Unlocking Diagnostic Potential: Integrating Multivariate Chemical Patterns for Kidney Disease Recognition

[1]Dr. Indira S P, [2]Prema S P

[1]Associate Professor, [2]Lecturer
Department of Chemistry
Davanagere, Karnataka, India.

Abstract- **CKD (Chronic Kidney Disease) is a condition where the kidneys gradually lose their ability to filter waste and excess fluids from the blood. It is typically a long term and progressive condition that can lead to various health complications. Treatment involves managing underlying causes making lifestyles changes and sometimes dialysis or kidney transplantation. Detecting CKD in its early stages is crucial for saving millions of millions lives. The importance of this article is to developing a diagnosis system to detect kidney disease of chronic by estimating the statistical analysis of chemical pattern of Glomerular Filtration Rate (GFR) and vitamin-'D' based on modified supervised learning algorithms.**

Keywords: **Medicine, Recognition, Chemical Patterns, Feature Selection, CKD,Machine Learning, Glomerular Filtration Rate and Vitamin- 'D'.**

I INTRODUCTION

Kidneys play a crucial role in maintaining the body's internal stability by filtering blood, regulating blood pressure, eliminating waste and excess water, as well as contributing to red blood cell production and vitamin D metabolism. They are described as reddish-brown bean-shaped organs located in the back of the abdomen, each approximately 4 to 5 inches long. Kidney disease can impair the kidneys' ability to perform their functions effectively, leading to issues such as inefficient blood filtration, improper water balance, high blood pressure, anemia due to reduced red blood cell production, and compromised vitamin D metabolism. In the context of chronic kidney disease (CKD), features refer to specific characteristics or factors that can influence the disease. These features include factors such as blood pressure (BP), which medical professionals utilize to diagnose and manage CKD, thereby tailoring treatment plans based on individual situations. The article mentions the importance of developing a recognition system to detect CKD by analyzing statistical features such as Glomerular Filtration Rate (GFR) and vitamin D levels. This involves the application of modified supervised learning algorithms to estimate and predict these factors accurately. The research aims to predict metabolic indices with high accuracy, which can be combined with standard risk assessment methods to enhance the detection and management of CKD. Fig-1.shows the Factors affecting the chronic kidney disease.

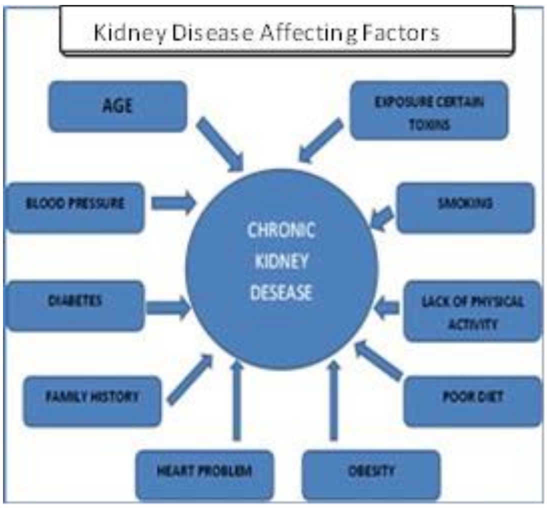

Fig.1. Kidney Disease Affecting Factors.

- Blood Pressure (BP): Elevated blood pressure is a significant risk factor for CKD. Hypertension can lead to damage to the blood vessels in the kidneys, impairing their function over time.
- Glomerular Filtration Rate (GFR): GFR is a measure of how effectively the kidneys filter waste from the blood. A decrease in GFR indicates impaired kidney function and can be indicative of CKD.
- Vitamin D Metabolism: Proper metabolism of vitamin D is essential for bone health. In CKD, impaired kidney function can lead to abnormalities in vitamin D metabolism, contributing to bone disorders such as osteoporosis.
- Anemia: Anemia, or a decrease in red blood cell count, is a common complication of CKD. Impaired kidney function reduces the production of erythropoietin, a hormone necessary for red blood cell production.
- Fluid and Electrolyte Balance: The kidneys play a crucial role in maintaining the body's fluid and electrolyte balance. CKD can disrupt this balance, leading to fluid retention, electrolyte imbalances, and related complications.
- Proteinuria: Proteinuria, or the presence of protein in the urine, is a common sign of kidney damage. It indicates dysfunction in the glomeruli, the filtering units of the kidneys.
- Diabetes Mellitus: Diabetes is a leading cause of CKD. High blood sugar levels can damage the blood vessels in the kidneys, leading to diabetic nephropathy and progressive kidney damage.
- Smoking: Smoking is a modifiable risk factor for CKD. It can accelerate the progression of kidney disease and increase the risk of cardiovascular complications.
- Obesity: Obesity is associated with an increased risk of CKD. Excess body weight can contribute to hypertension, diabetes, and other conditions that can damage the kidneys.
- Family History: A family history of kidney disease or related conditions can increase an individual's risk of developing CKD.
- Medication Use: Certain medications, particularly non-steroidal anti-inflammatory drugs (NSAIDs) and some antibiotics, can cause kidney damage if used excessively or inappropriately.
- Age: Advanced age is a risk factor for CKD. As people age, the kidneys undergo natural structural and functional changes, making them more susceptible to damage.

Understanding and addressing these factors are crucial for the early detection, prevention, and management of chronic kidney disease. By identifying individuals at risk and implementing appropriate interventions, the progression of CKD can be slowed, and complications can be minimized.

II RELATED WORK

[1]The proposed work by M. H. A. Elhebir and A. Abraham in 2015 introduces a novel approach for comparing ensemble models. The study utilizes a case study focused on the learning process to demonstrate this approach. Specifically, they extract streams of event logs from a learning execution environment. These event logs are then formatted in a way that enables effective mining and enhanced process analysis of the collected data. Ensemble models in machine learning refer to the combination of multiple individual models to improve predictive performance. These models could be of different types or trained on different subsets of data. The idea is that by aggregating the predictions of multiple models, the ensemble can achieve better accuracy and robustness than any individual model. Ensemble models in machine learning refer to the combination of multiple individual models to improve predictive performance. These models could be of different types or trained on different subsets of data. The idea is that by aggregating the predictions of multiple models, the ensemble can achieve better accuracy and robustness than any individual model. In the context of the proposed study, the researchers likely compared various ensemble models to determine which ones perform best for analyzing the learning process based on the extracted event logs. They would have evaluated factors such as prediction accuracy, computational efficiency, and the ability to extract meaningful insights from the data. The utilization of event logs from a learning execution environment suggests a focus on educational data mining or learning analytics. By analyzing the sequences of events generated during the learning process, researchers can gain insights into student behavior, learning patterns, and the effectiveness of instructional strategies.

[2]The work by P. Kathuria and B. Wedro, published in 2016, provides a quick overview of chronic kidney disease (CKD) in the context of healthcare. Additionally, the authors propose the use of a Support Vector Machine (SVM) classification algorithm for diagnosing CKD.SVM is a powerful supervised learning algorithm used for classification and regression tasks. It works by finding the optimal hyper plane that separates data points belonging to different classes with the maximum margin. SVM has been widely used in various fields, including healthcare, for its ability to handle high-dimensional data and nonlinear relationships. In the context of diagnosing CKD, SVM can be employed to classify patients into different categories based on their medical features and symptoms, helping healthcare professionals make informed decisions about treatment and management. Furthermore, the authors mention the utilization of wrapper and filter approaches to reduce the dimensionality of the CKD dataset. These approaches are commonly used in feature selection, a process that aims to identify the most relevant features or variables that

contribute to the predictive accuracy of a model while reducing computational complexity and over fitting. Wrapper methods evaluate the performance of a subset of features by training and testing a model on different subsets and selecting the one that yields the best performance. On the other hand, filter methods assess the relevance of features independently of the classifier and select them based on certain criteria, such as correlation with the target variable or information gain. By applying wrapper and filter approaches, the dimensionality of the CKD dataset is reduced, improving the efficiency and effectiveness of the SVM classification algorithm for diagnosing CKD. This approach allows for more accurate and reliable predictions while minimizing the risk of over fitting and computational burden.
[3]The work by S. Saremi, S. Mirjalili, and A. Lewis, published in 2017, introduces the Grasshopper Optimization Algorithm (GOA) along with its theory and application in solving optimization problems. The Grasshopper Optimization Algorithm is a nature-inspired meta heuristic optimization algorithm that mimics the behavior of grasshoppers in nature. This algorithm is based on the swarming behavior of grasshoppers, where individuals communicate and collaborate to find the optimal solution to a problem. In the context of optimization, the GOA works by iteratively updating the position of grasshoppers in the search space, with each grasshopper representing a potential solution. The movement of grasshoppers is influenced by their own position, the positions of other grasshoppers, and the global best solution found so far. The GOA aims to balance exploration and exploitation, allowing it to efficiently explore the search space to find promising regions and exploit them to refine the solution further. This makes the GOA suitable for solving a wide range of optimization problems, including continuous, discrete, and mixed-variable optimization problems.
The proposed algorithm has been applied to various real-world optimization problems across different domains, including engineering, finance, and computational biology. By harnessing the collective intelligence of grasshoppers, the GOA offers a competitive alternative to traditional optimization algorithms, providing solutions that are often comparable or superior in terms of accuracy and efficiency. Overall, the Grasshopper Optimization Algorithm contributes to the field of optimization by providing a versatile and effective tool for solving complex optimization problems, ultimately leading to improved decision-making and resource utilization in various applications. The work by G.A. Afzali and S. Mohammadi, published in 2018 in the IET Information Security journal, introduces a privacy-preserving approach for big data mining, specifically focusing on association rule mining. The proposed technique utilizes fuzzy logic to anonymize sensitive data while still allowing for meaningful analysis.
[4]G.A. Afzali and S. Mohammadi, "Privacy preserving big data mining: association rule iding using fuzzy logic approach," Association rule mining is a data mining technique used to discover interesting relationships or associations between variables in large datasets. However, in many cases, the data used for mining may contain sensitive information that needs to be protected to comply with privacy regulations or ethical considerations. To address this challenge, Afzali and Mohammadi propose a data anonymization technique based on fuzzy logic. Fuzzy logic allows for the representation of uncertainty and imprecision in data, which is particularly useful for preserving privacy while still enabling meaningful analysis. The proposed approach likely involves transforming the original dataset into a fuzzy representation, where individual data points are fuzzified to hide sensitive information. This transformation aims to preserve the statistical properties and patterns of the data while obscuring sensitive attributes.By anonymizing the data using fuzzy logic, the authors aim to strike a balance between privacy protection and data utility. This allows organizations to perform data mining and analysis on sensitive datasets without compromising individuals' privacy.
[5]The work by C. Xiong, M. Su, Z. Jiang, and W. Jiang, published in 2019 in the Journal of Medical Systems, focuses on predicting the timing of hemodialysis using LVW (Local Vector Weighting) feature selection and ensemble learning techniques.Hemodialysis is a medical procedure used to remove waste products and excess fluids from the blood when the kidneys are unable to perform this function adequately. Predicting the optimal timing for initiating hemodialysis is crucial for improving patient outcomes and managing resources efficiently in healthcare settings. By leveraging LVW feature selection and ensemble learning techniques, the authors aim to develop a predictive model capable of accurately determining the optimal timing for initiating hemodialysis. This can help healthcare providers make informed decisions regarding patient care and resource allocation, ultimately leading to better outcomes for patients with kidney disease.
[6]The work by S. Ravizza et al., published in 2019 in Nature Medicine, focuses on predicting early risk using real-world data. The authors aim to develop predictive models that can identify individuals at risk of certain health conditions or events using real-world data sources.Real-world data refers to data collected outside of traditional clinical trials, including electronic health records, claims data, patient registries, wearable devices, and other sources. These data sources provide valuable insights into patients' health status, treatment patterns, and outcomes in real-world settings.
[7]The work by E.H. A. Rady and A. S. Anwar, published in 2019 in the journal Informatics in Medicine Unlocked, represents a significant effort to leverage data mining techniques for predicting kidney disease stages. The primary objectives of their study are to extract hidden information from medical datasets related to kidney disease and to develop predictive models that can assist healthcare professionals in diagnosing and managing the condition more

effectively.Data mining algorithms are instrumental in this endeavor, as they enable the analysis of large and complex datasets to uncover patterns, trends, and relationships that may not be readily apparent through traditional statistical methods. In the context of kidney disease, these algorithms can be applied to diverse sources of data, including medical records, patient demographics, laboratory test results, and imaging data, among others. By scrutinizing these data sources, data mining algorithms can identify key features or variables that play a crucial role in the progression of kidney disease and its various stages. the study represents a significant advancement in leveraging real-world data for predicting early risk and informing clinical decision-making. By harnessing the power of real-world data and predictive modeling, the authors aim to improve patient care and population health outcomes.

[8]The work by L. Jerlin Rubini and E. Perumal involves the development and simulation of a kidney disease classification method using MATLAB. This method likely aims to classify individuals into different categories or stages of kidney disease based on various features or biomarkers. By simulating the kidney disease classification method in MATLAB, the researchers aim to develop an accurate and reliable tool for diagnosing and categorizing kidney disease, which can potentially aid healthcare professionals in patient management and treatment decision-making.

III EXISTING TECHNOLOGY

The use of regression analysis in conjunction with blood test results has significantly improved the precision and dependability of predictions in illness diagnosis. Regression techniques, including linear regression, logistic regression, and other advanced regression models, are commonly employed to analyze the relationships between various factors in blood test results and the likelihood of developing a specific illness. The integration of regression analysis with blood test results enhances the diagnostic accuracy, precision, and reliability in illness prediction. By identifying patterns and relationships in complex datasets, regression techniques empower healthcare professionals to make informed decisions and provide personalized care to patients. Furthermore, an empirical evaluation of machine learning (ML) techniques for chronic kidney disease (CKD) prediction is conducted. This study aims to assess the performance of various ML algorithms in predicting the occurrence or progression of CKD. Scientific Reports aims to establish clinical thresholds for diagnosing iron deficiency by comparing the functional assessment of serum ferritin levels to population-based centiles [9][10].The study contributes to the advancement of non-invasive and cost-effective approaches to support the diagnosis of diabetes mellitus, potentially improving patient outcomes and healthcare resource utilization. Further the study contributes to the advancement of predictive analytics in healthcare by developing machine learning prediction models for chronic kidney disease using real-world data from national health insurance claims[11][12].The study contributes to the advancement of predictive analytics in healthcare by comparing intelligent machine learning methods for predicting kidney disease, with potential implications for improving diagnosis, prognosis, and treatment planning for patients with kidney disease. Next the study contributes to the advancement of diagnostic methods for CKD by utilizing machine learning algorithms and RFE techniques for accurate classification of individuals into CKD and non-CKD categories[13][14].

The study contributes to the advancement of predictive analytics in healthcare by developing machine learning models for predicting (end-stage kidney disease) ESKD in CKD patients, with potential implications for improving patient care and outcomes. the study contributes to the field of predictive analytics in healthcare by leveraging machine learning techniques for the prediction of chronic kidney disease. The development of accurate predictive models can potentially improve patient outcomes and inform clinical decision-making in the management of CKD[15][16].

The findings of the study have clinical implications for the early detection and management of CKD. By developing accurate predictive models using AI algorithms, healthcare providers can identify individuals at risk of CKD and monitor disease progression more effectively, leading to improved patient outcomes and personalized treatment strategies. Overall, the study contributes to the field of nephrology by leveraging AI algorithms for the prediction of CKD and its progression, potentially improving patient care and management strategies in the field of kidney diseases. The findings of the study underscore the transformative potential of AI algorithms in revolutionizing the early detection and management of chronic kidney disease. By harnessing the predictive power of AI, healthcare providers can identify at-risk individuals, personalize treatment approaches, and improve outcomes for patients with CKD. The study contributes to the understanding of risk prediction models for CKD and provides insights into their development, validation, and application in clinical practice. By summarizing the progress in this field, the study may inform future research efforts aimed at improving the early detection and management of CKD. The study contributes to the understanding of risk prediction models for CKD and provides insights into their development, validation, and application in clinical practice. By summarizing the progress in this field, the study may inform future research efforts aimed at improving the early detection and management of CKD[17][18].

The performance of the diagnostic prediction and classification models is evaluated using standard evaluation metrics, such as accuracy, sensitivity, specificity, area under the receiver operating characteristic curve (AUC-ROC), and other relevant measures. Cross-validation techniques may be applied to assess the generalization performance of the models. the study contributes to the growing body of research on predictive analytics in healthcare by leveraging machine learning algorithms for the risk prediction of chronic kidney disease. The development of accurate predictive models can potentially improve patient outcomes by enabling early intervention and personalized management strategies for individuals at high risk of CKD[19][20][21]. The study contributes to our understanding of CKD screening methods and their utility in identifying individuals at risk of kidney disease, particularly among specific populations such as staff members of healthcare institutions. The primary objective of the study is to enhance the classification accuracy of machine learning models by employing a technique called Recursive Feature Elimination (RFE) with cross-validation. The study aims to improve the performance of classification algorithms by systematically selecting the most informative features and optimizing model parameters through cross-validation[22][23].

IV OBJECTIVES

The objectives outlined involve utilizing machine learning techniques to analyze the blood creatinine indication and investigating the relationship between vitamin D and estimated glomerular filtration rate (eGFR). The eGFR is calculated using a specific formula.

- Utilizing Machine Learning Techniques: The first objective involves applying various machine learning techniques, including classification, regression, support vector machines (SVM), random forests, and light generalized estimating equations (GEE), to analyze the blood creatinine indication. These techniques can help in understanding patterns, correlations, and predictive relationships between blood creatinine levels and other variables of interest.
- Investigating the Connection between Vitamin D and eGFR: The second objective is to investigate the relationship between vitamin D levels and estimated eGFR. This involves analyzing data to determine if there is a correlation or association between vitamin D levels and eGFR values. The eGFR is typically calculated using a formula that takes into account factors such as serum creatinine levels, age, gender, and race.
- eGFR Calculation: As part of the analysis, the formula used to determine eGFR will be applied. The eGFR formula is commonly derived from the Modification of Diet in Renal Disease (MDRD) Study equation or the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation. These equations use variables such as serum creatinine levels, age, gender, and race to estimate kidney function.

By achieving these objectives, the study aims to gain insights into the relationship between blood creatinine indication, vitamin D levels, and estimated eGFR. This information can be valuable for understanding kidney function, identifying potential risk factors for kidney disease, and informing clinical decision-making related to kidney health.

V PROPOSED METHODOLOGY

(eGFR) and vitamin-'D' levels can have a complicated and multifaceted connection.The importance of estimated filtration rate of glomerular and vitamin-'D' levels are given below separately.

•A tool for evaluating kidney function is eGFR. It calculates how quickly the glomeruli, or the small blood veins in the kidneys, filter blood. generally, the serum creatinine level, age, and gender are used to compute the eGFR. Reduced kidney function is indicated by a decreased eGFR value, which may be a symptom of renal illness or malfunction.

•Vitamin -'D' is a necessary nutrient that is essential for maintaining bone health, boosting immunity function and controlling the metabolism of calcium and phosphate. Vitamin-'D2' (ergocalciferol) and vitamin-'D3' (cholecalciferol) are its two primary forms. Both forms of vitamin-'D' can be acquired through food sources or supplements. Vitamin-'D3' is predominantly created in the skin when it is exposed to sun light. In the liver and kidneys, vitamin-'D' undergoes a number of metabolic reactions before becoming its active form, called calcitriol.

Vitamin-'D' levels and eGFR are related in the following ways:

•Low vitamin-'D':- levels have been linked in certain studies to lower eGFR and a higher risk of developing renal disease, according to some research. Individuals with CKD have been found to be vitamin-'D' deficient, which is probably caused by a problem with vitamin-'D' absorption in the kidneys. It's crucial to remember that comorbid conditions including irritation, malnutrition, and low vitamin-'D' levels may also have an impact on the association between low eGFR and low vitamin-'D' levels.

•The metabolism and activation of vitamin-'D':- might be impacted by poor renal function, on the other hand. Since the kidneys are where calcitriol, the active form of vitamin-'D', is made, decreased renal function can result in insufficient amounts of active vitamin-'D'. Mineral and bone problems, secondary hyperparathyroidism, and renal osteodystrophy can all be influenced by this.

•Studies examining the relationship between vitamin-'D': administration and eGFR have produced conflicting findings. While other studies have not identified a strong correlation, others have hypothesized that vitamin-‘D’ intake may have a protective impact on kidney function. It’s important to keep in mind that these studies frequently differ in terms of research design, demographic characteristics, vitamin-‘D’ doses, and length of supplementation.

Algorithm

Input: Training dataset $S = \{(x_1, y_1), \ldots, (x_2, y_2), \ldots\ldots, (x_n, y_n)\}$ g, base learner h, the number of iterations T.

Step 1: Determine the mass of trial:x_i: $\mathrm{D}_1(i) = \frac{1}{n}$

Step 2: Select a training data subset X from S, fit h using X to get a delicate classifier h_t .

Step3: Let $\mathrm{n}+$ and $\mathrm{n}-$ indicate the positive and negative classes respectively.

Calculate the error rate ε_t of the base learner for both the positive class

$\varepsilon_t{}^n$: $\varepsilon_t = \frac{\varepsilon_t^p + \varepsilon_t^n}{2}$, where $\varepsilon_t^p = P[h_t(x_i) \neq y_i]$

$= \sum_{i=1}^{n+} \mathrm{D}_t(\mathrm{i})\mathrm{I}[\,\mathrm{h}_t(x_i) \neq y_i\,]$ and $\sum_{i=1}^{n-} \mathrm{D}_t(\mathrm{i})\mathrm{I}[\,\mathrm{h}_t(x_i) \neq y_i\,]$

Step 4: Calculate the mass of h_t : $\alpha_t = \frac{1}{2}\ln\frac{(1-\varepsilon_t)x^2}{\varepsilon_t}$

Step 5: Updation of masses of all the instances in S: $\mathrm{D}_{t+1}(i)) = \frac{\mathrm{D}_t(i)}{\mathrm{Z}_t}\exp\left(-\alpha_t\,\beta_i\,\gamma_i\,h_t(x_i)\right)$

Where Z_t is a normalization factor and is calculated as fallows

$\mathrm{Z}_t = \sum_{i=1}^{n} \mathrm{D}_t(\mathrm{i})\exp\left(-\alpha_t\,\beta_i\,\gamma_i\,h_t(x_i)\right)$ and β_i is calculated as is shown in step 6 :

Step 6: β_i is calculated as

$$\beta_i = \left\{\begin{array}{l} \frac{\mathrm{TP}_t}{\mathrm{FP}_t + \mathrm{TP}_t}\mathrm{C}_{10,}\ if\ \mathrm{y}_i = 1, \mathrm{h}_t(\mathrm{x}_i) = -1 \\ \frac{\mathrm{TN}_t}{\mathrm{FN}_t + \mathrm{TN}_t}\mathrm{C}_{01,}\ if\ \mathrm{y}_i = -1, \mathrm{h}_t(\mathrm{x}_i) = 1 \\ \frac{\mathrm{TP}_t}{\mathrm{FP}_t + \mathrm{TP}_t}\mathrm{C}_{11,}\ if\ \mathrm{y}_i = 1, \mathrm{h}_t(\mathrm{x}_i) = 1 \\ \frac{\mathrm{TN}_t}{\mathrm{FN}_t + \mathrm{TN}_t}\mathrm{C}_{00,}\ if\ \mathrm{y}_i = -1, \mathrm{h}_t(\mathrm{x}_i) = -1 \end{array}\right\}$$

where $\mathrm{TP}_t, \mathrm{TN}_t, \mathrm{FP}_t, \mathrm{FN}_t$ are true positive, negative false positive and falsenegative values for iteration $t.$ *and* negative values for iteration $t.$ *and* $\mathrm{C}_{10,}\mathrm{C}_{01,}\mathrm{C}_{11,}\mathrm{C}_{00,}$are additional values , where$\mathrm{C}_{10} > \mathrm{C}_{00,}\mathrm{C}_{01} > \mathrm{C}_{11,}$

Step 7: The final classification alue is obtained as follows: $\mathrm{H}(\mathrm{x}) = \mathrm{sign}\left(\sum_{t=1}^{T} \alpha_{t\,\mathrm{h}_t}(\mathrm{x})\right)$

Step 8: Output: the final robust value of H.

Step 9: Stop.

This algorithm appears to describe a variant of AdaBoost called Cost-Sensitive AdaBoost, where misclassification costs are taken into account. AdaBoost is particularly useful when dealing with imbalanced datasets or when the cost of misclassifying certain classes is higher than others

Based on the Kidney Disease: Improving Global Outcomes (KDIGO) classification, the categories for estimated glomerular filtration rate (eGFR) are as follows and Categories for estimated glomerular filtration rate (eGFR) based on the Kidney Disease: Improving Global Outcomes (KDIGO) classification. Table-1defines how the categories are defined:

Categories for estimated glomerular filtration rate (eGFR) based on the Kidney Disease: Improving Global Outcomes (KDIGO) classification. Here's how the categories are defined: • G1:Description: Normal or high eGFR Range: > 90 mL/min/1.73m² • G2:Description: Mildly decreased eGFR Range: 60 - 89 mL/min/1.73m² • G3a:Description: Mildly to moderately decreased eGFR Range: 45 - 59 mL/min/1.73m² • G3b:Description: Moderately to severely decreased eGFR Range: 30 - 44 mL/min/1.73m² • G4:Description: Severely decreased eGFR Range: 15 - 29 mL/min/1.73m² • G5:Description: Kidney failure	Categories for persistent albuminuria based on urine albumin-to-creatinine ratio (ACR) ranges. Here's how the categories are defined: ❖ Normal: • Male: < 2.5 mg/mmol • Female: < 3.5 mg/mmol ❖ Microalbuminuria: • Male: 2.5 - 25 mg/mmol • Female: 3.5 - 35 mg/mmol ❖ Macroalbuminuria: • Male: > 25 mg/mmol • Female: > 35 mg/mmo

eGFR Range: < 15 mL/min/1.73m	

Table-1 Relationship between eGFR and CKD

These classifications help in diagnosing and categorizing kidney function and albuminuria levels, aiding in the management and treatment of kidney-related conditions

- Table-1 Shows the Relation Between eGFR and CKD:
- Table-2 Shows the Description of the Features:
- Table 3: Shows the Feature Analysis:
- Fig.2. Statistical Data Analysis of Age:
- Fig.3 Statistical Data Analysis of Blood Pressure:
- Fig4. Statistical Data Analysis of Blood Glucose Random:
- Fig5. Statistical Data Analysis of Blood Urea:
- Fig.6. Statistical Data Analysis of Haemoglobin

Following table-2 shows the Description of the Features

Sl. No	Name of the feature	Description of the feature	Data type of the feature	Scale of the feature
F1	age	Age of the patient	Numerical	l age in years
F2	bp	Blood pressure	Numerical	mm/Hg
F3	sg	Specific gravity	Nominal	1.005, 1.010, 1.015, 1.020, 1.025
F4	al	Albumin	Nominal	0, 1, 2, 3, 4, 5
F5	su	Sugar	Nominal	0, 1, 2, 3, 4, 5
F6	rbc	Red blood cells	Nominal normal	normal, abnormal
F7	pc	Pus cell	Nominal	normal, abnormal
F8	pcc	Pus cell clumps	Nominal	present, not present
F9	ba	Bacteria	Nominal	present, not present
F10	bgr	Blood glucose random	Numerical	mgs/dl
F11	bu	Blood urea	Numerical	mgs/dl
F12	sc	Serum creatinine	Numerical	mgs/dl
F13	sod	Sodium	Numerical	mEq/L
F14	pot	Potassium	Numerical	mEq/L
F15	hemo	Hemoglobin	Numerical	gms
F16	pcv	Packed cell volume	Numerical	% -a percentage
F17	wc	White blood cell count	Numerical	cells/cumm
F18	rc	Red blood cell count	Numerical	millions/cmm
F19	htn	Hypertension	Nominal	yes, no
F20	dm	Diabetes mellitus	Nominal	yes, no
F21	cad	Coronary artery disease	Nominal	yes, no
F22	appet	Appetite	Nominal	good, poor
F23	pe	Pedal edema	Nominal	yes, no
F24	ane	Anemia	Nominal	yes, no
F25	class	class	Nominal	Yes ,no

Features in Table-2 are briefly describes the role of chemical patterns as follows:

- **F1-age-** Age of the patient (age)-in years
- **F2-blood presure(bp)-** Blood pressure is the force of blood pushing against the walls of arteries. Each time the heart beats pumps blood into the arteries. A normal **blood pressure** level is less than 120/80 mmHg. **-** (millimetres of mercury).
- **F3-Specific gravity(sg)-** estimates particles concentration in the urine and the urine's density relative to density of water. It indicates the hydration status of a patient together with the functional ability of the patient's kidney. Specific gravity (sg) **has no units**. That's because it is a ratio of two densities, so the units divide out, making sg a unit less number.
- **F4-Albumin(al)-** is a protein found in the blood. When the kidney is damaged, it allows albumin into the urine. Higher albumin levels in the urine could indicate the presence of CKD. The normal range is 3.5 to 5.5 g/dl(grams per deciliter) or 35-55 g/liter (grams per liter)
- **F5-Sugar(su)- sucrose**, which is a disaccharide made of two mono saccharides: **blood urea** indicates vital information about the functionality of the kidney. this test measures the urea nitrogen quantity in a patient's blood, and a high amount implies the kidneys are not functioning normally. While a random blood glucose test measures the amount of sugar circulating in a patient's blood, and a level of 200 mg/dL or above implies the patient has diabetes.
- **F6- Red blood cells (rbc)-** Red blood cells (RBC) in humans deliver oxygen to the body tissues. The average RBC count is 4.7 to 6.1 mcpm for men and 4.2 to 5.4 mcpm for women. A low RBC, also called anemia, is a common complication of CKD.
- **F7-Pus cells(pc)-** Pus cells in urine are **an indication of urinary tract infection** and in severe cases may indicate sepsis or any other health condition.
- **F8-Pus cell clumps(pcc**)- Pus cells are a collection of dead, white blood cells that accumulates when the body's immune system activates in response to an infection
- **F9-Bacteria(ba)-** A kidney infection (pyelonephritis) is a painful and unpleasant illness caused by bacteria travelling from bladder into one or both of kidneys
- **F10-Blood glucose random(bgr)-** A random glucose test measures the amount of glucose circulating in a person's blood. perform this test to determine whether a person suffered from diabetes.

F11-Blood urea(bu)- blood urea test measures the amount of urea nitrogen found in blood. Urea nitrogen is a waste product made when liver breaks down protein. It's carried in blood, filtered out by kidneys, and removed from body in urine- mgs/dl-milligrams per deciliter

- **F12- Serum creatinine** is a waste product produced by a person's muscles. The serum test is used to checking health of kidney in people at high risk
- **F13-sodium(sod)-** sodium is an electrolyte in the blood that helps the muscles and nerves work effectively. sodium amount is measured by patient's blood, and a very high or low amount may indicate a kidney problem, dehydration, or other medical condition.
- **F14- potassium(pot)-** Potassium is another electrolyte in the blood, and a very high or low amount could signal the presence of an underlying condition. White blood cells (WBC) protect the human body from invading pathogens. They are part of the body's immune system, protecting it from infections.The normal range is between 4000 and 11,000 per microliter of blood. Elevated WBC count is a popular indicator of the progression of CKD.
- **F15-hemoglobin (hemo)-** the part of rbc that carries oxygen throughout body
- **F16-packed cell volume (pcv)-** The packed cell volume (PCV) is a measurement of blood proportion. The value is expressed as a percentage.
- **F17-white blood cell(wc)-**Elevated white blood cell (WBC) count is a well-known predictor of kidney disease in chronic (CKD) progression
- **F18-Red blood cell (rc)-** When the kidneys are damaged, they produce less erythropoietin (EPO), a hormone that signals the bone marrow—a spongy tissue inside in most of bones—to make red blood cells. With less EPO, the body makes fewer red blood cells, and less oxygen is delivered to all organs and tissues
- **F19-Hypertension(htn)- Hypertension** is one of the leading causes of CKD due to the deleterious effects that increased BP has on kidney vasculature
- **F20- Diabetes mellitus (dm**)- Diabetes mellitus is a disorder in which the body does not produce enough or respond normally to insulin
- **F21- coronary artery disease (cad**)- is a narrowing or blockage of coronary arteries
- **F22- apitite(appet**)--a desire for food or drink
- **F23- A pulmonary embolism (PE**)- is a sudden blockage in a lung artery. It usually happens when a blood clot breaks loose and travels through the bloodstream to the lungs
- **F24- Acute necrotizing encephalopathy –(**ane)-**Acute Necrotizing Encephalopathy of Childhood** (ANEC) is a disease characterized by respiratory or gastrointestinal infection.
- **F25-class-class-present or not.**

Meanwhile, the data needs to be pre-processed to make it suitable for machine learning. Therefore,

- The attributes whose scales are ‘normal’ and ‘abnormal’ were transformed to 1 and 0, respectively.
- The attributes whose scales are ‘present’ and ‘not present’ were transformed to 1 and 0, respectively.
- Also, the ‘yes’ and ‘no’ scales were defined as 1 and 0, respectively.
- Lastly, the attribute with ‘good’ and ‘poor’ are also transformed to 1 and 0, respectively.

VI RESULTS AND DISCUSSION

The outcome of this study would be the identification of the important predictors that strongly correlate with eGFR. The complicated patterns and non-linear connections between these characteristics and kidney function can be found with the use of machine learning methods. The models can quantify each feature’s relative relevance and offer information on how it influences changes in eGFR levels. It’s crucial to keep in mind that the precise outcomes of using machine learning to identify factors affecting kidney function using the eGFR formula will depend on the dataset utilized, the quality and completeness of the data, the selection of machine learning algorithms, and the precise analysis and interpretation conducted. Therefore, the findings should be understood in combination with clinical knowledge and validated by more research.

For instance, the results of the research can show that age and serum creatinine level are the two most crucial determinants of eGFR. The model may discover that lower eGFR values are closely related to higher blood creatinine levels, indicating compromised kidney function. Similar to how eGFR declines with age, older age may also be found to be a major predictor of decreasing eGFR.The variables impacting kidney function may be investigated and predicted using machine learning approaches, which can provide important insights into the intricate connections and patterns in the data. A deeper knowledge of renal health and possible risk factors for kidney dysfunction may be gained by using machine learning algorithms to discover and analyse numerous aspects that affect kidney function. Following are some crucial topics for discussion: Identification of Key Predictors, Risk Assessment, Early Detection, and Interpretability of Features analysis of several factors, Non- Linear Relations, and Complex Patterns. The use of machine learning in identifying predictors of eGFR and understanding kidney function offers a promising avenue for research and clinical practice. However, it's essential to interpret the findings in conjunction with clinical knowledge, validate results through additional research, and consider the limitations and potential biases inherent in the dataset and analytical methods used.

	mean	sd	min	max	25%	50%	75%
Age	51.4833	16.9749	2	90	42	54	64
bp	76.4671	13.756	50	180	70	78	80
bgr	148.03	76.583	22	490	101	126	150
bu	57.426	49.987	1.5	391	27	44	61.75
hemo	12.526	2.815	3.1	17.8	10.87	12.53	14.625
pcv	38.884	8.762	9	54	34	38.77	44
pot	4.627	2.92	2.5	47	4	4.63	4.8
sod	137.52	9.908	4.5	163	135	137.53	141
rc	4.707	0.89	8	2.1	4.69	5.1	8
wc	8406.12	2823.3	2200	26400	6975	8377.63	9400

Table 3: Feature Analysis and Result Analysis

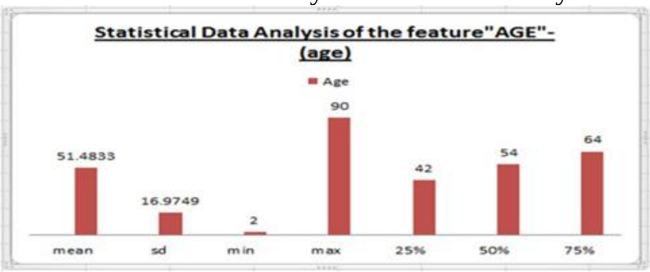

Fig.2. Statistical Data Analysis of Age

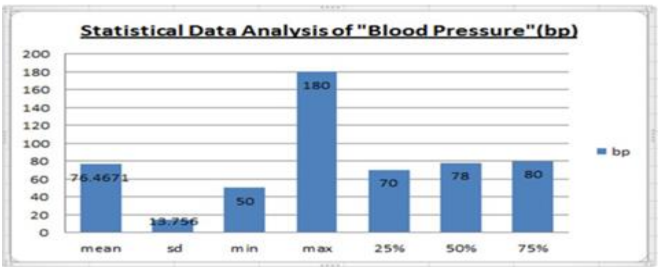

Fig.3Statistical Data Analysis of Blood Pressure

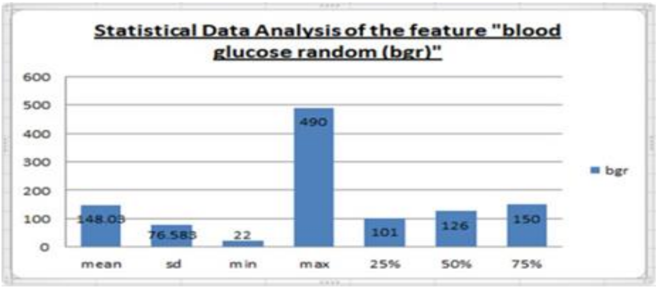

Fig.4. Statistical Data Analysis of Blood Glucose Random

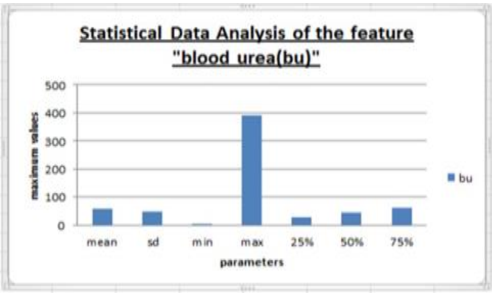

Fig.5. Statistical Data Analysis of Blood Urea

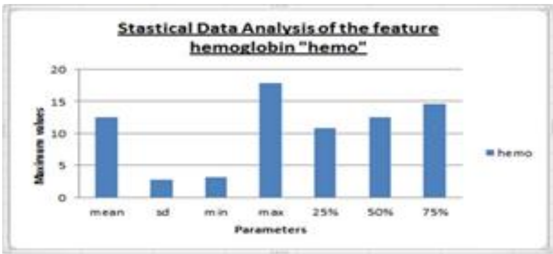

Fig.6. Statistical Data Analysis of Hemoglobin

VII CONCLUSION

In conclusion, machine learning algorithms hold tremendous promise in the field of nephrology for predicting and understanding kidney function. These algorithms can effectively analyze large and diverse datasets containing clinical parameters such as laboratory results, demographic information, medical history, and imaging data. By doing so, they can uncover complex patterns and non-linear relationships that may influence kidney function.

Through the application of machine learning, healthcare providers can develop predictive models that estimate estimated glomerular filtration rate (eGFR) more accurately. These models enable early detection of declining kidney function, allowing for timely intervention and personalized treatment strategies. Additionally, machine learning algorithms can provide insights into the factors that have a significant impact on kidney function, aiding in risk assessment and disease management. Overall, the integration of machine learning into nephrology has the potential to revolutionize patient care by improving diagnostic accuracy, treatment effectiveness, and ultimately, patient outcomes. Continued research and development in this area are essential to harness the full potential of machine learning in kidney health management.

REFERENCES:

[1] M. H. A. Elhebir and A. Abraham, proposed “A novel ensemble approach to enhance the performance of web server logs classification,”SoCPaR (2015)

[2]. P. Kathuria and B. Wedro, “Chronic kidney disease quick overview” IOP Publishing emedicinehealth,(2016). [3]. S. Saremi, S. Mirjalili, and A. Lewis, “Grasshopper optimisation algorithm: theory and application,” Adv. Eng. Softw., vol. 105, pp. 30–47,(2017

[4]. G. A. Afzali and S. Mohammadi, “Privacy preserving big data mining: association rule hiding using fuzzy logic approach,” IET Inf. Secur., vol. 12, no. 1, pp. 15–24, (2018)

[5]. C. Xiong, M. Su, Z. Jiang, and W. Jiang, “Prediction of hemodialysis timing based on LVW feature selection and ensemble learning,” J. Med. Syst., vol. 43, no. 1, pp. 1–8, (2019)

[6]. S. Ravizza et al., “Predicting the early risk of chronic kidney disease in patients with diabetes using real-world data,” Nat. Med., vol. 25, no. 1, pp. 57–59, (2019)

[7]. E.H. A. Rady and A. S. Anwar, “Prediction of kidney disease stages using data mining algorithms,” Informatics Med. Unlocked, vol. 15, p. 100178, (2019) [8]. L. Jerlin Rubini and E. Perumal, “Efficient classification of chronic kidney disease by using multi- kernel support vector machine and fruit fly optimization algorithm,” Int. J. Imaging Syst. Technol., vol. 30, no. 3, pp. 660–673, (2020)

[9]. B. Khan, “Empirical Evaluation of ML Techniques for CKD Prophecy,” IEEE Access, vol.8, March (2020), DOI:10.1109/ACCESS.2020.2981689

[10]. Sezgin, G. et al. (2020) “Clinical thresholds for diagnosing iron deficiency: Comparison of functional

assessment of serum ferritin to population-based centiles". Sci. Rep. Vol.10,no.1,pp-1–12-

[11]. Cardozo, G. et al.(2020) "Prediction of glycated hemoglobin based on routine blood count tests to support the diagnosis of diabetes mellitus". https://doi.org/10.21203/rs.3.rs-38916/v1

[12]. Krishnamurthy S., KS K., Dovgan E., et al. (2021). "Machine learning prediction models for chronic kidney disease using national health insurance claim data in Taiwan".Healthcare.Vol.9, no.5, pp.546, DOI: https://doi.org/10.3390/healthcare9050546

[13].Ifraz, G. M., Rashid, M. H., Tazin, T., Bourouis, S. Khan, M. M. (2021) "Comparative analysis for prediction of kidney diseaseusing intelligent machine learning methods". Comput. Math. Methods Med. https://doi.org/10.1155/2021/6141470.-

[14].Senan EM, Al-Adhaileh MH, Alsaade FW, Ald- hyani THH, Alqarni AA, Alsharif N, Uddin MI, Alahmadi AH, Jadhav ME, Alzahrani MY.J Healthc Eng. (2021) "Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques".DOI: 10.1155/2021/1004767.

[15].Qiong Bai, Chunyan Su, Wen Tang Yike Li. (2022) "Machine learning to predict end-stage kidney disease in chronic kidney disease".Vol.12, pp.1-8

[16] P. Chittora et al, "Prediction of Chronic Kidney Disease - A Machine Learning Perspective," *IEEE Access*, vol.9 January(2021) DOI 10.1109/ACCESS.2021.3053763

[17] Schena FP, Anelli VW, Abbrescia DI, Di Noia T.J Nephrol. (2022) "Prediction of chronic kidney disease and its progression by artificial intelligence algorithms".Vol.35, no.8, DOI: 10.1007/s40620- 022- 01302-3.

[18]. Zeng ZQ, Yang SC, Yu CQ, Zhang LX, Lyu J, Li LM.Zhonghua Liu Xing Bing Xue Za Zhi. (2023). "Progress in research of risk prediction model for chronic kidney disease".Vol.44, no.3, pp-49.

[19] Pankaj:Chittora, Sandeep;Chaurasia, PrasunChakraba rti, GauravKumawat; Tulika hakrabarti,(2021),Prediction of Chronic Kidney Disease - A Machine Learning Perspective, *IEEE Access (Volume: 9),Page(s): 17312 – 17334*,DOI: 10.1109/ACCESS.2021.3053763

[20] Poonia RC, et al. Intelligent Diagnostic Prediction and Classification Models for Detection of Kidney Disease. *Healthcare*. 2022;10:2.

[21] Yashfi SY. Risk Prediction Of Chronic Kidney Disease Using Machine Learning Algorithms. 2020.

[22] Molla MD, et al. Assessment of serum electrolytes and kidney function test for screening of chronic kidney disease among Ethiopian Public Health Institute staff members, Addis Ababa, Ethiopia. *BMC Nephrol.* 2020;21(1):494.

[23] Misra P, Yadav AS. Improving the classification accuracy using recursive feature elimination with cross-validation. *Int J Emerg Technol*. 2020;11(3):659–65.