EXPLORING PERSPECTIVES IN CHRONIC KIDNEY DISEASES USING MACHINE LEARNING

A PROJECT REPORT

Submitted by

MURUGAN M	**[20L231]**
THARUN K	**[20L257]**
VENKATESH K	**[20L259]**

in partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

IN

ELECTRONICS AND COMMUNICATION ENGINEERING

KARPAGAM COLLEGE OF ENGINEERING

COIMBATORE – 641 032

ANNA UNIVERSITY: CHENNAI 600 025

MAY 2024

KARPAGAM COLLEGE OF ENGINEERING

COIMBATORE – 641 032

ANNA UNIVERSITY: CHENNAI 600 025

BONAFIDE CERTIFICATE

Certified that this project report **"EXPLORING PERSPECTIVES IN CHRONIC KIDNEY DISEASES USING MACHINE LEARNING" is** the bonafide work of **MURUGAN M [20L231], THARUN K [20L257], VENKATESH K [20L259],** who carried out the project work under my supervision.

SIGNATURE
Dr.R.Sarankumar., M.E., Ph.D.
HEAD OF THE DEPARTMENT
Electronics and Communication Engineering
Karpagam College of Engineering
Coimbatore – 641 032

SIGNATURE
Dr.R.Sarankumar M.E.,Ph.D.
SUPERVISOR
Professor
Electronics and Communication Engineering
Karpagam College of Engineering
Coimbatore – 641 032

Certified that the candidate was examined during the viva voce examinations held on _______________

Signature of the Internal Examiner with date

Signature of External Examiner with date

DECLARATION

We here by declare that this Project report entitled **"EXPLORING PERSPECTIVES IN CHRONIC KIDNEY DISEASES USING MACHINE LEARNING"** submitted by us for the degree of **BE in Electronics and Communication Engineering at Karpagam College of Engineering, Coimbatore** is the record of original work done by us under the guidance and supervision of **Dr.R.Sarankumar M.E.,Ph.D. Professor** at the Department of Electronics and Communication Engineering, Karpagam College of Engineering, Coimbatore – 641032 and has not formed the basis for the award of any degree, or diploma or titles in this institution or any other Institution of higher learning.

Date:

Place: Coimbatore

Name and Signature of the Candidates

MURUGAN M	[20L231]
THARUN K	[20L257]
VENKATESH K	[20L259]

ACKNOWLEDGEMENT

First and foremost, praises and thanks to the almighty for her showers and blessings throughout our project work to complete it successfully.

We extend our gratitude to the Management of Karpagam College of Engineering, Coimbatore for the excellent infrastructure and support facilities to undergo the project work.

We a are very grateful to **Dr. V. KUMAR CHINNAIYAN**, the Principal and **Dr.R. SARANKUMAR** the HoD, **Electronics and Communication Engineering** for provided the facilities, support and permission to carried out our project work at our esteemed institution.

We record my sincere gratitude to our Project Coordinator **Dr.S.SOBANA** for giving inputs, encouragement for the continuous improvement during the progress and to complete this project work.

We would like to express our sincere gratitude to our Supervisor **Dr.R.SARANKUMAR** for the continuous support for our UG study, for him motivation and adequate guidance which helped us to achieve success in all our accomplishments and to complete this project work.

We also thank all the teaching faculty members and non-teaching Staff members of the department of **Electronics and Communication Engineering**, Karpagam College of Engineering, Coimbatore for their kindness and support.

We would like to thank **our parents, family members and friends** who sacrificed their time and energy to complete the project work successfully.

MURUGAN M [20L231]
THARUN K [20L257]
VENKATESH K [20L259]

ABSTRACT

Chronic kidney disease (CKD) is a global health problem with high morbidity and mortality rate, and it induces other diseases. Since there are no conspicuous side effects during the beginning phases of CKD, patients regularly neglect to see the illness. Early discovery of CKD empowers patients to get opportune treatment to enhance the movement of this infection. Machine learning models can successfully help clinicians accomplish this objective because of their quick and precise acknowledgment execution. In this assessment, Logistic regression system for diagnosing CKD. The CKD data set are got from public resources using for case studies**,** which has a tremendous number of missing characteristics. Missing qualities are generally found, all things considered, clinical circumstances since patients may miss a few estimations for different reasons. In this research work Decision Tree, Random Forest and Logistic regression are used as classifier to classify whether a person has chronic kidney disease or not. The parameters namely accuracy, precision and recall are used for performance analysis of the diagnosing CKD.

TABLE OF CONTENTS

REFERNCE

APPENDIX I

APPENDIX II

APPENDIX III

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

S NO	ABBREVIATIONS	EXPANSIONS
1	CKD	Chronic Kidney Disease
2	UCI	University of California Irvine
3	AI	Artificial intelligent
4	ML	Machine Learning
5	GFR	Glomerular Filtration Rate
6	TP	True Positive
7	TN	True Negative
8	FP	False Positive
9	FN	False Negative

CHAPTER 1

INRODUCTION

Chronic Kidney Disease (CKD) is a pervasive and critical health condition globally, exerting a significant toll on individuals' well-being, particularly in regions with low-to-middle income levels. Sadly, millions succumb to its consequences regularly, primarily due to the absence of adequate treatment options. CKD progresses through various stages, and the severity of its impact on health outcomes is closely intertwined with the stage at which it remains undiagnosed and untreated.

One of the most alarming aspects of CKD is its insidious nature. Often, symptoms do not manifest until the disease has advanced to later stages, complicating early detection and intervention efforts. This delayed recognition exacerbates the already dire consequences associated with CKD. The risk factors for CKD are manifold, encompassing a range of conditions such as diabetes, hypertension, heart disease, and a family history of kidney failure. The increasing prevalence of these risk factors globally has further fueled the CKD epidemic.

Untreated CKD poses grave risks to an individual's health, primarily by predisposing them to hypertension and, in severe cases, kidney failure. Hypertension, or high blood pressure, is a common consequence of CKD and, in turn, exacerbates the progression of kidney damage. As CKD advances, the kidneys' ability to filter waste and excess fluids from the bloodstream diminishes, leading to a cascade of adverse health effects. Eventually, this can culminate in kidney failure, a life-threatening condition requiring dialysis or kidney transplantation for survival.

The ramifications of CKD extend far beyond its immediate health implications, impacting individuals' quality of life and placing a significant burden on healthcare systems worldwide. Early detection and management of CKD are therefore paramount in mitigating its deleterious effects and improving patient outcomes. This necessitates a concerted effort to raise awareness about CKD risk factors, enhance screening initiatives, and ensure equitable access to essential healthcare services, particularly in underserved communities where the burden of CKD is most pronounced.Kidney disease facts

- 37 million American adults have CKD, and millions of others areat increased risk
- Early detection can help prevent the progression of kidneydisease to kidneyfailure
- Heart disease is the primary cause of death for all people withCKD

People who develop chronic kidney disease may have some or allof the following testsand measurements.

1.1 Kidney Function Tests and Indicators

Serum Creatinine: Creatinine is a waste product in your blood that comes from muscleactivity. It is normally removed from your blood by your kidneys, but when kidney function slows down, thecreatinine level rises. Your doctor should use the results of your serum creatinine test to calculate your GFR.

Glomerular Filtration Rate (GFR): Your GFR tells how much kidney function you have. It may be estimated from your blood level of creatinine. If your GFR falls below 30 you will need to seea kidney disease specialist (called a nephrologist). A GFR below15 indicates that you need to start a treatment for kidney failure. Your kidney diseasespecialist will speak to you about treatments for kidney failure, such as dialysis or kidney transplant.

Blood Urea Nitrogen (BUN): Urea nitrogen is a normal wasteproduct in your blood that comes from the breakdown of proteinfrom the foods you eat and from your body metabolism. It isnormally removed from your blood by your kidneys, but whenkidneyfunction slows down, the BUN level rises. BUN can alsorise if you eat more protein, and it can fall if you eat less protein. **Urine Protein:** When your kidneys are damaged, protein leaksinto your urine. A simpletest can be done to detect protein in yoururine. Persistent protein in the urine is an early sign of chronickidney disease.

Microalbuminuria: This is a sensitive test that can detect a smallamount of protein inthe urine.

Urine Creatinine: This test estimates the concentration of your urine and helps to give an accurate protein result. Protein-to- Creatinine Ratio: This estimates the amount of protein you excretein your urine in a day and avoids the need to collect a 24-hour sample of your urine.

Serum Albumin: Albumin is a type of body protein made from the protein you eat eachday. A low level of albumin in your bloodmay be caused by not getting enough protein or calories from your diet. A low level of albumin may lead to health problems such as difficulty fighting off infections. Ask your dietitian how to get the right amount of protein and calories from your diet.

nPNA: Your nPNA (normalized protein nitrogen appearance) is atest that may tell if you are eating enough protein. This measurement comes from lab studies that include aurine collectionand blood work. Your dietitian may ask for an accurate food record to go with data.

Subjective Global Assessment (SGA): Your dietitian may use SGA to help check forsigns of nutrition problems. The dietitian will ask you some questions about your daily diet and check your weight and the fat and muscle stores in your face, hands, arms, shoulders and legs. Ask your dietitian about your score on the SGA. If your score is toolow, ask how to improveit.

Hemoglobin: Hemoglobin is the part of red blood cells that carries oxygen from yourlungs to all parts of your body. Your hemoglobinlevel tells your doctor if you have anemia, which makes you feel tired and have little energy.

Hematocrit: Your hematocrit is a measure of the red blood cells your body is making.A low hematocrit can mean you have anemia.**TSAT and Serum Ferritin:** Your TSAT (pronounced tee-sat) and serum ferritin (pronounced ferry-tin) are measures of iron in your body. Your TSAT should be above20 percent and your serum ferritin should be above 100. This will help you build red blood cells. Your doctor will recommend iron supplements when neededto reach yourtarget levels.

Parathyroid Hormone (PTH): High levels of parathyroid hormone (PTH) may result from a poor balance of calcium and phosphorus in your body. This can cause bone disease. Ask your doctor if your PTH level is in the right range. Your doctor may order aspecial prescription form of vitamin D to help lower your PTH. Caution: Do not take over-the-counter vitamin D unless ordered by your doctor.

Calcium: Calcium is a mineral that is important for strong bones.Ask your doctor what your calcium level should be. To help balance the amount of calcium in your blood, your doctor may ask you to take calcium supplements and a special prescription form of vitamin D. Take only the supplements and medications recommended by your doctor.

Phosphorus: A high phosphorus level can lead to weak bones. Ask your doctor whatyour phosphorus level should be. If your level is too high, your doctor may ask you toreduce your intake of foods that are high in phosphorus and take a type of medicationcalled a phosphate binder with your meals and snacks.

Potassium: Potassium is a mineral in your blood that helps your heart and muscles workproperly. A potassium level that is too highor too low may weaken muscles and change your heartbeat. Whether you need to change the amount of high- potassium foodsin your diet depends on your stage of kidney disease. Ask your doctor what your potassiumlevel should be. Your dietitian can helpyou plan your diet to get the right amount of potassium.

Body Weight: Maintaining a healthy weight is important to your overall health. If you are losing weight without even trying, you may not be getting the right nutrition to stay healthy. Your dietitiancan suggest how to safely add extra calories to your diet if needed.On the other hand, if you are slowly gaining too much weight, youmay need toreduce calories and increase your activity level. A sudden weight gain can also be a problem. If it is accompanied byswelling, shortness of breath and a rise in blood pressure, it may be a sign of too much fluid in your body. Speak to your doctor if your weight changes noticeably.

Blood Pressure: Ask your doctor what your blood pressure shouldbe. If your blood pressure is high, make sure to follow all the stepsin your prescribed treatment, which may include taking high bloodpressure medications, cutting down on the amount of saltin your diet, losing excess weight and following a regular exercise program.

Total Cholesterol: Cholesterol is a fat-like substance found in your blood. A high cholesterol level may increase your chance of having heart and circulation problems. Formany patients, a good level for total cholesterol is below 200. If your cholesterol level istoo high, your doctor may ask you to make some changes in your diet and increase your activity level. In some cases, medications are also used.

HDL Cholesterol: HDL cholesterol is a type of "good" cholesterol that protects yourheart. For many patients, the target level for HDLcholesterol is above 40.

LDL Cholesterol: LDL cholesterol is a type of "bad" cholesterol.A high LDL level may increase your chance of having heart and circulation problems. For many patients, the target level for LDL cholesterol is below 100. If your LDL levelis too high, your doctor may ask you to make some changes in yourdiet and increase your activity level.

Triglyceride: Triglyceride is a type of fat found in your blood. A high triglyceride levelalong with high levels of total and LDL cholesterol may increase your chance of heart and circulation problems.

Glomerulonephritis: Glomerulonephritis is a group of diseases that cause inflammationand damage the kidney's filtering units. These disorders are the third most common type of kidney disease.

Inherited diseases: Polycystic kidney disease, or PKD, is a common inherited diseasethat causes large cysts to form in the kidneys and damage the surrounding tissue.

Kidney and urinary tract abnormalities before birth: Malformations that occur as ababy develops in its mother's womb.For example, a narrowing may occur that preventsnormal outflow of urine and causes urine to flow back up to the kidney. This causes infections and may damage the kidneys.

Autoimmune diseases: When the body's defense system, the immune system, turnsagainst the body, it's called an autoimmune disease. Lupus nephritis is one such autoimmune disease that results in inflammation (swelling or scarring) of the small blood vessels that filter wastes in your kidney

Other causes: Obstructions caused by kidney stones or tumors can cause kidney damage. An enlarged prostate gland in men or repeated urinary infections can also causekidney damage

1.2 Stages of CKD

- **Stage 1:** There is kidney damage but the normal function ismaintained. This is why the condition can go unnoticed for a considerable time as there are not manysymptoms.

In Stage 1 of Chronic Kidney Disease (CKD), individuals experience kidney damage, but their kidneys are still able to function normally. This stage can often go unnoticed for a considerable period because there may be few to no symptoms present. Despite the underlying damage to the kidneys, their ability to filter waste and regulate fluids and electrolytes remains intact, allowing individuals to maintain normal kidney function. As a result, routine screenings and diagnostic tests are typically necessary to detect CKD at this early stage, emphasizing the importance of proactive healthcare monitoring, especially for individuals with known risk factors such as diabetes or hypertension. Early detection and intervention during Stage 1 can help slow the progression of CKD and prevent further complications, underscoring the significance of timely medical attention in managing this chronic condition.

- **Stage 2:** We begin to see a small loss in kidney function. The patient may noticemild symptoms such as fluid retention and highblood pressure.

In Stage 2 of Chronic Kidney Disease (CKD), there is a gradual decline in kidney function, indicating further progression from Stage 1. While the loss of kidney function is still relatively mild, individuals may begin to experience subtle symptoms such as fluid retention and high blood pressure. Fluid retention can manifest as swelling in the legs, ankles, or around the eyes, while high blood pressure may cause headaches, fatigue, or shortness of breath.

These symptoms, though mild, can serve as early warning signs of kidney dysfunction and should prompt individuals to seek medical evaluation. However, it's important to note that many people with Stage 2 CKD may still not experience noticeable symptoms, highlighting the importance of regular check-ups.

Managing Stage 2 CKD typically involves lifestyle modifications, such as dietary changes to reduce sodium intake and control blood pressure, as well as monitoring for any further decline in kidney function. By addressing risk factors and implementing appropriate interventions at this stage, healthcare providers can help slow the progression of CKD and minimize the risk of complications in the long term.

- **Stage 3:** Here we see mild to moderate loss of kidney function with more severesymptoms such as back pain and fatigue.

In Stage 3 of Chronic Kidney Disease (CKD), there is a progression to mild to moderate loss of kidney function, indicating a more advanced stage of the condition compared to Stages 1 and 2. At this stage, individuals may experience more severe symptoms, including back pain and fatigue.

Back pain can occur as a result of kidney inflammation or swelling, known as nephritis, which can lead to discomfort or pain in the lower back region. Fatigue is a common symptom in CKD due to the kidneys' reduced ability to effectively filter waste products from the blood, leading to an accumulation of toxins and metabolic by-products in the body. Additionally, anemia, a condition characterized by low red blood cell count, commonly occurs in CKD, contributing to feelings of tiredness and weakness.

While symptoms in Stage 3 CKD are more pronounced compared to earlier stages, they may still be overlooked or attributed to other causes, emphasizing the importance of regular medical monitoring and evaluation. Treatment at this stage focuses on managing symptoms, slowing the progression of kidney damage, and addressing underlying risk factors such as hypertension and diabetes through lifestyle modifications and medication. Early intervention during Stage 3 CKD can help improve outcomes and delay the need for more aggressive treatments such as dialysis or kidney transplantation in later stages of the disease.

- **Stage 4:** There is severe kidney function loss with symptoms suchas anaemia and bone disease.

In Stage 4 of Chronic Kidney Disease (CKD), there is severe loss of kidney function, representing a significant progression from earlier stages. At this advanced stage, individuals may experience debilitating symptoms such as anemia and bone disease. Anemia occurs due to a decrease in the production of red blood cells, which are responsible for carrying oxygen throughout the body. In CKD, the kidneys' impaired ability to produce erythropoietin, a hormone crucial for red blood cell production, can lead to anemia. Symptoms of anemia may include fatigue, weakness, dizziness, and shortness of breath, further compromising an individual's quality of life.

Additionally, CKD can disrupt the balance of minerals such as calcium and phosphorus in the body, leading to bone disease. The kidneys play a crucial role in maintaining the balance of these minerals, and their dysfunction can result in weakened bones, increased risk of fractures, and bone pain.

- **Stage 5:** Kidney function is either close to breakdown or lost completely. Patientswill need dialysis to get rid of toxins and will suffer from symptoms such as shortness of breath and nausea and vomiting.

In Stage 5 of Chronic Kidney Disease (CKD), also known as end-stage renal disease (ESRD), kidney function is severely compromised, nearing complete failure or already lost entirely. At this critical stage, patients require renal replacement therapy, such as dialysis, to eliminate toxins and waste products from the blood, as the kidneys are no longer able to perform this vital function adequately. Individuals with Stage 5 CKD often experience debilitating symptoms, including shortness of breath, nausea, and vomiting. Shortness of breath can occur due to fluid buildup in the lungs, known as pulmonary edema, as a result of fluid retention and impaired kidney function. Nausea and vomiting may result from the accumulation of toxins in the bloodstream, leading to gastrointestinal disturbances and general malaise.

Dialysis serves as a life-sustaining treatment for individuals with Stage 5 CKD, helping to remove excess fluids and waste products from the body and restore electrolyte balance. However, despite the benefits of dialysis, it can be associated with its own set of challenges and complications, including vascular access issues, infection risks, and dietary restrictions.

For patients with Stage 5 CKD, kidney transplantation may offer the best long-term solution, providing improved quality of life and survival compared to dialysis. However, transplantation is not always readily available or suitable for all patients, necessitating ongoing management and support to optimize outcomes and minimize complications associated with ESRD medical attention in managing this chronic condition.

Stage	Description	eGFR	Kidney Function
1	Possible kidney damage (e.g., protein in the urine) with ***normal*** kidney function	90 or above	90-100%
2	Kidney damage with ***mild loss*** of kidney function	60-89	60-89%
3a	***Mild to moderate*** loss of kidney function	45-59	45-59%
3b	***Moderate to severe*** loss of kidney function	30-44	30-44%
4	***Severe loss*** of kidney function	15-29	15-29%
5	Kidney ***failure***	Less than 15	Less than 15%

Table 1.1 Stages of CKD

-

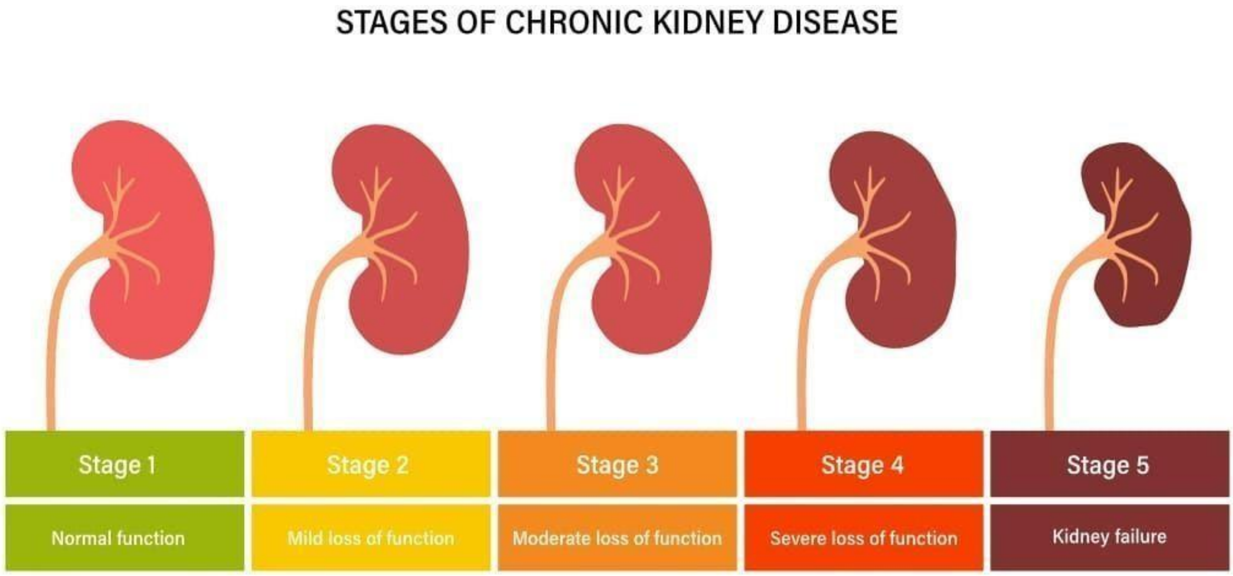

Fig 1.1 Stages of CKD

1.3 Symptoms of CKD

Your symptoms depend on how severe your kidney disease is.Symptoms of mild kidney disease:

- Needing to urinate several times at night
- Symptoms of moderate kidney problems
- Feeling tired, weak, confused, and less alert
- Being less hungry
- Shortness of breath
- Feeling sick to your stomach
- A bad taste in your mouth
- With severe kidney problems you may also have:
- Muscle weakness, twitching, cramping, and pain
- Not being able to feel things in some parts of your body
- Restless legs syndrome, which is when you have pain or other bad feelings inyour legs, usually at night, that make you need tomove them
- Problems with thinking, confusion, and seizures
- Severe shortness of breath
- Sometimes, a white powder on the skin left behind after yousweat

1.4 Complications

Chronic kidney disease can affect almost every part of your body.Potentialcomplications include:

- Fluid retention, which could lead to swelling in your arms and legs, high blood pressure, or fluid in your lungs (pulmonaryedema)

- A sudden rise in potassium levels in your blood (hyperkalemia), which couldimpair your heart's function and can be life-threatening
- Anemia
- Heart disease
- Weak bones and an increased risk of bone fractures
- Decreased sex drive, erectile dysfunction or reduced fertility
- Damage to your central nervous system, which can cause difficulty concentrating,personality changes or seizures
- Decreased immune response, which makes you more vulnerable toinfection
- Pericarditis, an inflammation of the saclike membrane that envelops your heart(pericardium)
- Pregnancy complications that carry risks for the mother and the developing fetus
- Irreversible damage to your kidneys (end-stage kidney disease), eventuallyrequiring either dialysis or a kidney transplantfor survival

We procured a standard data set from the UCI machinerepository for chronic Kidney Disease. CKD if predicted early andaccurately, can benefit patients in many ways. It increases the probability of a successful treatment while also adding years to theperson's life. This project aims to predict kidney disease by using some of the selected machine learning algorithms and feature selection methods. The objective is to collect the combination of different feature and then have used it as input to the machine learning algorithms. The algorithms have been implemented on thebasis of selected features and then we compare their performances.

Machine learning

Machine learning (ML) is a branch of artificial intelligence focused on developing algorithms that can learn and improve from experience and data without being explicitly programmed. These algorithms analyze and learn patterns from sample data, known as training data, to build predictive models or make decisions in various applications. Unlike traditional programming where explicit rules are defined, ML algorithms adapt and evolve based on the data they receive.

ML finds extensive applications across diverse fields, including medicine, email filtering, speech recognition, and computer vision. In medicine, ML algorithms are used to analyze patient data and assist in disease diagnosis, treatment planning, and drug discovery. Email filtering systems utilize ML to classify emails as spam or non-spam based on their content and user interactions. Speech recognition technologies employ ML to transcribe spoken words into text or recognize commands in virtual assistants. Additionally, ML plays a crucial role in computer vision tasks, enabling machines to interpret and analyze visual information from images or videos.

Overall, ML algorithms offer powerful tools to tackle complex problems and automate decision-making processes in situations where conventional programming approaches may be impractical or insufficient. As the volume and complexity of data continue to grow, the importance and applications of machine learning are expected to expand further across various domains.

1.5.Types of machine learning

Classical machine learning is often categorized by how an algorithm learns to become more accurate in its predictions. There are four basic approaches:supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning. The type of algorithm data scientists choose to use depends on what type of data they want to predict.

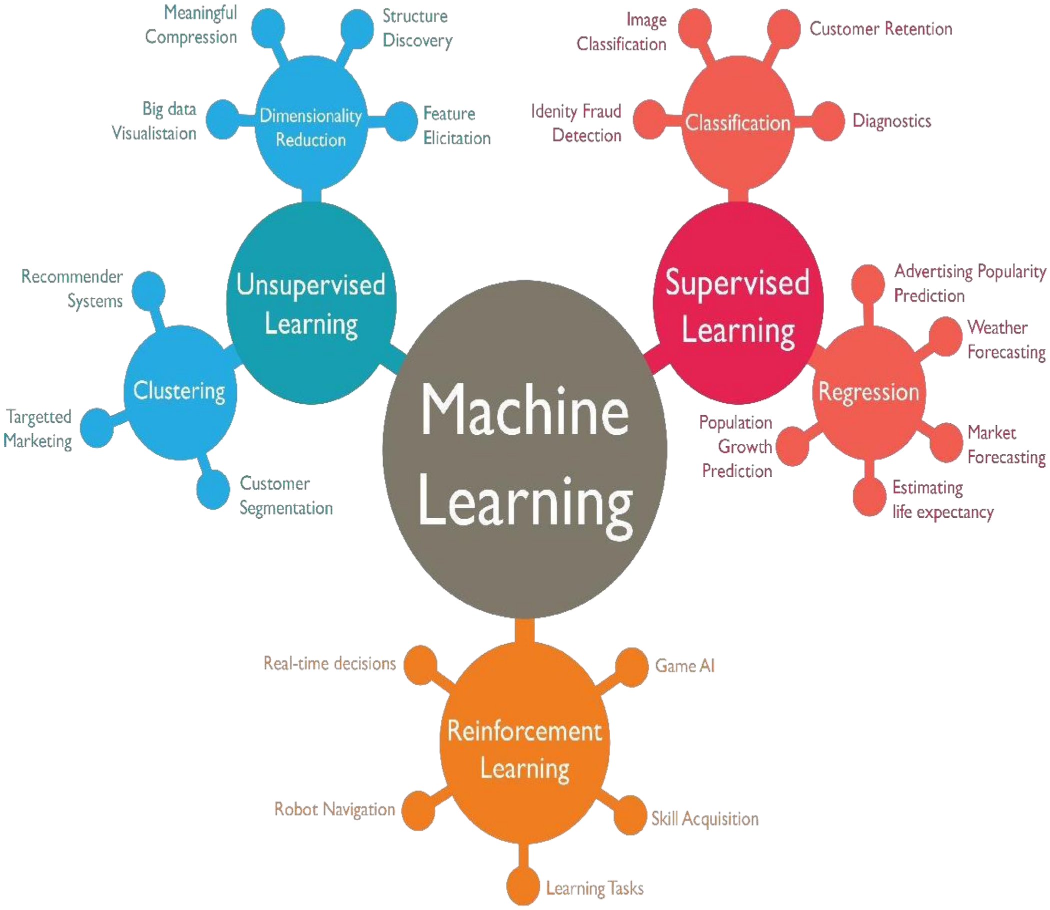

Fig 1.2 Types of Machine Learning

- **Supervised learning:** In this type of machine learning, data scientists supply algorithms with labeled training data and define the variables they want the algorithm to assess for correlations. Both the input and the output of the algorithm is specified.
- **Unsupervised learning:** This type of machine learning involves algorithms that train on unlabeled data. The algorithm scans through data sets looking for any meaningful connection. The datathat algorithms train on as well as the predictions or recommendations they output are predetermined.
- **Semi-supervised learning:** This approach to machine learning involves a mix of the two preceding types. Data scientists may feed an algorithm mostly labeled training data, but the model is free to explore the data on its own and develop its own understanding of the data set.
- **Reinforcement learning:** Data scientists typically use reinforcement learning to teach a machine to complete a multi-stepprocess for which there are clearly defined rules. Data scientists program an algorithm to complete a task and give it positive or negative cues as it works out how to complete a task. But for the most part, the algorithm decides on its own what steps to take alongthe way.

CHAPTER 2

Literature review

This section provides a comprehensive review of the relevant literature on the prediction of CKDs using a range of machine learning and deep learning techniques. Within this review, we also highlight the limitations identified in previous works, shedding light on areas where further research and improvement are needed:

M.A. Islam et al. (2023)8 conducted a study on the early detection of CKD using machine learning. They worked with a dataset of 400 cases, featuring 24 attributes 13 categorical and 11 numerical. After preprocessing, Principal Component Analysis (PCA) was applied to determine key features for CKD prediction. The XgBoost classifier outperformed other algorithms, reaching an accuracy of 98.33% with the original data and improving to 99.16% after PCA was applied. Other classifiers also achieved an accuracy of 98.33% before PCA.

R. Sawhney et al. (2023) developed AI models to predict and assess CKD, using a dataset with 400 cases and 24 features, both categorical and numerical. They utilized a Multilayer Perceptron (MLP) with backpropagation, integrating two feature extraction and three feature selection techniques to improve efficiency. The Artificial Neural Network (ANN) model they created outperformed other classifiers, achieving a perfect testing accuracy of 100%, significantly higher than the Logistic Regression (LR)and Support Vector Machine (SVM) scores of 96% and 82%, respectively.

Alsekait et al. (2023)6 developed an ensemble deep learning model to predict CKD using a dataset of 400 cases with 24 features. The process involved data preprocessing, including label encoding and outlier detection, followed by feature selection through methods like mutual information and Recursive Feature Elimination (RFE). The model used a stacked approach combining RNN, LSTM, and GRU models, with a Support Vector Machine (SVM) for meta-learning. This model achieved high performance metrics, with an accuracy, precision, recall, and

F1 score all around 99.69%.

Arif M.S. et al. (2023)10 designed a machine learning model to predict CKD, incorporating advanced preprocessing, feature selection using the Boruta algorithm, and hyperparameter optimization. Their method involved iterative imputation for missing values and a novel sequential data scaling technique that included robust scaling, z-standardization, and min-max scaling. The model, tested on the UCI CKD dataset with 400 cases and 24 features, achieved an accuracy rate of 100% using k-Nearest Neighbors (KNN) algorithm and grid-search CV for optimization.

Poonia RC et al. (2022)11 developed a feature-based model for kidney disease detection using a dataset with 400 cases and 24 features. They employed machine learning algorithms like KNN, ANN, SVM, and Naive Bayes, along with Recursive Feature Elimination (RFE)and chi-Square tests for feature selection. The model was evaluated on a dataset with healthy and diseased individuals. A logistic regression model, optimized with Chi-Square selected features, achieved the highest accuracy of 98.75%.

Pal. S. (2022)12 developed a model to predict CKD using a dataset of 400 instances with 24 features from the UCI machine learning repository. The study applied Logistic logistic Regression (LR), Decision Tree (DT), and Support Vector Machine (SVM) classifiers, and improved model performance with a bagging ensemble method. The Decision Tree (DT) classifier achieved the highest accuracy of 95.92%, which increased to 97.23% after implementing the bagging method.

K.M. Almustafa (2021)13 proposed a classification system for kidney diseases using a dataset of 400 cases with 24 features. Various machine learning classifiers were tested, with J48 and Decision Tree (DT) performing best, achieving 99% accuracy. Post-feature selection, these classifiers, along with Naive Bayes and KNN, showed improved accuracy, indicating the effectiveness of feature selection in enhancing model performance.

Ilyas et al. (2021)14 developed a decision tree-based diagnosis system for CKDs using a dataset with 400 cases and 24 features. They utilized J48 and Random Forest (RF) algorithms, highlighting the importance of the Glomerular Filtration Rate (GFR) calculated using the CKD-EPI equation. The model was assessed across five CKD stages, with J48 and RF performing well, particularly in early stages, with accuracies up to 98%. The performance slightly decreased in more advanced stages.

DATA SET

This project used a dataset from the UCI machine learning repository. We haverecords of 100 patients in this dataset. The data has 25attributes as follows:

No	Attributes	Type of Attribute	Explanation
1	age	numerical	age
2	bp	numeric	blood pressure
3	sg	nominal	specific gravity
4	al	nominal	albumin
5	su	nominal	sugar
6	rbc	nominal	red blood cells
7	pc	nominal	pus cell
8	pcc	nominal	pus cell clumps
9	ba	nominal	bacteria
10	bgr	numeric	blood glucose random
11	bu	numeric	blood urea
12	sc	numeric	serum creatinine
13	sod	numeric	sodium
14	pot	numeric	potassium
15	home	numeric	hemoglobin
16	pcv	numeric	packed cell volume
17	wc	numeric	white blood cell count
18	rc	numeric	red blood cell count
19	htn	nominal	hypertension
20	dm	nominal	diabetes mellitus
21	cad	nominal	coronary artery disease
22	appet	nominal	appetite
23	pe	nominal	pedal edema
24	ane	nominal	anemia
25	class	nominal	class

Table 2.1 Dataset Description

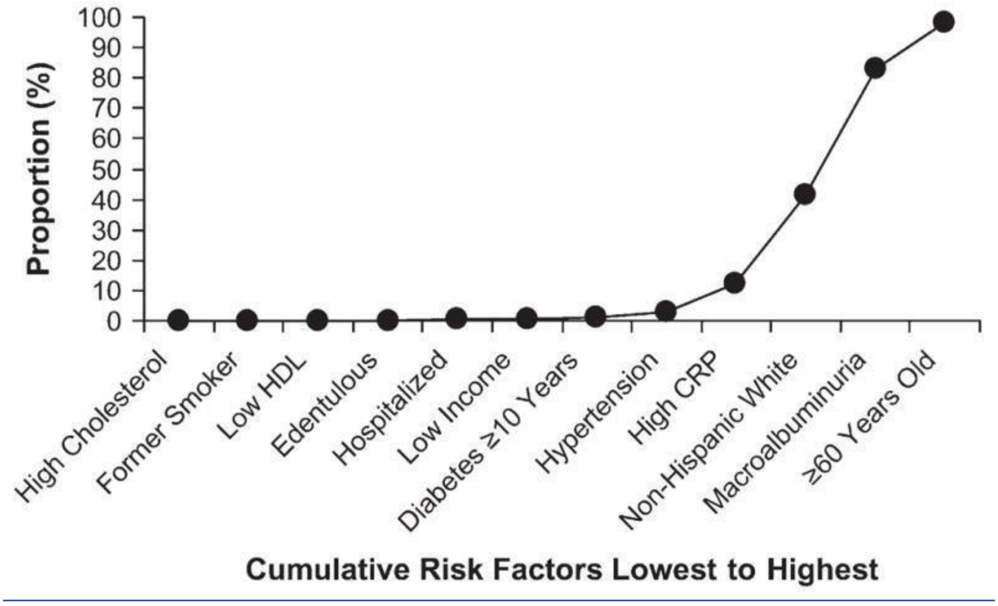

Fig 2.1 Proportion of individuals with chronic kidney disease.

Cumulative risk factors lowest to highest.

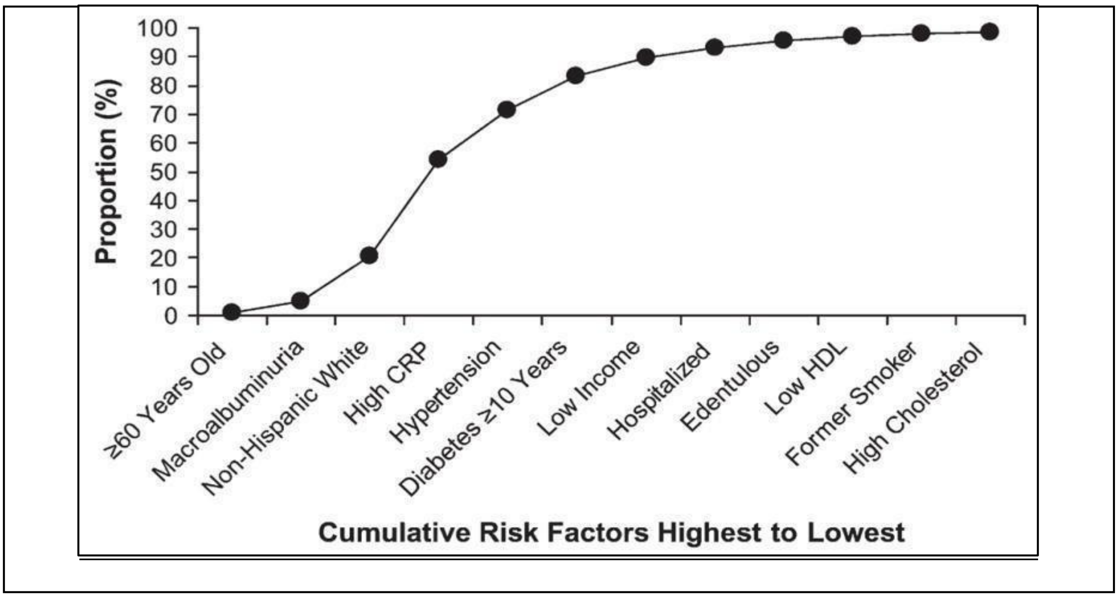

Fig 2.2 Proportion of individuals with chronic kidney disease.

Cumulative risk factors highest to lowest.

CHAPTER 3

PROPOSED SYSTEM

Data preprocessing could be a strategy that is utilized to change over the raw information into a clean data set. It is the basicstep to train every machine learning classifier algorithm. Data Cleaning: This involves handling missing values, outliers, and noisy data. Missing values can be filled using techniques like mean, median, or mode imputation, or by using more advanced methods such as interpolation or predictive modeling. Outliers and noisy data can be detected and either removed or corrected.

Data Transformation: Data transformation involves converting data into a more suitable format for analysis. This may include scaling features to a similar range (e.g., normalization or standardization), transforming categorical variables into numerical representations (e.g., one-hot encoding), or handling skewed distributions through techniques like logarithmic transformation.

Feature Selection/Extraction: Feature selection involves choosing the most relevant features for model training, while feature extraction involves creating new features from the existing ones. This helps reduce dimensionality and improve model performance by focusing on the most informative features.

Data Integration: In cases where data is collected from multiple sources, data integration involves combining and merging datasets to create a unified dataset for analysis. This ensures that all relevant information is available for model training.

Data Splitting: Before training a machine learning model, the dataset is typically divided into training, validation, and testing sets. This ensures that the model is trained on one subset, validated on another, and evaluated on a separate unseen subset to assess its generalization performance.This technique concludes such actions as

- Handle missing values
- Recalling of the data set,
- Transform into binary data

- Standardize of the data set. When the data set included attributeswith varying scales,recalling is used to scale the data set.

The binary transformation has been applied to con- vert thevalue into 0 and 1. All values of every attribute are considered as for above the threshold and as 0 for below the threshold. Standardized method ensures that each attribute has mean 0 and standard deviation 1.

3.1.DATA MINING AND PREPROCESSING

The dataset this peoject used which received from the internet source needs to be cleaned as it has null or NA values for various attributes for a giveninstance. Missing values in the dataset like NA's or blank values are removed by using Pandas library "dropna" and "fillna", whichdrops either column or rows with missing data and replaces NA's with the mean values of that attribute respectively.

Pandas Library: Pandas is a powerful Python library for data manipulation and analysis. It provides data structures and functions to efficiently handle structured data, such as tables and time series.

dropna() Method: This method is used to remove rows or columns with missing values (NaN or NA) from a DataFrame. By default, it drops rows containing any missing values, but you can specify axis=1 to drop columns instead. This helps in cleaning the dataset by eliminating instances with incomplete information.
fillna() Method: This method is used to fill missing values (NaN or NA) in a DataFrame with specified values. One common approach is to fill missing values with the mean of the respective attribute. This helps in retaining data completeness while handling missing values appropriately.
Missing Values (NaN or NA): Missing values occur when no data is stored for a particular attribute in a dataset. They can arise due to various reasons such as data entry errors, equipment malfunction, or non-response in surveys. Handling missing

values is crucial for data preprocessing to ensure accurate and reliable analysis and modeling.

Attributes Details: This refers to the specific details or characteristics of the attributes in the dataset. It includes information such as attribute names, data types, possible values, and descriptions of what each attribute represents. Understanding the attributes is essential for data cleaning, transformation, and analysis tasks

3.2.FEATURE SELECTION METHODS

The performance of every model is heavily impacted with the use of feature selection models. They drastically change the attributes that will be used to train the model as they remove attributes that do not have high impact on the data or negatively impact the data. It is a core concept in machine learning.

Feature selection and Data cleaning are supposed to be the very first thing to do inmodel designing. In this paper we will applytwo feature selection algorithms to obtain important features according to those. Then we would run our classification algorithms on the attributes are deemed important and relevant byboth the selection

The dataset utilized in this research was obtained from the publicly available CKD collection hosted on the UCI Machine Learning Repository.15 This dataset, with a file size of 43 KB, is specifically tailored for research in the field of medical diagnosis, particularly focusing on CKD. This dataset, donated on July 2, 2015, by L. Rubini, P. Soundarapandian, and P. Eswaran, is specifically designed for predicting CKD and was collected over approximately 2 months from a hospital. The dataset is characterized by the following features

a.Multivariate nature: It includes various types of data.

b. Associated tasks: Primarily for classification tasks.

c. Feature types: Real-valued features.

d. Number of instances: 400.

e. Number of features: 25, including demographic, clinical, and laboratory data.

The dataset contains 24 features plus a class attribute, with a mix of 11 numerical and 14 nominal variables. The class label indicates the presence or absence of CKD; '1' denotes a positive CKD diagnosis, while '0' signifies a negative diagnosis. Presence of missing data can significantly impact the performance and accuracy of machine learning models. The dataset seems to have missing values (NaN) in several columns. Handling these missing values through techniques like imputation or deletion will be crucial. Columns like rbc, pc, pcc, ba, htn, dm, cad, appet, pe, ane, and classification appear to be categorical. Depending on the machine learning algorithm used, these may need to be encoded into numerical values.

FLOW DIAGRAM

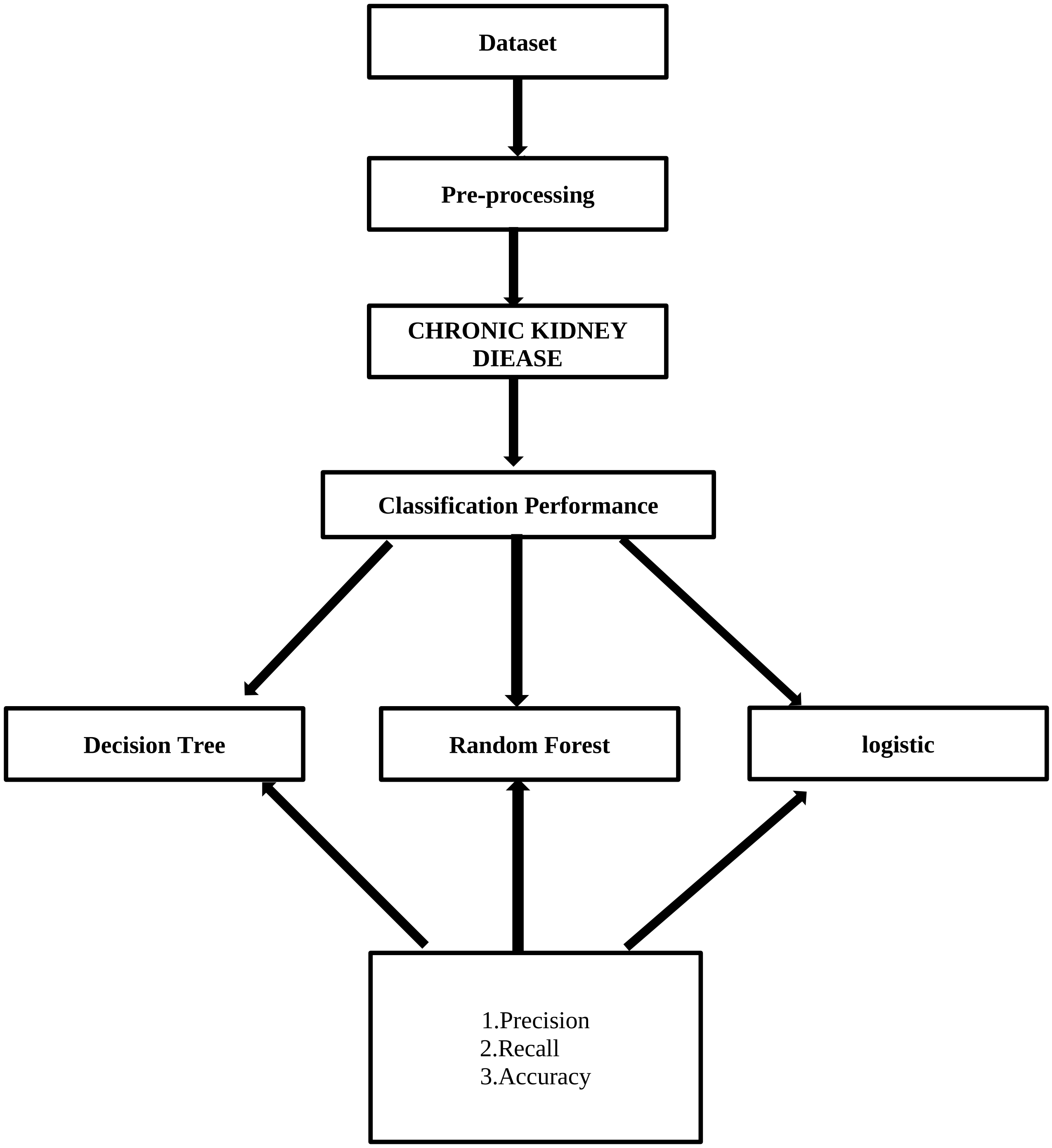

Fig 3.1 Flow Diagram

CHAPTER 4

CLASSIFICATION ALGORITHM

Classification refers to mapping data of the dataset into predefined groups or classes (predicting the class of unclassified data). Classification is a process related to categorization where weclassify a collection of data into different groups. Here we use binary classification that has two classes. So a person would eithertest positive or negative for CKD. In this research work Decision Tree, Random Forest and Logistic regression are used as classifierto classify whether a person has chronic kidney disease or not.

Binary classification is a specific type of classification where there are only two possible classes or outcomes. In this scenario, the classes are "positive" (indicating the presence of CKD) and "negative" (indicating the absence of CKD).

4.1.Decision Tree:

In decision trees Classification models are built in a tree structure. The dataset is splitinto smaller subsets and we incrementally establish an associated decision tree. The final outcome of the algorithm is a tree which has decision nodes .Theleaf nodes depicts theclassification or class. Decision nodes have two or more branches. The uppermostdecision node in a tree is called root node. Decision trees can work with both numerical and categorical data.

Tree Structure: Decision tree classification models are built in a tree-like structure, where each internal node represents a decision based on an attribute, each branch represents the outcome of that decision, and each leaf node represents a class label or classification.

Dataset Splitting: The dataset is split into smaller subsets based on the values of attributes. The decision tree algorithm identifies the most informative attribute at each step to split the dataset into subsets that are as pure as possible in terms of class.

Incremental Building: The decision tree is incrementally built by recursively partitioning the dataset into subsets based on the values of attributes. At each step, the algorithm selects the attribute that best separates the data into homogeneous classes or minimizes impurity.

Decision Nodes and Leaf Nodes: Decision nodes are the internal nodes of the tree where decisions are made based on attribute values. These nodes have two or more branches, each corresponding to a possible value of the attribute. Leaf nodes are the terminal nodes of the tree that represent the final classification or class label.

Root Node: The uppermost decision node in a decision tree is called the root node. It represents the initial decision based on the most informative attribute in the dataset.

Handling Numerical and Categorical Data: Decision trees can handle both numerical and categorical data. For numerical attributes, the algorithm selects a threshold to split the data into two subsets, while for categorical attributes, it creates branches for each category.

Classification at Leaf Nodes: The final outcome of the decision tree algorithm is a tree structure where each leaf node represents a class label or classification decision. When a new instance is presented to the trained decision tree, it traverses the tree from the root node to a leaf node based on the attribute values of the instance, and the class label associated with the leaf node is assigned as the predicted class.

- With training data D., starts with single node N.
- N becomes leaf if all the data in D belongs to same class. Otherwise attribute ‘A’ isselection method based on the splittingcriterion.
- The instance in ‘D’ is partitioned accordingly.
- Apply algorithm recursively to each subset in ‘D’ to all theother subset in ‘D’ to form decision tree.

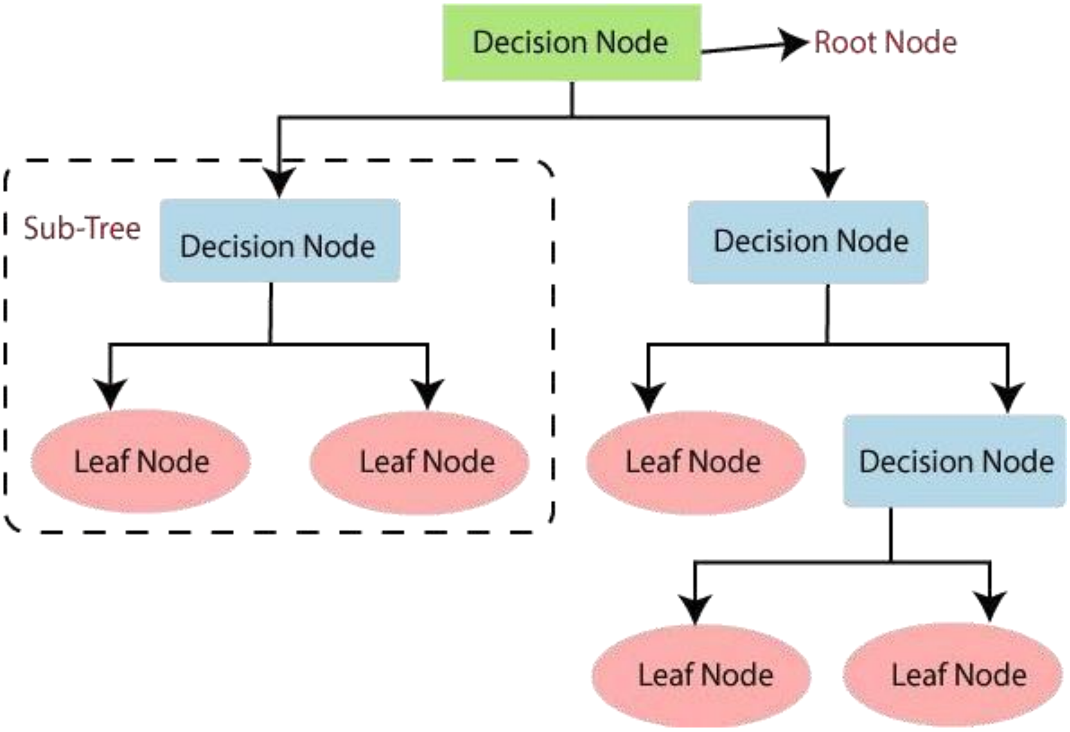

Fig 4.1 Decision Tree

4.2.Random forest:

This algorithm creates a forest of decision trees. It is also a supervised learningclassification algorithm. The number of trees inthe forest is an indication of the forest's robustness. Also the accuracy of the algorithm is directly proportional to the number oftrees in the forest.

Forest of Decision Trees: Random Forest creates an ensemble of decision trees, where each tree is trained independently on a random subset of the training data and a random subset of the features. These decision trees collectively form a "forest" of trees.

Supervised Learning: Random Forest is a supervised learning algorithm, meaning it requires labeled data for training. Each instance in the dataset is associated with a class label, and the algorithm learns to classify new instances based on patterns observed in the training data.

Number of Trees: The number of trees in the forest is a hyperparameter that can be adjusted by the user. Increasing the number of trees typically leads to a more robust and accurate model, as it reduces overfitting and improves generalization to unseen data.

Robustness of the Forest: The robustness of the Random Forest is indicated by the number of trees in the forest. Having a larger number of diverse trees helps to capture a wider range of patterns and variations in the data, making the model more robust to noise and outliers.

Accuracy and Number of Trees: The accuracy of the Random Forest algorithm is generally directly proportional to the number of trees in the forest, up to a certain point. Adding more trees tends to improve the accuracy of the model by reducing variance and improving the stability of the predictions. However, there may be diminishing returns in accuracy with increasing numbers of trees, and computational resources should be considered.

- First we randomly select ‘p’ features from the total ‘q’ features (where p<<q)
- From the selected’ ‘p’ features, we need to calculate a node, referred as ‘d’ using the method best split point. Using the best split, we need to split the node into daughter nodes.
- Then we repeat the above steps till a number ‘l’ is reached
- A forest of ‘n’ number of trees is built by applying the above steps‘n’ number of times.

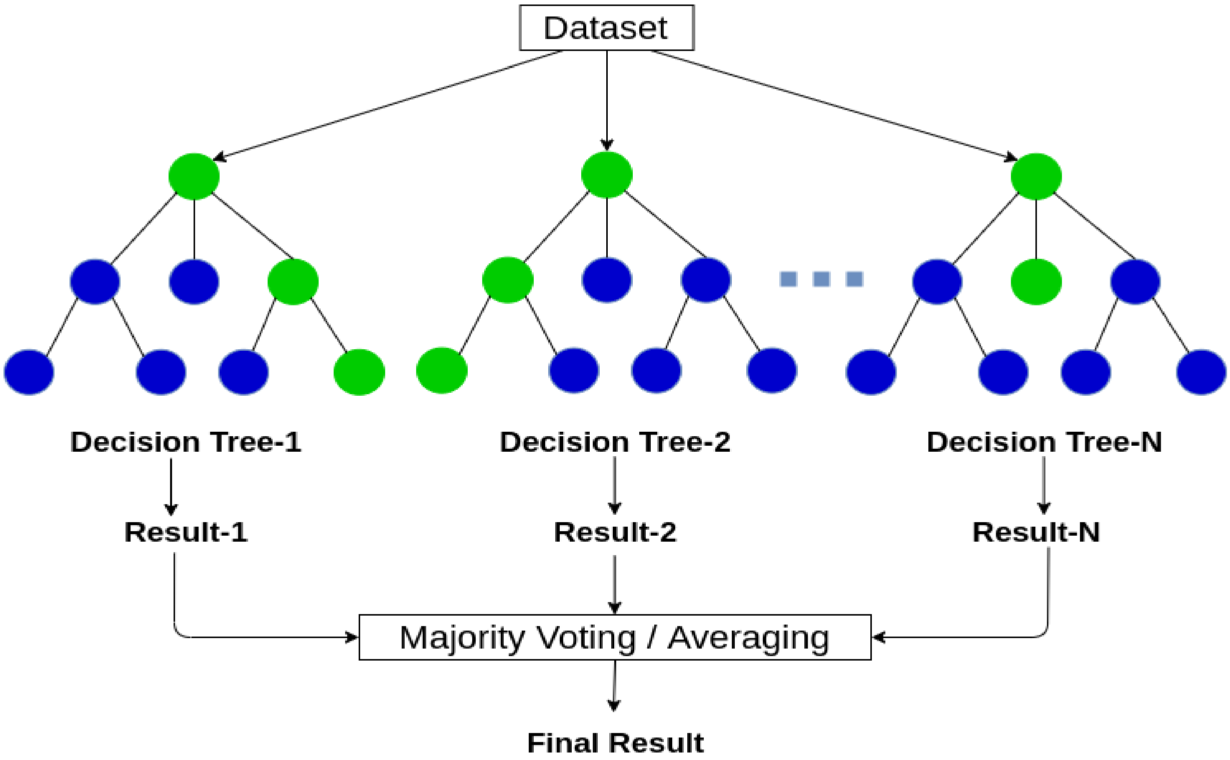

Fig 4.2 Random Forest

4.3.Logistic Regression:

The core method or the heart of logistic regression is the logisticfunction, which is also referred to as the sigmoid function .Another type of regression, i.e. linear regression predicts continuous dependent variable. Logistic regression on the other hand predicts categorical dependent variables with the use of a set containing independent variables.

Logistic Function (Sigmoid Function): The logistic function, also known as the sigmoid function, is the core method used in logistic regression. It maps any real-valued number to a value between 0 and 1, making it suitable for modeling probabilities. The formula for the logistic function is :$f(x) = 1/1+e^{-x}$ where e is the base of the natural logarithm, and x is the input variable.

Prediction of Categorical Dependent Variables: Unlike linear regression, which predicts continuous dependent variables, logistic regression predicts categorical dependent variables. The dependent variable in logistic regression is binary or ordinal (e.g., yes/no, 0/1, low/medium/high), and logistic regression models the probability that an instance belongs to a particular category.

Independent Variables: Logistic regression uses a set of independent variables (also known as features or predictors) to make predictions about the probability of the categorical outcome. These independent variables can be numerical or categorical and are used to estimate the log-odds of the event being observed.

Estimating Probabilities: Logistic regression estimates the probability that a given instance belongs to a particular category using the logistic function. It models the relationship between the independent variables and the log-odds of the categorical outcome through a linear combination of the predictors, transformed by the logistic function.

Decision Boundary: In logistic regression, a decision boundary is used to separate instances belonging to different categories. The decision boundary is determined based on the estimated probabilities, and instances with probabilities above a certain threshold are classified into one category, while those below the threshold are classified into the other category.

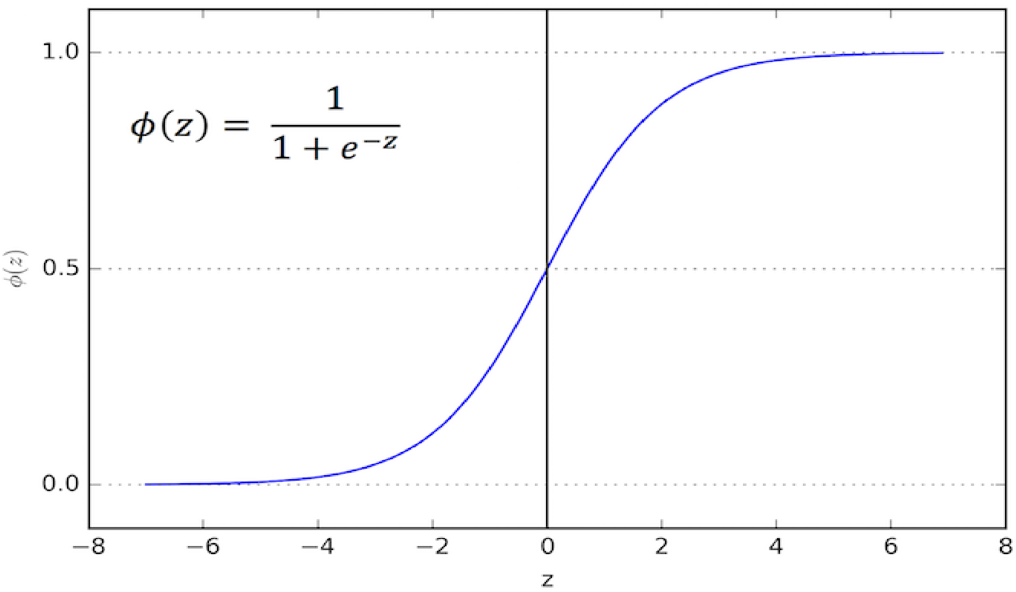

Fig 4.3 Logistic Regression

CHAPTER 5

RESULT AND DISCUSSION

To measure the performance of our algorithms we will use tree different evaluation functions. In these functions TP refers to casesthat were positive and were predicted positive by the algorithms. TN as negative cases predicted negative. FP are casespredicted positive but were actually negative. FN represents cases that were predicted negative but were actually positive.

True Positives (TP): True positives are cases where the actual class is positive (e.g., presence of a disease, occurrence of an event), and the algorithm correctly predicts them as positive.

True Negatives (TN): True negatives are cases where the actual class is negative (e.g., absence of a disease, non-occurrence of an event), and the algorithm correctly predicts them as negative.

False Positives (FP): False positives are cases where the actual class is negative, but the algorithm incorrectly predicts them as positive. It's also known as a Type I error. For example, predicting a healthy person as having a disease.

False Negatives (FN): False negatives are cases where the actual class is positive, but the algorithm incorrectly predicts them as negative. It's also known as a Type II error. For example, predicting a diseased person as healthy.

Accuracy:

It is defined as the ratio of correctly predicted observation(truenegative + true positive) to the total observations:

$$\text{Accuracy}=\frac{TP+TN}{TP+FP+FN+TN}$$

Precision:

Precision defines the proportion of predicted positives that areactually positive. It is therefore the ratio of true positives(TP) to the number of cases predicted positive(TP + FP):

$$\text{Precision}=\frac{TP}{TP+FP}$$

Recall:

Recall defines the proportion of actual positives that were predicted positive. It is the ratio of true positives(TP) to the actualpositives (TP + FN):

$$\text{Recall}=\frac{TP}{TP+FN}$$

COMPARISION CHART

5.1.ACCURACY

The final accuracy of the three classification algorithms used are 98.48%, 99.0% and 97.61%

S. No.	Algorithm	Accuracy %
1	Decision Tree	98
2	Random Forest	99
3	Logistic Regression	97

Table 5.1.Accuracy

Fig 5.1 Accuracy Comparison

In terms of accuracy:

- Logistic Regression offers decent accuracy and is suitable for linearly separable data with well-separated classes.

- Decision Trees can achieve high accuracy by capturing complex non-linear relationships in the data but may overfit if not pruned properly.

- Random Forests tend to provide high accuracy by combining predictions from multiple decision trees, making them suitable for a wide range of classification tasks, including those with complex data patterns and imbalanced classes.

5.2.PRECISION

The final precision of the three classification algorithms we used are:98.0 %,95.12% and 98.82%

S. No.	Algorithm	Accuracy %
1	Decision Tree	98
2	Random Forest	95
3	Logistic Regression	98

Table 5.2.Precision

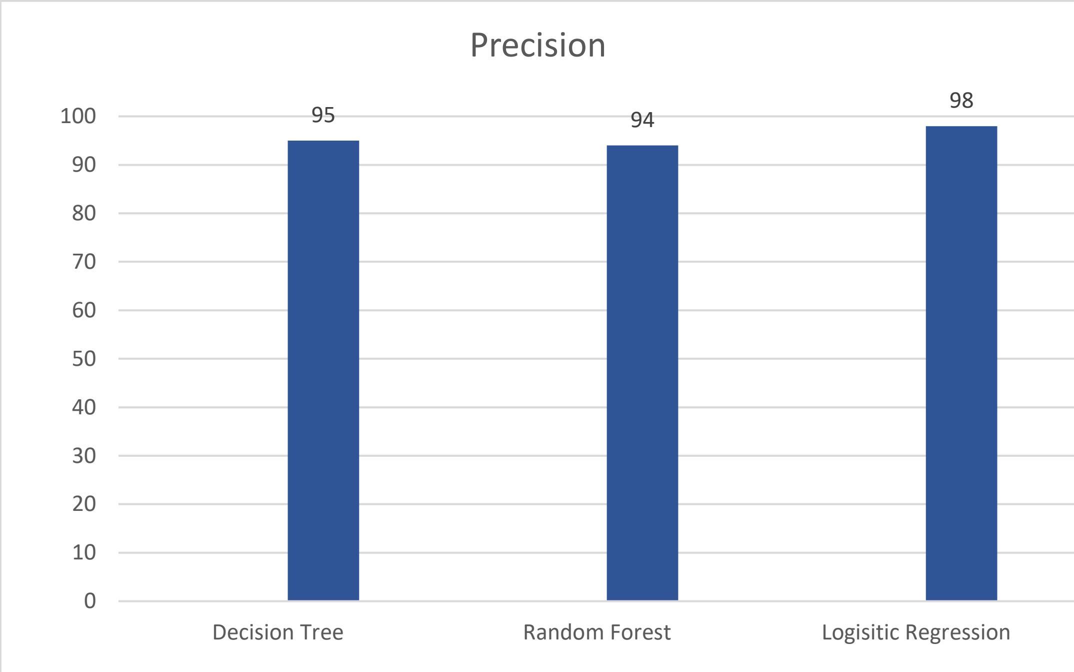

Fig 5.2 Precision Comparison

In terms of precision:

- Logistic Regression can provide good precision, but its performance in recall may be limited by its linear nature.
- Decision Trees can achieve high precision and recall, especially when they are pruned appropriately to avoid overfitting.
- Random Forests tend to offer high precision and recall, making them suitable for a wide range of classification tasks, including those with imbalanced classes or complex data patterns.

5.3.RECALL

The final recall of the three classification algorithms we used are: 95.61%, 94.29% and 98.0%

S. No.	Algorithm	Accuracy %
1	Decision Tree	95
2	Random Forest	94
3	Logistic Regression	98

Table 5.3.Recall

Fig 5.3 Recall Comparison

In terms of recall:

- Random forests tend to perform well in terms of recall because they can capture complex relationships in the data, handle noise and outliers effectively, and reduce overfitting.
- Decision trees can also perform well in terms of recall, especially when the data has complex interactions between predictors and the target variable. However, they may be more prone to overfitting compared to random forests.
- Logistic regression may not perform as well in terms of recall compared to decision trees and random forests, especially when the relationships in the data are non-linear and complex.

CHAPTER 6

CONCLUSION

Employing machine learning algorithms for the classification of Chronic Kidney Disease (CKD) offers a promising approach for diagnosis and prediction. Through the utilization of various algorithms such as Decision Trees, Random Forests, Logistic Regression, and K-Nearest Neighbors, valuable insights can be derived from the CKD dataset to aid in medical decision-making.

Each algorithm brings its own strengths to the table. Decision Trees provide interpretable results, making them useful for understanding the decision-making process. Random Forests excel in handling complex datasets and reducing overfitting. Logistic Regression offers simplicity and efficiency, especially for binary classification tasks like CKD diagnosis. K-Nearest Neighbors leverages the similarity of instances to make predictions, which can be beneficial in scenarios where the decision boundaries are not linear.

Through meticulous data preprocessing, algorithm selection, model training, and evaluation, the best-performing classifier can be identified and deployed for practical use. Performance metrics such as accuracy, precision, recall, and F1-score serve as benchmarks for assessing the effectiveness of the models.

By harnessing the power of machine learning, healthcare professionals can enhance their diagnostic capabilities, improve patient outcomes, and contribute to the advancement of medical Research in the field of chronic kidney disease.

FUTURE SCOPE

- Feature selection based on attributes (age , gender., etc.,) . Featureselectionis the process of reducing the number of input variables when developing apredictive model.
- It is desirable to reduce the number of input variables to both reduce thecomputational cost of modelling and, in some cases, to improve the performance of the model.
- Extracting feature vectors or predictors could remove variables that are neither useful for prediction nor related to response variables and thus prevent these unrelated variables the models to make an accurate prediction .
- Here in, we used optimal subset regression and LR to extract the variablesthat are most meaningful to the prediction.

REFERENCES

1. M. Z. U. Haq, M. Z. Khan, H. U. Rehman et al., “An adaptive topology management scheme to maintain network connectivity in Wireless Sensor Networks,” *Sensors*, vol. 22, no. 8, p. 2855, 2022.
2. M. Fayaz, G. Mehmood, A. Khan, S. Abbas, M. Fayaz, and J. Gwak, “Counteracting selfish nodes using reputation based system in mobile ad hoc networks,” *Electronics*, vol. 11, no. 2, p. 185, 2022.
3. S. Khan, M. Z. Khan, P. Khan, G. Mehmood, A. Khan, and M. Fayaz, “An ant-hocnet routing protocol based on optimized fuzzy logic for swarm of UAVs in FANET,” *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 6783777, 12 pages, 2022.
4. S. Khan, M. Z. Khan, P. Khan, G. Mehmood, A. Khan, and M. Fayaz, “An ant-hocnet routing protocol based on optimized fuzzy logic for swarm of UAVs in FANET,” *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 6783777, 12 pages, 2022.
5. M. Ghassemi, T. Naumann, P. Schulam, A. L. Beam, I. Y. Chen, and R. Ranganath, ‘‘A review of challenges and opportunities in machine learning for health,’’ in Proc. AMIA JointSummits Transl. Sci., 2020, p. 191.
6. J. Qin, L. Chen, Y. Liu, C. Liu, C. Feng, and B. Chen, ‘‘A machine learning methodology for diagnosing chronic kidney disease,’’ IEEE Access, vol. 8, pp.20991–21002, 2020
7. L. Kilvia De Almeida, L. Lessa, A. Peixoto, R. Gomes, and J. Celestino, ‘‘Kidney failure detection using machine learning techniques,’’ in Proc. 8th Int. Workshop ADVANCEs ICT Infrastructures Services, 2020, pp. 1–8.
8. S. Shankar, S. Verma, S. Elavarthy, T. Kiran, and P. Ghuli, ‘‘Analysis and prediction of chronic kidney disease,’’ Int. Res. J. Eng. Technol., vol. 7, no. 5, May 2020, pp. 4536–4541.
9. P. G. Scholar, ‘‘Chronic kidney disease prediction using machine learning,’’ Int. J.Eng. Res. Technol., vol. 9, no. 7, pp. 137–140, 2020.

10. A. S. A. Rabby, R. Mamata, M. A. Laboni, Ohidujjaman, and S. Abujar, "Machine learning applied to kidney disease prediction: comparison study," in *Proceedings of the 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kanpur, India, July 2019.
11. S. Revathy, B. Bharathi, P. Jeyanthi, and M. Ramesh, "Chronic kidney disease prediction using machine learning models," *International Journal of Engineering and Advanced Technology*, vol. 9, no. 1, pp. 6364–6367, 2019.
12. M. S. Basarslan and F. Kayaalp, "Performance analysis of fuzzy rough set-based and correlation-based attribute selection methods on detection of chronic kidney disease with various classifiers," in *Proceedings of the 2019 Scientific Meeting on Electrical-Electronics and Biomedical Engineering and Computer Science (EBBT)*, IEEE, Istanbul, Turkey, April 2019.
13. M. Kumar, "Prediction of chronic kidney disease using random forest machine learning algorithm," *International Journal of Computer Science and Mobile Computing*, vol. 5, pp. 24–33, 2016.
14. M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, "Intelligent heart disease prediction system using random forest and evolutionary approach," *Journal of network and innovative computing*, vol. 4, pp. 175–184, 2016.
15. S and D. S, "Data Mining Classification Algorithms for Kidney Disease Prediction," Int. J. Cybern. Informatics, vol. 4, no. 4, pp. 13– 25, Aug. 2015, doi: 10.5121/ijci.2015.4402.
16. A. Nway Oo, "Classification of Chronic Kidney Disease (CKD) Using Rule based Classifier and PCA," Int. J. Adv. Manag. Technol. Eng. Sci., vol. 8, no. 4, pp. 728– 733, 2018.
17. Dua, Dheeru and Graff, and Casey, "UCI Machine Learning Repository," 2017.[Online].

APPENDIX I

BANNARI AMMAN INSTITUTE OF TECHNOLOGY

Stay Ahead

An Autonomous Institution, Affiliated to Anna University, Approved by AICTE, Accredited by NAAC with 'A+' Grade, Sathyamangalam- 638 401, Erode(dt), Tamilnadu, India

Certificate of Participation

This is to certify that

Prof/Dr/Mr/Ms. **MURUGAN M** from **KARPAGAM COLLEGE OF ENGINEERING** has presented a paper titled **EXPLORING PERSPECTIVES IN CHRONIC KIDNEY DISEASES USING MACHINE LEARNING** in **2nd International Conference on Innovative Trends in Engineering and Sciences (ICITES '24)** held during 26-27 March 2024 at Bannari Amman Institute of Technology, Sathyamangalam.

Organizer Secretary

Convener

Principal

A1924IC853

Stay Ahead

BANNARI AMMAN INSTITUTE OF TECHNOLOGY

An Autonomous Institution, Affiliated to Anna University, Approved by AICTE, Accredited by NAAC with 'A+' Grade,
Sathyamangalam- 638 401, Erode(dt), Tamilnadu, India

Certificate of Participation

This is to certify that

Prof/Dr/Mr/Ms. **THARUN K** from **KARPAGAM COLLEGE OF ENGINEERING** has presented a paper titled **EXPLORING PERSPECTIVES IN CHRONIC KIDNEY DISEASES USING MACHINE LEARNING** in **2nd International Conference on Innovative Trends in Engineering and Sciences (ICITES '24)** held during 26-27 March 2024 at Bannari Amman Institute of Technology, Sathyamangalam.

Organizer Secretary | Convener | Principal

A1924IC550

BANNARI AMMAN INSTITUTE OF TECHNOLOGY

Stay Ahead

An Autonomous Institution, Affiliated to Anna University, Approved by AICTE, Accredited by NAAC with 'A+' Grade,
Sathyamangalam- 638 401, Erode(dt), Tamilnadu, India

Certificate of Participation

This is to certify that

Prof/Dr/Mr/Ms. **VENKATESH K** from **KARPAGAM COLLEGE OF ENGINEERING** has presented a paper titled **EXPLORING PERSPECTIVES IN CHRONIC KIDNEY DISEASES USING MACHINE LEARNING** in **2nd International Conference on Innovative Trends in Engineering and Sciences (ICITES '24)** held during 26-27 March 2024 at Bannari Amman Institute of Technology, Sathyamangalam.

V. S

Organizer Secretary

Convener

Principal

A1924IC318

APPENDIX II

KARPAGAM COLLEGE OF ENGINEERING

20L259_VENKATESH K ECE_B <20l259@kce.ac.in>

Intimation for abstract ICITES-24-351 approval - Reg.,

International Conference <icites@bitsathy.ac.in> Sat, Mar 23, 2024 at 9:37 AM
To: 20l259@kce.ac.in

Greetings from ICITES 2024

We are pleased to inform you that your paper based on your abstract has been **accepted** by the ***2nd International Conference on Innovative Trends in Engineering and Sciences (ICITES) -2024*** during March 26 - 27, 2024. The Committee now needs to have confirmation from you that you will be able to submit your full paper to us by 24 March 2024 and that you will be able to present your paper in a 15 minute time slot during Open Session at the Conference.

When we receive confirmation (Registration) from you, we will send you the detailed schedule for your presentation. Form that will provide the conference organisers with the full details of your paper so that the information can be incorporated into the conference programme. We are looking forward to hearing from you.

NOTE:

Abstract ID: **ICITES-24-351**

Kindly complete the registration as quickly as possible

Registration Link: https://forms.gle/FBbsXmazWNtH43wQ9

Conference WhatsApp Group:**https://chat.whatsapp.com/LwxVLHIlaKyJ2PKWfJd5Vh**

Mode of Presentation:

https://forms.gle/fgK4NkNnhZnwX29N6

Online Payment Details:

Bank Name : **Axis Bank**

Branch Name: **Sathyamangalam**

Acc. Name : **BIT - Centre for Advanced Studies**

Acc. No : **911010036862582**

IFSC : **UTIB0000368**

Fee Details (Including 18% GST):

Industry Delegate: Rs.1200/-

Research Scholar & Faculty Members:Rs.1000/-

UG & PG Students: Rs.600/- (Per Team -Max of 3 Nos),

With Warmest Regards

Best Regards,

BANNARI AMMAN INSTITUTE OF TECHNOLOGY
Stay Ahead

Dr ANANDHA MOORTHY A | ASP
Mechanical Engineering
Bannari Amman Institute of Technology
Sathyamangalam, Erode - 638 401
+91 94884 70455 | 04295 226000
anandhamoorthya@bitsathy.ac.in

APPENDIX III

EXPLORING PERSPECTIVES IN CHRONIC KIDNEY DISEASES USING MACHINE LEARNING

by Venkatesh

Submission date: 26-Jan-2024 11:53PM (UTC-0500)
Submission ID: 2278783078
File name: b11_finalconference2024.docx (200.76K)
Word count: 2575
Character count: 15247

ORIGINALITY REPORT

10%	7%	8%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	H. Muthukrishnan, V. Sneharathna, Jeevanantham Arumugam, A. Panneerselvam, K Lalitha., D. Vinothini. "Machine Learning For Chronic Kidney Disease Detection", 2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT), 2022 Publication	3%
2	Afia Farjana, Fatema Tabassum Liza, Partha Pratim Pandit, Madhab Chandra Das, Mahadi Hasan, Fariha Tabassum, Md. Helal Hossen. "Predicting Chronic Kidney Disease Using Machine Learning Algorithms", 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), 2023 Publication	2%
3	Rohit K. Singla, Matthew Kadatz, Robert Rohling, Christopher Nguan. "Kidney Ultrasound for Nephrologists: A Review", Kidney Medicine, 2022 Publication	1%