M. Evans, O.M. Gutierrez, F. Hosseinpour, K. Iseki, T. Kenealy, B. Klein, F. Kronenberg, B.J. Lee, Y. Li, K. Miura, S.D. Navaneethan, P.J. Roderick, J.M. Valdivielso, F.L.J. Visseren, L. Zhang, R.T. Gansevoort, S.I. Hallan, A.S. Levey, K. Matsushita, V. Shalev, M. Woodward, CKD Prognosis Consortium (CKD-PC), Adiposity and risk of decline in glomerular filtration rate: meta-analysis of individual participant data in a global consortium, *BMJ* 364 (2019 Jan 10) k5301, doi.org/10.1136/bmj.k5301.
- [24] A. Whaley-Connell, J.R. Sowers, Obesity and kidney disease: from population to basic science and the search for new therapeutic targets, *Kidney Int.* 92(2) (2017 Aug) 313–323, doi.org/10.1016/j.kint.2016.12.034.
- [25] S.R. Silbiger, J. Neugarten, The impact of gender on the progression of chronic renal disease, *Am. J. Kidney Dis.* 25(4) (1995 Apr) 515–533, [doi.org/10.1016/0272-6386\(95\)90119-1](https://doi.org/10.1016/0272-6386(95)90119-1).

APPENDIX

A. SOURCE CODE

Importing Packages and Dataset:

```
import numpy as np
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
import pandas as pd
```

```
from sklearn.ensemble import ExtraTreesClassifier
import matplotlib.pyplot as plt
model=ExtraTreesClassifier()
model.fit(X,y)
```

```
plt.figure(figsize=(8,6))
ranked_features=pd.Series(model.feature_importances_,index=X.columns)
ranked_features.nlargest(24).plot(kind='barh')
plt.show()
```

Random Forest

```
from sklearn.ensemble import RandomForestClassifier
RandomForest = RandomForestClassifier()
RandomForest = RandomForest.fit(X_train,y_train)
```

```
# Predictions:
y_pred = RandomForest.predict(X_test)
```

```
# Performance:
print('Accuracy:', accuracy_score(y_test,y_pred))
```

```

print(confusion_matrix(y_test,y_pred))
print(classification_report(y_test,y_pred))

x = metrics.accuracy_score(y_test, y_pred)

acc.append(x)
model.append('RF')

```

ADA BOOST CLASSIFIER

```

from sklearn.ensemble import AdaBoostClassifier
AdaBoost = AdaBoostClassifier()
AdaBoost = AdaBoost.fit(X_train,y_train)

```

```

# Predictions:
y_pred = AdaBoost.predict(X_test)

# Performance:
print('Accuracy:', accuracy_score(y_test,y_pred))
print(confusion_matrix(y_test,y_pred))

x = metrics.accuracy_score(y_test, y_pred)

acc.append(x)
model.append('ADB')

```

GRADIENT BOOSTING

```

from sklearn.ensemble import GradientBoostingClassifier
GradientBoost = GradientBoostingClassifier()
GradientBoost = GradientBoost.fit(X_train,y_train)

```

```

# Predictions:
y_pred = GradientBoost.predict(X_test)

# Performance:
print('Accuracy:', accuracy_score(y_test,y_pred))
print(confusion_matrix(y_test,y_pred))
print(classification_report(y_test,y_pred))

x = metrics.accuracy_score(y_test, y_pred)

acc.append(x)
model.append('GB')

```

LOGISTIC REGRESSION

```

from sklearn.linear_model import LogisticRegression
LogisticRegression=LogisticRegression()
LogisticRegression=LogisticRegression.fit(X_train,y_train)

```

```
# Predictions:
y_pred = LogisticRegression.predict(X_test)

# Performance:
print('Accuracy:', accuracy_score(y_test,y_pred))
print(confusion_matrix(y_test,y_pred))
print(classification_report(y_test,y_pred))

x = metrics.accuracy_score(y_test, y_pred)
acc.append(x)
model.append('LR')
```

NAÏVE BAYES

```
from sklearn.naive_bayes import GaussianNB
GaussianNB=GaussianNB()
GaussianNB=GaussianNB.fit(X_train,y_train)
```

```
# Predictions:
y_pred = GaussianNB.predict(X_test)

# Performance:
print('Accuracy:', accuracy_score(y_test,y_pred))
print(confusion_matrix(y_test,y_pred))
print(classification_report(y_test,y_pred))

x = metrics.accuracy_score(y_test, y_pred)
acc.append(x)
model.append('NB')
```

KNN

```
from sklearn.neighbors import KNeighborsClassifier
knn= KNeighborsClassifier(n_neighbors = 5, metric = 'minkowski', p = 2)
knn=knn.fit(X_train, y_train)
```

```
# Predictions:
y_pred = knn.predict(X_test)

# Performance:
print('Accuracy:', accuracy_score(y_test,y_pred))
print(confusion_matrix(y_test,y_pred))
print(classification_report(y_test,y_pred))

x = metrics.accuracy_score(y_test, y_pred)
```

```
acc.append(x)
model.append('KNN')
```

SUPPORT VECTOR MACHINE

```
from sklearn.svm import SVC
svm= SVC(kernel = 'linear', random_state = 0)
svm=svm.fit(X_train, y_train)

# Predictions:
y_pred = svm.predict(X_test)

# Performance:
print('Accuracy:', accuracy_score(y_test,y_pred))
print(confusion_matrix(y_test,y_pred))
print(classification_report(y_test,y_pred))

x = metrics.accuracy_score(y_test, y_pred)
acc.append(x)
model.append('SVM')
```

KERNAL SVM

```
from sklearn.svm import SVC
ksvm = SVC(kernel = 'rbf', random_state = 0)
ksvm=ksvm.fit(X_train, y_train)

# Predictions:
y_pred = ksvm.predict(X_test)

# Performance:
print('Accuracy:', accuracy_score(y_test,y_pred))
print(confusion_matrix(y_test,y_pred))
print(classification_report(y_test,y_pred))

x = metrics.accuracy_score(y_test, y_pred)
acc.append(x)
model.append('kSVM')
```

DECISION TREE

```
from sklearn import tree
tree=tree.DecisionTreeClassifier()
dtree=tree.fit(X_train,y_train)
```

```
# Predictions:
y_pred = dtree.predict(X_test)

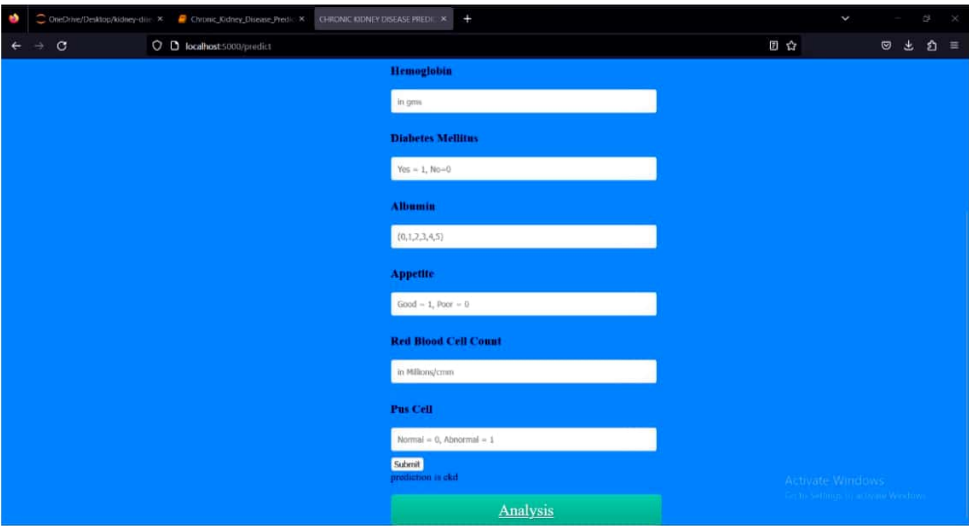
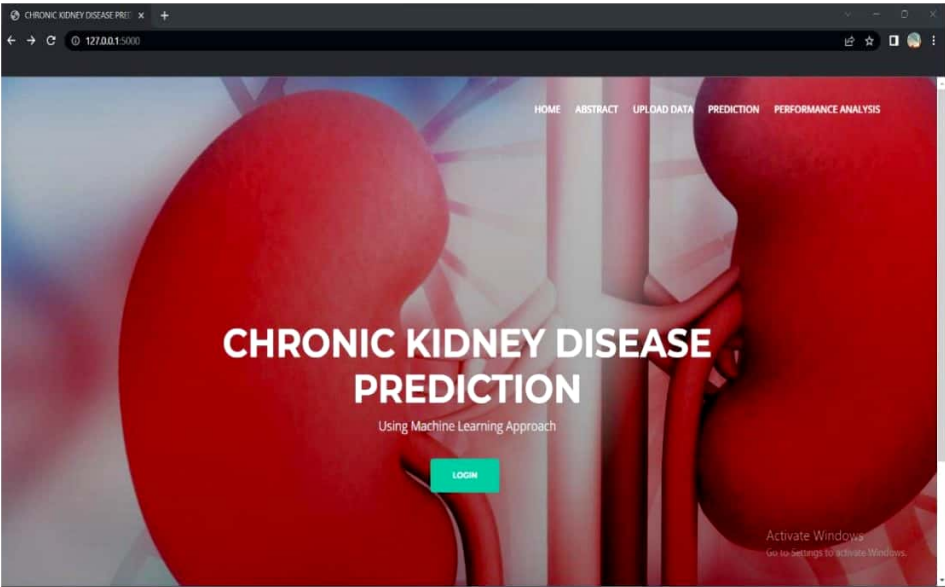
# Performance:
print('Accuracy:', accuracy_score(y_test,y_pred))
print(confusion_matrix(y_test,y_pred))
print(classification_report(y_test,y_pred))

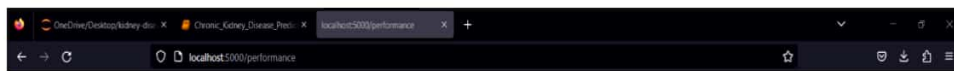
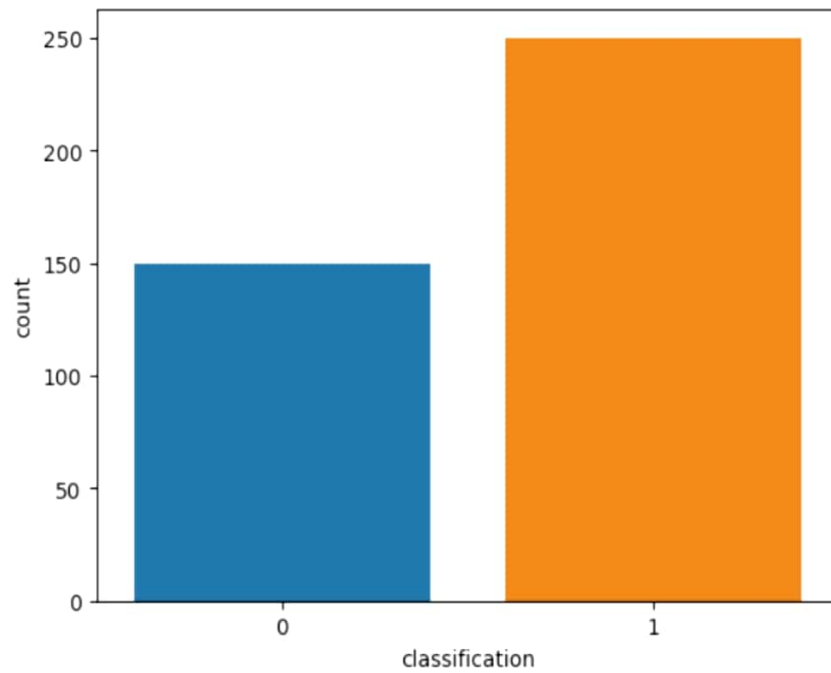
x = metrics.accuracy_score(y_test, y_pred)
acc.append(x)
model.append('DB')
```

ACCURACY COMPARISION

```
plt.figure(figsize=[10,5],dpi = 100)
plt.title('Accuracy Comparison')
plt.xlabel('Accuracy')
plt.ylabel('Algorithm')
sns.barplot(x = acc,y = model,palette='dark')
```

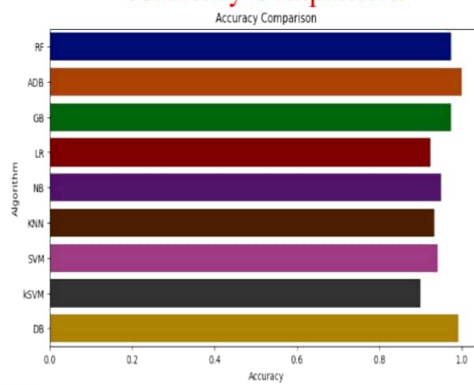
B.SCREEN SHOTS:





PERFORMANCE ANALYSIS

Accuracy Comparison



Predict

Logout

Activate Windows
Go to Settings to activate Windows.

Risk Prediction of Renal Failure Using Machine Learning

B.Lakshmi Vignesh, R.Jeberson Retna Raj, I.Hemanth Kumar

Department of Information
Technology Sathyabama Institute
of Science and Technology
Chennai, India

jebersonretnaraj.it@sathyabama.ac.in

Abstract - With the system of expertise development in the life and fitness of communities of assistant scientists, the perfect belief of clinical facts is completed through early ailment detection, affected person care and community presentations. When the beauty of medical matters is imperfect, the precision of notion is decreased. In addition, clean manifestations of nice local sicknesses are discovered in different areas, which cause the deplorable ailment. The proposed framework affords device learning algorithms to successfully are expecting the occurrence of a couple of sicknesses in societies with frequent diseases. Accumulated experiments with smooth tissue modified fashions in scientific data from medical facilities. To overcome issues within complete facts, a hidden model of the view is used to extract the lacking statistics. This is an experiment with a close to- permanent brain infarction. Using based and unstructured facts from the scientific center, the device makes use of mastering algorithms. It is in all likelihood to predict diseases based totally on records evaluation, together with Covid-19, chronic kidneysickness, and coronary heart disease. To our information, within the vast scope of statistical evaluation, no present day paintings is relevant to the look at of both. With ordinary scoring algorithms, the calculation accuracy of our proposed rule reaches 94.8%, and the convergence charge is better than that of the random probability prediction rule set when using a getting to know technique.

Keywords— *Renal failure, Risk prediction, Electronic health record, Health bigdata, Machine learning*

I. INTRODUCTION

The Earth is experiencing a technological crimson blip, and then the demands for intelligence and precision arise. Today, human beings are probable addicted to the Internet, but they do not worry approximately their bodily health. People ignore a small problem and do now not go to the health

facility, which finally leads to a severe contamination. Using this growing technology, our fundamental aim is to develop advice that predicts couple of illnesses based totally on sufferers' signs without traveling hospitals/medical doctors. Machine mastering is a subset of AI this is especially worried with the examiner of algorithms that are stepped forward by means of information and experience. Apparatus learning two steps, i.e., Training and trying out. Machine mastering presents a powerful platform inside the medical subject to remedy diverse health troubles an awful lot faster. There are two varieties of device getting to know, predictive gaining knowledge of and unsupervised studying. In cautious studies we shape a model with properly-categorized data. On the other hand, an embedded learning model learns from disjoint facts. The goal is to derive an excellent system getting to know algorithm that is powerful and correct for disorder prediction. In this newsletter, the idea of gadget getting to know control is used to predict sicknesses. The main feature will be system getting to know, in which we are able to use a gadget learning algorithm so as to be capable of appropriately expect diseases at an early degree and enhance patient care.

II. LITERATURE SURVEY

In [1], chronic kidney disease (CKD) has grown to be a serious global health issue. Between 1990 and 2017, a study by the GBD Chronic Kidney Disease Collaboration sought to determine the burden of CKD on a global, regional, and country scale. The study discovered that, with a sharp rise in fatalities and disability-adjusted life years (DALYs) between 1990 and 2017, CKD is now the biggest cause of death and disability in the world. 35.8 million DALYs and 1.2 million fatalities were attributed to CKD in 2017. The research also emphasized that CKD affects high-income countries more so than low- and middle-income countries. To lessen the effects of CKD, risk factors like obesity, diabetes, and hypertension must be addressed. Additional research from the US and China. Other research from the US and China has also shown how common CKD is and how it affects healthcare costs

and quality of life, highlighting the need for a worldwide strategy to handle this problem.

In [2], In the United States, end-stage renal disease (ESRD) is a significant public health concern with considerable effects on patients and healthcare systems. Somnath Pal reviewed the literature and looked at the main ESRD causes in the US. The study found that diabetes and hypertension, which together make up about two-thirds of cases, are the two main causes of ESRD. About 44% of instances of ESRD are caused by diabetes, which is also the most common cause. Diabetes-related high blood sugar levels can gradually harm the kidneys' filtration function, resulting in kidney failure. With 28% of cases, hypertension is the second most typical cause of ESRD. The kidneys' ability to filter waste and High blood pressure can make it difficult to retain fluids because it harms the kidney's blood vessels. Glomerulonephritis, polycystic kidney disease, and autoimmune illnesses are some additional less frequent causes of ESRD. Compared to other racial and ethnic groups, African Americans, Hispanics, and Native Americans have a higher chance of having ESRD. The study emphasizes how controlling diabetes and hypertension is crucial for delaying the onset and progression of ESRD. Early kidney disease detection and treatment can lessen the burden on patients and healthcare systems by helping to delay or avoid the requirement for kidney transplantation or dialysis.

In [3], A national registry with the mission of tracking the incidence, treatment, and prognosis of The China Kidney Illness Network (CK-NET) is a network for kidney illness in China. The CK-NET 2015 annual data report by L. Zhang et al. provides a comprehensive analysis of the network's data. The study found that chronic kidney disease (CKD) is very common in China, with an estimated 119.5 million cases in 2015. Hypertension, followed by diabetes and glomerulonephritis, was the most frequent cause of CKD. The study also showed differences in CKD incidence and prevalence by geography, with a greater burden in rural and western sections of the nation. The research emphasizes the necessity of better CKD care and prevention in China. The prevalence of CKD in the nation could be decreased with the aid of measures including early detection and treatment of hypertension and diabetes, lifestyle modifications, and public awareness campaigns. The management and results in China are those with end-stage renal disease (ESRD) also revealed by the CK-NET registry. According to the study, hemodialysis was the most popular form of treatment for those with ESRD, with a high prevalence of catheter use and insufficient dialysis dosage. The emphasizes the need of raising the availability and caliber of

dialysis services for ESRD patients in China.

In [4], The prevention and treatment of chronic kidney disease (CKD), a severe public health concern, depend on identifying those who are at risk of developing it and monitoring its development. J.B. Echouffo Tcheugui and A.P. Kengne conducted a thorough investigation to evaluate how effectively risk models predicted CKD and its progression. The study discovered 22 risk models for CKD and 12 risk models for CKD progression. The models' performance varied greatly; some had strong calibration and discriminatory power, while others did not. According to the study, the majority of risk models contained well-known risk factors such age, sex, race, diabetes, and hypertension. The scientists came to the conclusion that better risk models are required, ones that take into account novel risk variables and unusual markers including albuminuria, glomerular filtration rate, and genetic markers. They also emphasized how crucial it is to test how well risk models operate across various populations and contexts. The study emphasizes the present flaws and deficiencies in risk prediction models for CKD development. To determine people who are more likely to acquire CKD and to enable focused interventions and prevention measures, risk models should be improved. The study emphasizes the necessity of ongoing work in this field to enhance CKD care and prevention

In [5], Early detection and therapy of Stopping the progression of chronic kidney disease (CKD), a severe public health issue, is essential. H. Bang and colleagues developed the Screening for Occult Renal Disease (SCORED) model, a simple CKD prediction tool, to aid in identifying patients who are at risk of developing CKD. NHANES III, the Third National Health and Nutrition Examination Survey, offered information on more than 14,000 people, and this information was used to create the SCORED model. A different cohort of more than 5,000 people was used for validation. Age, sex, race, and a history of hypertension are four readily available characteristics that the model uses to predict the likelihood of getting CKD. The study discovered that the SCORED model performed better in identifying people at risk than other prediction models because it had strong discriminatory power and calibration. The study discovered that the SCORED model beat previous prediction models in identifying those who were at risk of developing CKD and had good discriminatory power and calibration. The SCORED model is a helpful tool for primary care doctors to identify people at risk of CKD because of its simplicity and use of widely available information, according to the authors. The creation and validation of the SCORED model show that it

is possible to identify people who are at risk of developing CKD by utilizing straightforward prediction models. By using such models in primary care settings, early interventions could be made to stop the course of CKD and assist in identifying people at risk. The study emphasizes how crucial it is to carry out further research in order to create and validate prediction models for CKD.

In [6], Kshirsagar and colleagues set out to develop an easy-to-use algorithm that could predict the incidence of renal illness in a certain neighborhood. The method was developed using data from the Screening for Occult Renal Disease (SCORED) project, and it was validated using data from the Atherosclerosis Risk in Communities (ARIC) study. Age, gender, race, blood pressure, presence of diabetes, and status of smoking, and history of cardiovascular disease are just a few of the seven readily available clinical factors that were incorporated to the algorithm. It demonstrated strong prediction accuracy with a C-statistic of 0.81 in the SCORED dataset and 0.79 in the ARIC dataset. Finding those at a high risk of developing renal disease is crucial for early intervention and risk factor control. This may contribute to lessening this strain.

In [7], C.S. Fox et al.'s 2010 study looked into the use of a multi-marker approach in determining the likelihood of developing chronic kidney disease (CKD) and microalbuminuria. The Framingham Offspring Study, which tracked a cohort of people for 8 years, provided the researchers with the data they used. The analysis comprised 2,345 participants in total. The multi-marker method utilized in the study involves evaluating a combination of more recent biomarkers, such as homocysteine and fibrinogen, as well as more established risk factors for CKD, such as blood pressure and glucose levels. In comparison to employing only traditional risk factors, the researchers discovered that this strategy was able to greatly enhance the prediction of incident CKD and microalbuminuria. The study also discovered certain indicators, such as cystatin C and urinary albumin-to-creatinine ratio, that were highly predictive of CKD and microalbuminuria. These markers might be helpful in identifying those who are highly susceptible to developing CKD or microalbuminuria, enabling earlier intervention and management.

In [8], Hippisley-Cox and Coupland set out to develop and verify prediction models for men and women in England and Wales with chronic kidney disease (CKD). They made using the QResearch database, which holds the medical records of more than 7 million patients who are registered with general practitioners. The models, which were

externally evaluated by utilizing a different cohort of patients, contained variables such as age, sex, ethnicity, smoking status, diabetes, hypertension, and past cardiovascular disease. With AUCs of 0.87 and 0.89 for men and women, respectively, and satisfactory calibration, which means that the projected hazards were congruent with the risks encountered by the validation group, the QKidney scores demonstrated good discriminatory power. In order to identify high-risk CKD patients for early therapies or screening programmes and to support collaborative decision-making, the QKidney scores could be employed. More investigation is required.

In [9], calculating the likelihood that Taiwanese people will develop chronic kidney disease (CKD), Chien et al. built a prediction model. Data from a community-based cohort study with over 12,000 people aged 35 to 64 was used by the authors. At baseline and during follow-up visits, the participants' demographic data, medical history, and laboratory test results were gathered. To build their prediction model, the authors used demographic and clinical factors such as age, sex, body mass index, blood pressure, diabetes, smoking status, and laboratory results. Several statistical techniques, such as Performance of the model was assessed using calibration plots and the area under the receiver operating characteristic curve (AUC). An AUC of 0.77 indicated from the findings that the prediction model had good discriminating. The model's calibration was also accurate, which means that the anticipated and observed hazards were closely related. The scientists came to the conclusion that people in Taiwan who are at a high risk of having CKD might be identified using their prediction model. This study emphasizes the significance of early CKD risk factor identification since it enables prompt therapies to stop or delay the onset of CKD. By identifying high-risk people for focused therapies, the prediction model could also be used to optimize healthcare resource allocation and enhance patient outcomes. Further research is necessary to see whether the approach can be applied to other populations and healthcare settings.

In [10], the goal of the work by Ando et al. was to develop a straightforward model to estimate the likelihood that HIV-infected people are at risk for developing chronic kidney disease (CKD). To create and validate their prediction model, the scientists used data from 1,067 HIV-positive people in Japan, who were split into two groups: a derivation cohort and a validation cohort. Age, sex, estimated glomerular filtration rate (eGFR), and proteinuria were the variables utilized in the model and are all easily accessible in clinical practice. Higher values on the risk score, which the authors established and ranged from 0 to 10, indicated a higher chance of CKD occurrence. The model

demonstrated good predictive accuracy with an area under the curve (AUC) of 0.83 and 0.77 in the derivation and validation cohorts, respectively. The risk score was also effective at identifying patients who had a high chance of developing CKD throughout a 5-year period. This study highlights the value of CKD early identification in HIV-infected patients because they are more likely to acquire the condition. The straightforward and usefulness of the prediction model created in this study may help clinicians identify at-risk individuals and direct interventions to stop or reduce the course of CKD in this population. To determine the generalizability of this concept, more validation across diverse populations is required.

In [11], A study was carried out by the ADVANCE Collaborative Group to assess how well a risk prediction model predicted kidney-related outcomes those who have type 2 diabetes. The ADVANCE project, which included more than 11,000 participants with type 2 diabetes from 20 different countries, provided data on which the model was built. The model considered systolic blood pressure, age, sex, the duration of diabetes, smoking status, serum creatinine, and urinary albumin-creatinine ratio among other risk factors. The model's propensity to forecast a composite renal disease outcome was evaluated. The study found that the prediction model had good calibration and discrimination performance. It showed strong discriminative performance in both the derivation and validation cohorts, with c-indices of 0.78 and 0.75, respectively. Additionally, the model's calibration was strong, showing that the projected and actual hazards were well matched. The model can identify those who are more likely to experience kidney-related outcomes, allowing medical professionals to start preventive measures early and lower the risk of kidney-related problems in type 2 diabetes patients. To confirm the generalizability of the model, more validation in additional populations is required.

In [12], O'Seaghdha et al. investigated the potential of a genetic risk score (GRS) to predict the prevalence of stage 3 chronic kidney disease (CKD) in the general population. Using information from the Framingham Offspring Study, they developed a GRS based on genetic variations associated with CKD and evaluated its effectiveness in predicting CKD stage 3 over a median follow-up of 8.2 years. The GRS alone had a low AUC of 0.59 and little predictive power. The predictive value significantly increased when combined with conventional risk factors, such as age, sex, diabetes, and had an AUC of 0.79. hypertension, and smoking status. The work emphasizes the significance of include genetic data in CKD risk prediction models in addition to traditional risk variables. In this study, O'Seaghdha

et al. investigated the potential of a genetic risk score (GRS) to predict the prevalence of stage 3 chronic kidney disease (CKD) in the general population. Using information from the Framingham Offspring Study, they developed a GRS based on genetic variations associated with CKD and evaluated its effectiveness in predicting CKD stage 3 over a median follow-up of 8.2 years. The GRS alone had a low AUC of 0.59 and little predictive power. When combined with conventional risk factors such as age, sex, diabetes, hypertension, and smoking status, the predictive value, with an AUC of 0.79, significantly improved. The study highlights the need of include genetic information in CKD risk prediction models together with conventional risk factors. Further study develop individualized CKD risk prediction algorithms.

In [13], the goal of Tangri et al.'s The goal of the study was to create a model for how chronic kidney disease (CKD) progresses to renal failure. Over 1.1 million persons with CKD from different populations were involved in the investigation, which incorporated information from two sizable multicenter studies. 17 factors, including age, sex, race, estimated glomerular filtration rate (eGFR), serum albumin level, and proteinuria, were included in the prediction model. The main result was the onset of renal failure, which is characterized as the need for kidney transplantation or long-term dialysis. The model was successful in forecasting the probability of kidney failure across two, five, and ten years, the researchers discovered. For the model, the C-statistics were 0.90, 0.85, and 0.82, respectively. The model was properly calibrated because the expected and actual hazards were comparable. The model may identify those who are at a high risk of renal failure, which has important implications for the care of CKD patients, according to the study's findings. It is possible to stop or delay the progression of CKD. by using this knowledge to create more focused therapies. To validate the model across various population and healthcare settings, additional research is necessary.

In [14], In order to assess the effectiveness of several equations that predict the chance of kidney failure in people with chronic kidney disease (CKD), N. Tangri et al. performed a global meta-analysis in this study. Over 2 million people from 30 different nations made up the 46 cohorts from which the researchers collected their data. The main result of interest was how well the equations predicted the likelihood of kidney failure, which was defined such involves the requirement for kidney replacement therapy or passing away due to renal failure. By assessing the discrimination and calibration capabilities of 16 different equations, including both clinical and laboratory-based

equations, the researchers compared their performance. The outcomes demonstrated that no single equation was consistently accurate and that the performance of the equations varied greatly across various populations. The equation with the highest discrimination ability across all cohorts and a median the creatinine-cystatin C equation in the CKD-EPI had an area under the receiver operating characteristic curve (AUC) of 0.90. The study's findings have significant implications for the clinical care of people with CKD, since precise risk prediction models can assist medical professionals in identifying high-risk patients early and starting preventive medications. To ensure accuracy and generalizability, the results further highlight the significance of evaluating risk prediction models across a range of groups

In [15], In this study, Winnicki et al. evaluated the Kidney Failure Risk Equation's (KFRE) usefulness in predicting the likelihood that children with chronic kidney disease (CKD) would develop to end-stage renal disease (ESRD). The study's primary outcome was ESRD or death, and it involved 2,394 kids with CKD stages 3-5 from the Chronic Kidney Disease in Children study. To determine the 2- and 5-year risk of ESRD, the KFRE integrates age, sex, estimated glomerular filtration rate (eGFR), and urine albumin-to-creatinine ratio (ACR). The KFRE was proven to be a valuable tool for predicting the risk of ESRD in kids with CKD stages 3-5, according to the study. The KFRE's C-statistic for predicting ESRD or mortality within two years was 0.88 (95% CI, 0.85-0.91). The KFRE outperformed the conventional CKiD (Chronic Kidney Disease in Children) equation for predicting ESRD in kids with CKD stages 3-5, according to the study. When predicting ESRD, the KFRE fared better than the CKiD equation in terms of discrimination and calibration. According to the study, the KFRE can be a helpful tool for clinical decision-making because it accurately predicts the chance of ESRD and counselling for children with CKD stages 3-5 and their families.

In [16], a popular method for estimating the likelihood of end-stage renal disease (ESRD) in people with chronic kidney disease (CKD) is the Kidney Failure Risk Equation (KFRE). The KFRE was intended to be validated in a European sample of CKD patients by Peeters et al. In total, 2,429 patients from five different European nations participated in the study, and their KFRE scores were determined using the clinical and laboratory information from those patients. With The results of the study showed that the KFRE was useful in predicting the risk of ESRD in European CKD patients, with an area under the receiver operating characteristic curve (AUC) of 0.86 (95% CI, 0.84-

0.88) for predicting ESRD within 2 years and an AUC of 0.83 (95% CI, 0.81-0.85) for predicting ESRD within 1 year. The Modification of Diet in Renal Disease (MDRD) equation and the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation were determined to be subpar compared to the KFRE. The results of the study show that the KFRE is a valid and accurate approach for assessing the risk of ESRD in patients with CKD in Europe. Using the KFRE, doctors can identify individuals who are most at risk for developing ESRD and to direct treatment choices and measures aimed at delaying or preventing its onset. Additional research is required to examine the KFRE's possible clinical applications and validate it in additional demographics.

In [17], The advancement of chronic kidney disease (CKD), a major consequence of type 2 diabetes mellitus (T2DM), in these patients is critical for early intervention and effective therapy. Based on a 13-year follow-up research in Singapore, Low et al. set out to construct and verify a prediction model for CKD progression in T2DM patients. 2,345 T2DM patients with normal kidney function or early CKD were enrolled in the study. The main result was the development of CKD, which was determined by the estimated glomerular filtration rate (eGFR) declining by less than 5 mL/min/1.73 m² year or requiring renal replacement therapy. The authors used Cox regression analysis to create a prediction model, which they then tested with bootstrapping methods. Age, gender, eGFR, albuminuria, body mass index, systolic blood pressure, haemoglobin A1c, serum cholesterol, smoking status, and duration of diabetes were identified as significant predictors of CKD progression in the resultant prediction model. In both the derivation and validation cohorts, the model demonstrated strong calibration and discrimination. The suggested forecasting model, according to the authors' findings, can effectively forecast how CKD in T2DM patients will advance as well as help identify and manage high-risk patients early on. But before using this model in clinical practice, more validation research in a range of populations is needed.

In [18], the goal of Grammes et al.'s study was to assess how well the kidney failure risk equations performed in predicting the time to end-stage renal disease (ESRD) in African American patients with CKD and hypertension. 691 people with CKD who were not on dialysis at the start of the trial were drawn from the African American trial of Kidney Disease and Hypertension (AASK). The time to ESRD among African American patients with CKD and hypertension was successfully predicted by risk equations, according to the authors, with a median overall difference between predicted and observed

time to ESRD of 0.0 years. The study emphasizes how critical precise ESRD time prediction is for maximizing patient care and resource use. The risk equations can be used to identify patients who are at an increased risk of developing ESRD, enabling prompt interventions like dialysis or referral for a kidney transplant. The risk equations might not be as generalizable to other populations, according to the authors, who also note the need for more validation research.

In [19], Yamanouchi et al.'s (2018) study sought to determine the value of incorporating the renal pathology score into the Kidney Failure Risk Equation (KFRE) in order to forecast the advanced diabetic nephropathy development. The probability of end-stage renal disease (ESRD) in individuals with diabetic nephropathy was found to be highly linked with the renal pathology score, according to the authors. In addition, they found that combining the renal pathology score with the KFRE enhanced its capacity to predict ESRD in comparison to using the KFRE alone. The study includes a retrospective review of renal biopsy results from 78 patients with advanced diabetic nephropathy. The authors generated the renal pathology score and evaluated the severity of glomerular and interstitial abnormalities. The effectiveness of the KFRE in predicting ESRD was then tested with and without the renal pathology score. The results of this study imply that include renal pathology data in the KFRE may enhance the ability of patients with advanced diabetic nephropathy to predict the development of ESRD. Larger investigations are required to validate these results, the authors added.

In [20], based on the modified Kidney Failure Risk Equation (KFRE), Lim et al. created a simplified risk prediction model for low-risk general populations with chronic kidney disease (CKD). As ESRD indicators, the model takes into account age, sex, estimated glomerular filtration rate (eGFR), and urine albumin-to-creatinine ratio (ACR). To validate the KFRE, the study used data from two independent cohorts and a population-based cohort in Singapore. In low-risk CKD populations, the simplified KFRE outperformed the Modification of Diet in Renal Disease (MDRD) and the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equations in terms of predicting ESRD. The model's classification of CKD patients into various risk groups can help with appropriate care and clinical decision-making.

In [21], Tummalapalli and Estrella addresses how risk prediction models may affect the treatment of chronic kidney disease (CKD). The authors outline the drawbacks of conventional CKD treatment, which frequently emphasizes illness stage rather than risk assessment. They contend that risk

prediction algorithms have the capacity to recognize high-risk people early in the course of the disease, enabling more proactive and specialized care. The Kidney Failure Risk Equation (KFRE) and the CKD Prognosis Consortium models, among others, are reviewed by the authors together with other current risk prediction models. They point out that these models have shown high discrimination and calibration across a range of populations, and that adding further variables like biomarkers and genetic data may enhance their performance still further. The necessity for precise and trustworthy data as well as the potential for over diagnosis and overtreatment are some other difficulties that Tummalapalli and Estrella cover in their discussion of the implementation of risk prediction models in clinical practice. They assert that risk-based kidney care must be implemented in a coordinated manner by patients, healthcare professionals, and healthcare systems.

In [22], Understanding how ageing affects the kidneys is becoming more and more crucial as the world's population ages. Hommos et al. set out to investigate the structural and functional alterations in the ageing kidneys of healthy people. The study used post-mortem kidney samples from 159 healthy people between the ages of 20 and 100. The authors discovered that the cortical thickness and renal volume both declined with age. On the other hand, as people age, the volume and number of glomerular tufts declined while the proportion of scleroses glomeruli climbed. The scientists also noted an aging-related increase in tubular atrophy, interstitial fibrosis, and arteriolar halitosis. The research proved that glomerular sclerosis and the loss of functioning nephrons are inevitable aspects of human ageing.

In [23], the relationship between obesity and deterioration in glomerular filtration rate (GFR) utilising a meta-analysis of participant-specific data from 36 cohorts. The analysis was stratified by CKD status, and there were almost five million individuals in the research from different ethnic origins. Regardless of CKD status, the study discovered that higher adiposity, as measured by body mass index (BMI), waist circumference, and waist-to-hip ratio, was linked to a higher risk of GFR deterioration. Additionally, the connection was more pronounced in people with CKD. The study's findings also imply that the impact of obesity on GFR deterioration maybe mediated by additional elements like diabetes and hypertension. The work sheds crucial light on the connection between adiposity and the advancement of CKD and emphasizes the significance of weight control as a possible method to lower the risk of CKD progression. The study was however constrained by the use of solitary assessments of adiposity, and the

possible influence of weight fluctuation over time on the development of CKD was not investigated.

In [24], An important public health issue is the rising prevalence of obesity and its links to numerous chronic diseases. Obesity is another a well-known risk factor for the onset and progression of renal disease. A thorough analysis of the epidemiological, clinical, and basic scientific data demonstrating the relationship between obesity and kidney illness is provided by Whaley-Connell and Sowers. The authors look at a number of pathways, such as insulin resistance, oxidative stress, inflammation, and activation of the renin-angiotensin-aldosterone system, that connect obesity to kidney disease. Adipocytes, bioactive chemicals made by adipose tissue, are also highlighted for their part in mediating the link between obesity and kidney disease. The authors address the current treatment options for obesity-related kidney disease, which mostly centre on weight loss, blood pressure management.

In [25], A major public health concern, chronic kidney disease (CKD) is more common among women. Silbiger and Neugarten looked into how gender affected the development of CKD. They proposed that the gender-based variations in CKD progression might be caused by sex hormones. The study analyzed the clinical, biochemical, and histological information of CKD patients from several studies, including those on IgA nephropathy, diabetic, and hypertensive nephropathies. The study found that gender differences existed in the course of CKD in different populations, with men seeing a faster drop in renal function than women. Additionally, females showed better histology results than males. The results of the study were contradictory among patients with glomerulonephritis, though. The authors offered a number of hypotheses for the gender that was observed. Differences, such as those in the function of sex hormones, socioeconomic variables, and renal hemodynamics.

III. ARCHITECTURE

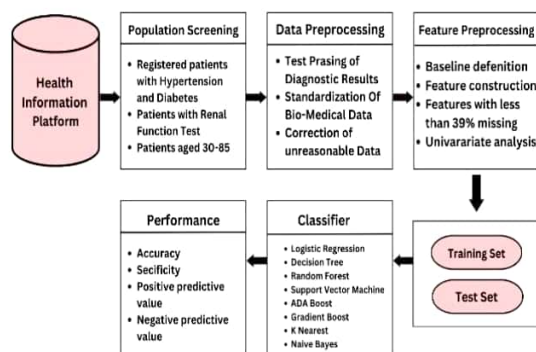


FIG.1 System Architecture

IV. PROPOSED SYSTEM

V.

The proposed system will collect all valid details of patient like age, specific gravity, potassium, hemoglobin, wbc and rbc count etc and other required medical information through a clinical data file which in format of .CSV since data in tabular form is easier to predict in through machine learning algorithm like Support vector machine(SVM),K Nearest Neighbor (KNN), Decision Tree and Random Forest In this system ,since we use KNN so even if there is empty values in columns that doesn't stop the system from running and the system got higher accuracy compared to other.

VI. PROPOSED ALGORITHMS

1. Logistic Regression
2. Decision Tree
3. Random Forest
4. Support Vector Machine
5. ADA Boost
6. Gradient Boost
7. K Nearest
8. Naïve Bayes

Logistic Regression

A supervised learning technique used in machine learning for classification applications is logistic regression. It is frequently used for binary classification situations when there are only two possible outcomes for the output variable (such as whether a customer will leave or not). The logistic function, which converts any input value to a number between 0 and 1, is used in the method to describe the relationship between the input variables and the likelihood of the output variable. In order to minimise a cost function, the logistic regression algorithm uses an optimisation method like gradient descent. The approach modifies the input variable weights during the training phase in order to reduce the cost function and boost prediction accuracy.

Decision Tree

In machine learning, a decision tree is a supervised learning technique used for both classification and regression tasks. According to the value of the input characteristics, the input data is recursively divided into smaller subsets, and this process is repeated until the subsets are pure or nearly pure in terms of the output variable. Decision trees can handle category and numerical data and are easy to understand. They are frequently employed in a variety of applications, including fraud detection, customer segmentation, and medical diagnostics. The overfitting and excessive variance that decision trees may experience can be reduced utilising approaches like pruning and ensemble methods.

Random Forest

A well-liked ensemble learning technique for classification and regression tasks in machine learning is called random forest. Multiple decision trees are built using various randomly selected subsets of the input data, and their forecasts are then combined to get a final prediction. A strong method that can handle categorical and numerical data and resists overfitting is random forest. It is frequently utilised in a variety of applications, including picture classification, fraud detection, and customer segmentation. Random forest can be computationally expensive and may not be appropriate for real-time applications, which is its fundamental drawback.

Support Vector Machine

A common supervised learning technique used in machine learning for both classification and regression applications is the Support Vector Machine (SVM). SVM operates by identifying the hyperplane that best divides the data into distinct groups or accurately predicts the target variable. SVM is a potent algorithm with a solid theoretical base that can handle highly dimensional and complex data. It is frequently utilised in a variety of applications, including bioinformatics, text classification, and picture classification. SVM may, however, be computationally expensive for large datasets and sensitive to the selection of hyperparameters.

ADA Boost

A well-liked ensemble learning technique for classification and regression tasks in machine learning is called AdaBoost (Adaptive Boosting). To attain high accuracy on the input data, weak learners (classifiers or regressors) are combined to create a strong learner. Powerful and resistant to overfitting, AdaBoost is an algorithm that can handle category and numerical data. It may also be used with any weak learner, such as decision trees or SVMs, and even with a small number of weak learners, it can achieve excellent accuracy. AdaBoost may be computationally expensive for large datasets and might be susceptible to noisy data and outliers.

Gradient Boost

Another well-liked ensemble learning approach for classification and regression applications in machine learning is gradient boosting. It is an iterative method that creates a series of ineffective learners (usually decision trees) and combines them into an effective learner that can accurately predict the target variable. A potent approach that can handle categorical and numerical data and is resistant to overfitting is gradient boosting. It is capable of handling non-linear interactions between the input and output variables and can achieve high accuracy even with a limited number of weak learners.

Gradient boosting is frequently utilised in numerous fields, including image classification, anomaly detection, and web search ranking.

KNN

A well-liked supervised learning technique used in machine learning for both classification and regression applications is the k-Nearest Neighbours (k-NN) algorithm. In order to forecast the class label or target variable of a new data point, it first locates the k nearest neighbours of a given data point in the feature space and uses their average values or class labels to do so. The k-NN technique is a straightforward and understandable method for handling categorical and numerical data without making any assumptions about the distribution of the data itself. It is extensively utilised in a variety of applications, including anomaly detection, recommendation systems, and picture recognition.

Naïve Bayes

Supervised learning algorithm for classification tasks in machine learning is naive bayes. Based on the Bayes theorem, it makes the assumption that given the class label, the features are conditionally independent. The Nave Bayes algorithm is a straightforward and effective computational tool for handling category and numerical data. It can be quickly trained on small to medium-sized datasets and updated with fresh information. Naive Bayes is widely utilised in numerous fields, including sentiment analysis, text classification, and email spam filtering.

VII. EXPERIMENTAL RESULTS

The overall system design objective is to provide a better performing and efficient system. That has higher accuracy than the existing system and has better performance parameters. It also provides faster run time for predicting the result. The training and testing modules which we have used helps us in getting higher accuracy. dimensionality to improve the presentation. By Taking the data set into the consideration we developed a risk prediction model of renal failure. With this model the patients suffering from chronic kidney disease can know whether they have ckd or not. This Model can identify the chronic kidney disease with the highest accuracy of 97%



Fig.2 Home Page

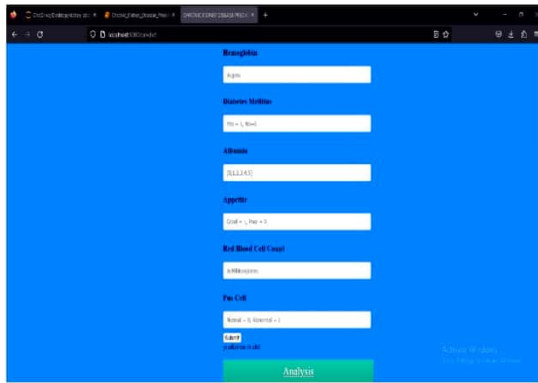


Fig.3 Process of Predicting CKD

VIII. PERFORMANCE TABLE

TABLE I. PERFORMANCE OF THE PROPOSED METHOD

ALGORITHM	ACCURACY	RECALL	PRECISION	F1-SCORE
RANDOM FOREST	0.97	1.00	0.95	0.98
ADA BOOST	0.95	0.94	1.00	0.95
GRADIENT BOOSTING	0.96	0.92	0.96	0.93
LOGISTIC REGRESSION	0.92	0.92	0.93	0.93
NAIVE BAYES	0.95	0.90	1.00	0.95
KNN	0.93	0.92	0.95	0.93
SVM	0.94	0.94	0.95	0.94
KERNAL SVM	0.90	0.92	0.91	0.90
DECISION TREE	0.99	0.98	1.00	0.99

The performance of our technique and other indicators for detecting renal failure are compared in this graph. 9 techniques and 4 metrics have been taken into account when executing our plan for renal failure. First, we used Random Forest, whose accuracy displayed a range of percentages for several metrics. The accuracy figures were 97%, 0.95 percent for precision, 0.98 percent for F1 Score, and 1% for recall. The following metrics are Ada boost, which have accuracy levels of 0.95 percent, precision percentages of 1 percent, F1 score percentages of 0.95 percent, and recall percentages of 0.94 percent. The Accuracy number is 0.96%, the Precision percentage I 0.92% is the percentage value. Then follows the Logistic Regression techniques, with percentage values for accuracy, precision, and F1 Score of 0.92%, 0.93%, and 0.92 respectively. And recall value is 0.92% The Naive Bayes method is the next step, and its metrics percentages for Accuracy are 0.95%, Precision is 1%, F1 Score value is 0.95%, and Recall is 0.90%.then comes the KNN(K nearest neighbour) methods and the metrics percentage values for Accuracy is 0.93%, precision percentage is 0.95%, the percentage value for F1 Score is 0.93%.and recall value is 0.92% and our proposed method is Decision Tree and also have the highest percentage

amongst the considered methods and the Accuracy value is 0.99%, precision percentage is 1%, F1 Recall percentage value is 0.98% and score 0.99%.

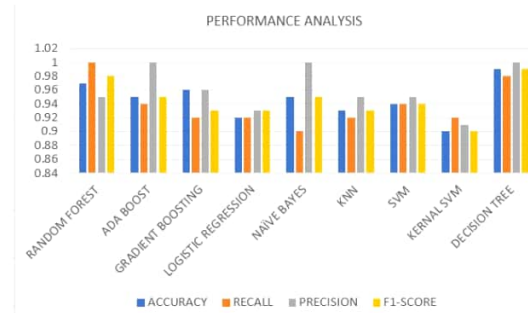


Fig.4. Performance Comparision Graph

This graph shows the comparison of performances of our method and the other metrics used to detect the renal failure. We have taken 9 methods and 4 metrics into consideration for our execution in renal failure. First method We have taken Random Forest where it's Accuracy shows different percentages in different metrics here the values for accuracy is 97%, percentage for precision is 0.95%, percentage for F1 Score is 0.98% and the percentage for Recall is 1%. Then comes the next metrics Ada boost where the accuracy level is 0.95%, Precision percentage is 1%, F1 Score percentage is 0.95% and the percentage of Recall is 0.94%. Next is the Gradient boost the Accuracy value is 0.96%, precision percentage is 0.96%, F1 Score 0.93% and Recall percentage value is 0.92%. then comes the Logistic Regression methods and the metrics percentage values for Accuracy is 0.92%, precision percentage is 0.93%, the percentage value for F1 Score is 0.93%.and recall value is 0.92% Next comes the Naive bayes method and the metrics percentage for Accuracy is 0.95%, precision percentage is 1%, F1 Score value percentage is 0.95% and the Recall percentage value is 0.90%.then comes the KNN(K nearest neighbor) methods and the metrics percentage values for Accuracy is 0.93%, precision percentage is 0.95%, the percentage value for F1 Score is 0.93%.and recall value is 0.92% and our proposed method is Decision Tree and also have the highest percentage amongst the considered methods and the Accuracy value is 0.99%, precision percentage is 1%, F1 Score 0.99% and Recall percentage value is 0.98%.

IX. CONCLUSION

For the identification of malaria, automated malaria detection, and quantification of malaria infection, we have provided a method and created an algorithm. Additionally, we created a machine

learning training method that can be applied to the detection of other parasite types in addition to malaria, and we also discussed how to improve the prediction value of the results. We conclude that the newly developed system of image processing technologies is suitable for parasite detection after taking into account the aforementioned methodology. The best methods for identifying cells and separating them from the rest of the image that is not of interest are those that use morphological segmentation techniques. Machine learning techniques were used to characterise the different types of plasmodium parasites, and cubic SVM emerged as the most effective method due to its high level of accuracy.

X. REFERENCES

- [1] GBD Chronic Kidney Disease Collaboration, Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017, *Lancet* 10225(395) (2020 February 29) 709–733, doi.org/10.1016/S0140-6736(20)30045-3, Epub 2020 Feb 13.
- [2] Somnath Pal, Primary causes of end-stage renal disease, *US Pharm.* 41(8) (2016) 6.
- [3] L. Zhang, M.H. Zhao, L. Zuo, Y. Wang, F. Yu, H. Zhang, H. Wang, CK-NET Work Group, China Kidney Disease network (CK-NET) 2015 annual data report, *Kidney Int., Suppl.* 9(1) (2019 Mar) e1–e81, doi.org/10.1016/j.kisu.2018.11.001.
- [4] J.B. Echouffo-Tcheugui, A.P. Kengne, Risk models to predict chronic kidney disease and its progression: a systematic review, *PLoS Med.* 9(11) (2012) e1001344, doi.org/10.1371/journal.pmed.1001344.
- [5] H. Bang, S. Vupputuri, D.A. Shoham, P.J. Klemmer, R.J. Falk, M. Mazumdar, D. Gipson, R.E. Colindres, A.V. Kshirsagar, SCReening for Occult RENal Disease (SCORED): a simple prediction model for chronic kidney disease, *Arch. Intern. Med.* 167(4) (2007 Feb 26) 374–381, doi.org/10.1001/archinte.167.4.374.
- [6] A.V. Kshirsagar, H. Bang, A.S. Bombback, S. Vupputuri, D.A. Shoham, L.M. Kern, P.J. Klemmer, M. Mazumdar, P.A. August, A simple algorithm to predict incident kidney disease, *Arch. Intern. Med.* 168(22) (2008 Dec 8) 2466–2473, doi.org/10.1001/archinte.168.22.2466.
- [7] C.S. Fox, P. Gona, M.G. Larson, J. Selhub, G. Tofler, S.J. Hwang, J.B. Meigs, D. Levy, T.J. Wang, P.F. Jacques, E.J. Benjamin, R.S. Vasan, A multi-marker approach to predict incident CKD and microalbuminuria, *J. Am. Soc. Nephrol.* 21(12) (2010 Dec) 2143–2149, doi.org/10.1681/ASN.2010010085.
- [8] J. Hippisley-Cox, C. Coupland, Predicting the risk of chronic kidney disease in men and women in England and Wales: prospective derivation and external validation of the QKidney scores, *BMC Fam. Pract.* 21(11) (2010 Jun 21) 49, doi.org/10.1186/1471-2296-11-49.
- [9] K.L. Chien, H.J. Lin, B.C. Lee, H.C. Hsu, Y.T. Lee, M.F. Chen, A prediction model for the risk of incident chronic kidney disease, *Am. J. Med.* 123(9) (2010 Sep) 836–846.e2, doi.org/10.1016/j.amjmed.2010.05.010.
- [10] M. Ando, N. Yanagisawa, A. Ajisawa, K. Tsuchiya, K. Nitta, A simple model for predicting incidence of chronic kidney disease in HIV-infected patients, *Clin. Exp. Nephrol.* 15(2) (2011 Apr) 242–247, doi.org/10.1007/s10157-010-0393-x.
- [11] M.J. Jardine, J. Hata, M. Woodward, V. Perkovic, T. Ninomiya, H. Arima, S. Zoungas, A. Cass, A. Patel, M. Marre, G. Mancia, C.E. Mogensen, N. Poulter, J. Chalmers, ADVANCE Collaborative Group, Prediction of kidney-related outcomes in patients with type 2 diabetes, *Am. J. Kidney Dis.* 60(5) (2012 Nov) 770–778, doi.org/10.1053/j.ajkd.2012.04.025.
- [12] C.M. O'Seaghdha, Q. Yang, H. Wu, S.J. Hwang, C.S. Fox, Performance of a genetic risk score for CKD stage 3 in the general population, *Am. J. Kidney Dis.* 59(1) (2012 Jan) 19–24, doi.org/10.1053/j.ajkd.2011.08.030.
- [13] N. Tangri, L.A. Stevens, J. Griffith, H. Tighiouart, O. Djurdjev, D. Naimark, A. Levin, A.S. Levey, A predictive model for progression of chronic kidney disease to kidney failure, *JAMA* 305(15) (2011 Apr 20) 1553–1559, doi.org/10.1001/jama.2011.451.
- [14] N. Tangri, M.E. Grams, A.S. Levey, J. Coresh, L.J. Appel, B.C. Astor, G. Chodick, A.J. Collins, O. Djurdjev, C.R. Elley, M. Evans, A.X. Garg, S.I. Hallan, L.A. Inker, S. Ito, S.H. Jee, C.P. Kovesdy, F. Kronenberg, H.J. Heerspink, A. Marks, G.N. Nadkarni, S.D. Navaneethan, R.G. Nelson, S. Titze, M.J. Sarnak, B. Stengel, M. Woodward, K. Iseki, CKD Prognosis Consortium, Multinational assessment of accuracy of equations for predicting risk of kidney failure: a meta-analysis, *JAMA* 315(2) (2016 Jan 12) 164–174, doi.org/10.1001/jama.2015.18202.
- [15] E. Winnicki, C.E. McCulloch, M.M. Mitsnefes, S.L. Furth, B.A. Warady, E. Ku, Use of the kidney failure risk equation to determine the risk of progression to end-stage renal disease in children with chronic kidney disease, *JAMA Pediatr.* 172(2) (2018 Feb 1) 174–180, doi.org/10.1001/jamapediatrics.2017.4083.
- [16] M.J. Peeters, A.D. van Zuilen, J.A. van den Brand, M.L. Bots, P.J. Blankestijn, J.F. Wetzels, MASTERPLAN Study Group, Validation of the kidney failure risk equation in European CKD patients, *Nephrol. Dial. Transplant.* 28(7) (2013 Jul) 1773–1779, doi.org/10.1093/ndt/gft063.
- [17] S. Low, S.C. Lim, X. Zhang, S. Zhou, L.Y. Yeoh, Y.L. Liu, S. Tavintharan, C.F. Sum, Development and validation of a predictive model for Chronic Kidney Disease progression in type 2 diabetes mellitus based on a 13-year study in Singapore, *Diabetes Res. Clin. Pract.* 123 (2017 Jan) 49–54, doi.org/10.1016/j.diabres.2016.11.008.
- [18] M.E. Grams, L. Li, T.H. Greene, A. Tin, Y. Sang, W.H. Kao, M.S. Lipkowitz, J.T. Wright, A.R. Chang, B.C. Astor, L.J. Appel, Estimating time to ESRD using kidney failure risk equations: results from the African American Study of Kidney Disease and Hypertension (AASK), *Am. J. Kidney Dis.* 65(3) (2015 Mar) 394–402, doi.org/10.1053/j.ajkd.2014.07.026.
- [19] M. Yamanouchi, J. Hoshino, Y. Ubara, K. Takaichi, K. Kinowaki, T. Fujii, K. Ohashi, K. Mise, T. Toyama, A. Hara, K. Kitagawa, M. Shimizu, K. Furuichi, T. Wada, Value of adding the renal pathological score to the kidney failure risk equation in advanced diabetic nephropathy,

- PLoS ONE 13(1) (2018 Jan 16) e0190930, doi .org /10 .1371 /journal .pone .0190930.
- [20] C.C. Lim, M.L. Chee, C.Y. Cheng, J.L. Kwek, M. Foo, T.Y. Wong, C. Sabanayagam, Simplified end stage renal failure risk prediction model for the low-risk gen-eral population with chronic kidney disease, PLoS ONE 14(2) (2019 Feb 22) e0212590, doi .org /10 .1371 /journal .pone .0212590.
 - [21] S.L. Tummalapalli, M.M. Estrella, Predicting risk of kidney disease: is risk-based kidney care on the horizon?, JAMA 322(21) (2019 Nov 8) 2079–2081, doi.org /10 .1001 /jama .2019 .17378.
 - [22] M.S. Hommos, R.J. Glassock, A.D. Rule, Structural and functional changes in human kidneys with healthy aging, J. Am. Soc. Nephrol. 28(10) (2017 Oct) 2838–2844, doi .org /10 .1681 /ASN.2017040421.
 - [23] A.R. Chang, M.E. Grams, S.H. Ballew, H. Bilo, A. Correa, M. Evans, O.M. Gutierrez, F. Hosseinpah, K. Iseki, T. Kenealy, B. Klein, F. Kronenberg, B.J. Lee, Y. Li, K. Miura, S.D. Navaneethan, P.J. Roderick, J.M. Valdivielso, F.L.J. Visseren, L. Zhang, R.T. Gansevoort, S.I. Hallan, A.S. Levey, K. Matsushita, V. Shalev, M. Woodward, CKD Prognosis Consortium (CKD-PC), Adiposity and risk of decline in glomeru-lar filtration rate: meta-analysis of individual participant data in a global con-sortium, BMJ 364 (2019 Jan 10) k5301, doi .org /10 .1136 /bmj .k5301.
 - [24] A. Whaley-Connell, J.R. Sowers, Obesity and kidney disease: from population to basic science and the search for new therapeutic targets, Kidney Int. 92(2) (2017 Aug) 313–323, doi .org /10 .1016 /j .kint .2016 .12 .034.
 - [25] S.R. Silbiger, J. Neugarten, The impact of gender on the progression of chronic renal disease, Am. J. Kidney Dis. 25(4) (1995 Apr) 515–533, doi.org /10 .1016 /0272 -6386(95)90119 -1.