Section: Research Paper
ISSN 2063-5346

Ensemble Machine Learning Models for Accurate Prediction of Kidney Diseases

Dillip Narayan Sahu[1], Satheesh Kumar [2*]

[1]Research Scholar, Department of Computer Science and Application, OPJS University, Churu, Rajasthan, India

[2*] Professor, Department of Computer Science and Engineering, Chaitanya (Deemed to be University), Telangana, India.

*Corresponding Author Email: satish4info@gmail.com

DOI: 10.48047/ecb/2023.12.si4.1692

Abstract

The study focuses on renal problems, which have emerged as a significant public health concern, and the significance of prompt management and better patient outcomes. The study suggests using machine learning techniques to anticipate kidney diseases in order to address this problem. The study's dataset includes a variety of clinical, demographic, and laboratory data from different patient populations. Utilizing feature engineering and preprocessing procedures, the input data's quality and relevance are increased. A group of machine learning models, including Random Forest, Gradient Boosting, and Support Vector Machines, are subsequently developed. Utilizing each particular technique's benefits while minimizing its drawbacks is the aim. The findings show that the ANFIS model achieves a precision of 100% compared to the NB-CbH model's 90.2%. This suggests favorable outcomes in the early detection of chronic kidney disease (CKD). The suggested model may help nephrologists make wise choices when necessary, potentially enhancing patient quality of life on a global scale. To further increase the predictor model's accuracy and classification rate, the study contends that deep learning techniques can be used in the future. The incorporation of deep learning into this predictor model should result in even more precise and reliable predictions for renal illnesses. Deep learning has demonstrated tremendous potential in a variety of medical applications.

Keywords: Ensemble learning, Machine learning, Prediction, Kidney diseases, Chronic kidney disease (CKD) and Diagnosis

1. Introduction

A rapidly spreading non-communicable illness, chronic kidney disease (CKD) has a large global mortality and health impact. A 2019 data analysis estimates that 755 million individuals worldwide, including 418 million women and 337 million men, are affected by CKD. With 17.5% of the world's population, India has a significant CKD burden[1]. The illness develops when the kidneys gradually are unable to carry out their crucial blood filtration task. "Chronic" illness is the word used to describe this continuing kidney impairment. Waste buildup in the blood causes a number of health issues as the urinary organs are harmed or compromised[2]. Numerous signs of CKD include hypertension, bone thinning, decreased nerve function, abnormal blood counts, and an elevated risk of heart and blood vessel problems. Early diagnosis of CKD and the detection of risk factors can help stop the evolution of the condition and reduce its negative effects on people's health. Glomerular Filtration Rate (GFR), which assesses renal functionality for the patient, is one of the important metrics used to gauge the severity of CKD[3]. The GFR value, which depends on elements including age, blood count, gender, race, and patient-specific characteristics, declines as the kidney disease gets worse. Based on the GFR measurement, CKD is divided into five stages that can be effectively analyzed utilizing machine learning techniques. Based on the GFR values acquired, these algorithms can offer clinicians useful information that can help them make better judgments[4]. Along with machine learning algorithms, clinical decision support systems (CDSS) can help doctors in emergency scenarios involving the management of CKD. Figure.1 depicts the course of CKD, showing the stages and effects on renal function over time.

Figure.1: CKD progression

The chief causes of chronic kidney disease (CKD), which frequently results from the loss of kidney function, include hypertension, diabetes, and cardiovascular disease (CVD). Hypertension is a significant cause of CKD cases in metropolitan environments. Other important causes of CKD include obesity, diabetes, aging, and CVD. Diabetes and high blood pressure are two of the disease's main causes[5]. Renal function is commonly evaluated in clinical medicine via tests of serum creatinine levels. The Glomerular Filtration Rate (GFR) is assessed by periodic measurements of creatinine drawn from urine samples. Serum creatinine can be influenced by a number of variables, including body weight, sex, age, nutrition, medications, and muscle mass, hence it has limits as a GFR predictor. Equations like the "Cockroft-Gault (CG) equation" and the "MDRD equation" are used to calculate GFR, which is regarded as the "Gold Standard" for assessing kidney function[6]. Given that it is simple to use and requires inputs for sex, race, weight, and age to calculate GFR, some practitioners choose the MDRD equation[7]. The MDRD equation, however, takes a long time and has large inaccuracies. Clinical Decision Support Systems (CDSS), which manage enormous volumes of patient data and data analytics, are essential for controlling CKD. The inclusion of machine learning algorithms in CDSS can assist in addressing these issues and improve forecast accuracy[8]. By combining CDSS and machine learning, healthcare providers can make decisions that are more effective and more successfully manage CKD. Figure.2 displays the general block diagram of a CDSS, showcasing its numerous elements and functionalities in aiding clinical decision-making[9]. Healthcare professionals can enhance the prediction and management of CKD by utilizing the capabilities of machine learning and CDSS, improving patient outcomes and fostering more efficient healthcare practices.

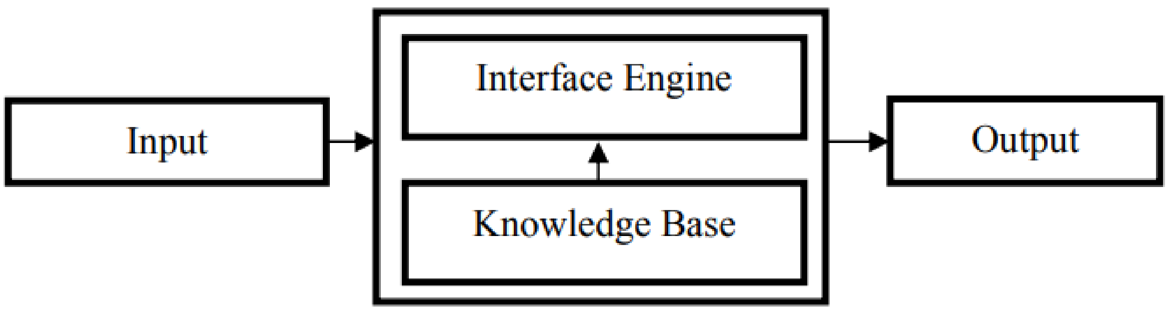

Figure.2: General Block Diagram of CDSS

Indeed, employing Clinical Decision Support Systems (CDSS) has a number of potential advantages that could greatly advance medical procedures: Enhancing Test Ordering and

Patient Safety: CDSS can assist healthcare professionals in making more precise and research-based judgments regarding the ordering of tests and available treatments[10]. The CDSS can recommend suitable tests and treatments by examining patient data and medical literature, which lowers the possibility of needless or inaccurate testing and enhances patient safety. Improved Service Quality: CDSS gives healthcare personnel access to the most recent clinical data, recommendations, and best practices. By doing this, they may be sure that they are treating patients with the most up-to-date information and knowledge[11]. The availability of pertinent information and suggestions enables doctors to make quick decisions that improve patient outcomes and patient satisfaction. Reducing Unnecessary Laboratory Investigations: By recommending the best tests in light of the patient's symptoms and medical background, CDSS can help to streamline the diagnosis procedure[12]. Healthcare companies can reduce expenses while enhancing the effectiveness of patient treatment by minimizing pointless or redundant laboratory investigations. Facilitating Data-Driven Decision-Making: To examine huge amounts of patient data, CDSS uses data analytics and machine learning algorithms[13]. This makes it possible for medical providers to make data-driven choices, which results in more precise diagnoses and individualized treatment strategies. Supporting Clinical Workflow: CDSS smoothly integrates into the clinical workflow, giving medical practitioners real-time assistance while they consult with patients and make decisions[14]. By ensuring that doctors have access to vital information at the moment of service, improved clinical results are produced. Promoting evidence-based medicine: CDSS bases its recommendations on evidence-based standards, protocols, and best practices. By adhering to evidence-based medicine, medical professionals may make sure that their choices are in line with the most recent, verified medical information, improving the standard of care given to patients. Overall, the deployment of CDSS in healthcare settings can result in more effective and efficient decision-making, greater patient safety, improved service quality, and cost savings through resource usage that is optimized[15]. As technology develops, CDSS is anticipated to become more and more important in the healthcare industry, altering how medical choices are made and improving patient care.

2. Literature Survey

In order to anticipate Chronic Kidney Disease (CKD) at an earlier stage, researchers have conducted a number of studies and analyses. This section provides a thorough summary of those studies and analyses. A important global health problem, CKD affects a sizeable section of the population and is becoming more common. For improved care and the reduction of complications, early CKD prediction is essential[16]. To predict CKD, a number of machine learning algorithms and methodologies have been investigated. While some studies employed clinical information and laboratory tests, several investigations concentrated on employing imaging techniques. Glomerular filtration rate (GFR), a crucial indicator of kidney health, is used to categorize CKD into various stages. To calculate GFR and forecast CKD phases, machine learning algorithms like the CKD-EPI and Cockroft-Gault equations have been applied. It is highlighted the advantages of employing clinical decision support systems (CDSS) to predict CKD, including better test ordering, patient safety, service quality, and cost savings. CDSS facilitates the diagnosis process and helps healthcare providers make decisions based on the best available evidence[17]. The part also covers the difficulties and complications of using CDSS and machine learning to predict CKD. To ensure the accuracy of predictions, preprocessing and handling missing data are crucial stages. The study also highlights the significance of screening and early detection while acknowledging the rising prevalence of CKD in India, particularly as a result of diabetes and hypertension. However, there are still certain issues and difficulties to be resolved despite the positive results of machine learning models and CDSS in CKD prediction[18]. More precise and well-modeled predictor designs are required, especially in randomized controlled trials, according to the section. The literature evaluation also enables researchers to propose additional data analysis and implementation as well as identify research gaps. Overall, the part offers insightful information about how CKD is now predicted using machine learning and CDSS, and it emphasizes the value of early detection and preventive steps in treating this serious health issue. Additionally, it identifies areas for additional study and advancements in CKD predictive model development. In order to anticipate Chronic Kidney Disease (CKD) at an earlier stage, researchers have conducted a number of studies and analyses[19]. This section provides a thorough summary of those studies and analyses. A important global health problem, CKD affects a sizeable section of the population and is becoming more common. For improved care and the reduction of complications, early CKD prediction is essential. To predict CKD, a number of machine learning algorithms and methodologies have been investigated[20]. While some studies employed clinical information and laboratory tests, several investigations concentrated on employing imaging techniques. Glomerular filtration rate (GFR), a crucial indicator of kidney health, is used to categorize CKD into various stages. To calculate GFR and forecast CKD phases, machine learning algorithms like the CKD-EPI and Cockroft-Gault equations have been applied[21]. It is highlighted the advantages of employing clinical decision support systems (CDSS) to predict CKD, including better test ordering, patient safety, service quality, and cost savings. CDSS facilitates the diagnosis process and helps healthcare providers make decisions based on the best available evidence. The part also covers the difficulties and complications of using CDSS and machine learning to predict CKD. To ensure the accuracy of predictions, preprocessing and handling missing data are crucial stages. The study also highlights the significance of screening and early detection while acknowledging the rising prevalence of CKD in India, particularly as a result of diabetes and hypertension. However, there are still certain issues and difficulties to be resolved despite the positive results of machine learning models and CDSS in CKD prediction. More precise and well-modeled predictor designs are required, especially in randomized controlled trials, according to the section[22]. The literature evaluation also enables researchers to propose additional data analysis and implementation as well as identify research gaps. Overall, the part offers insightful information about how CKD is now predicted using machine learning and CDSS, and it emphasizes the value of early detection and preventive steps in treating this serious health issue. Additionally, it identifies areas for additional study and advancements in CKD predictive model development.

3. Machine Learning

A branch of artificial intelligence known as machine learning (ML) is concerned with creating models and methods that enable computers to learn from data without being explicitly programmed. ML has demonstrated considerable promise in the medical industry for a number of applications, including medical prediction, medical quality improvement, and cost reduction. Nephrology-related clinical problems and challenges can be greatly aided by the application of machine learning (ML). Large datasets can be analyzed by ML algorithms, which can also find patterns and forecast kidney health and illness. Healthcare workers can gain from more precise diagnoses, individualized treatment strategies, and early kidney issue identification by implementing ML in nephrology. Depending on the needs of the modeling, many types of ML algorithms are categorized. Unsupervised learning, supervised learning, reinforcement learning, and semi-supervised learning are some of these categories. Finding patterns and relationships in data without preconceived labels is the focus of unsupervised learning. Labeled data are used in supervised learning to build models and make predictions. Through reinforcement learning, agents are taught to base decisions on rewards and penalties.

Unsupervised and supervised components are both included in semi-supervised learning. A thorough understanding of the ML domain is necessary for the creation and implementation of various ML algorithms. To effectively use ML for nephrology applications, researchers and healthcare practitioners must be well-versed in various statistical techniques and methodologies. To achieve accurate and trustworthy outcomes, the procedure includes data pretreatment, method selection, model training, and evaluation. The division of ML and ML techniques are probably depicted in Figures. 3 and 4, respectively, in the section, demonstrating the variety of ML approaches that can be used in the field of nephrology. Overall, using machine learning to nephrology has the potential to greatly enhance patient outcomes, simplify medical procedures, and offer insightful information to healthcare professionals. As the area develops, continuous ML method research and development will be essential for addressing clinical issues and improving the standard of treatment for kidney-related illnesses.

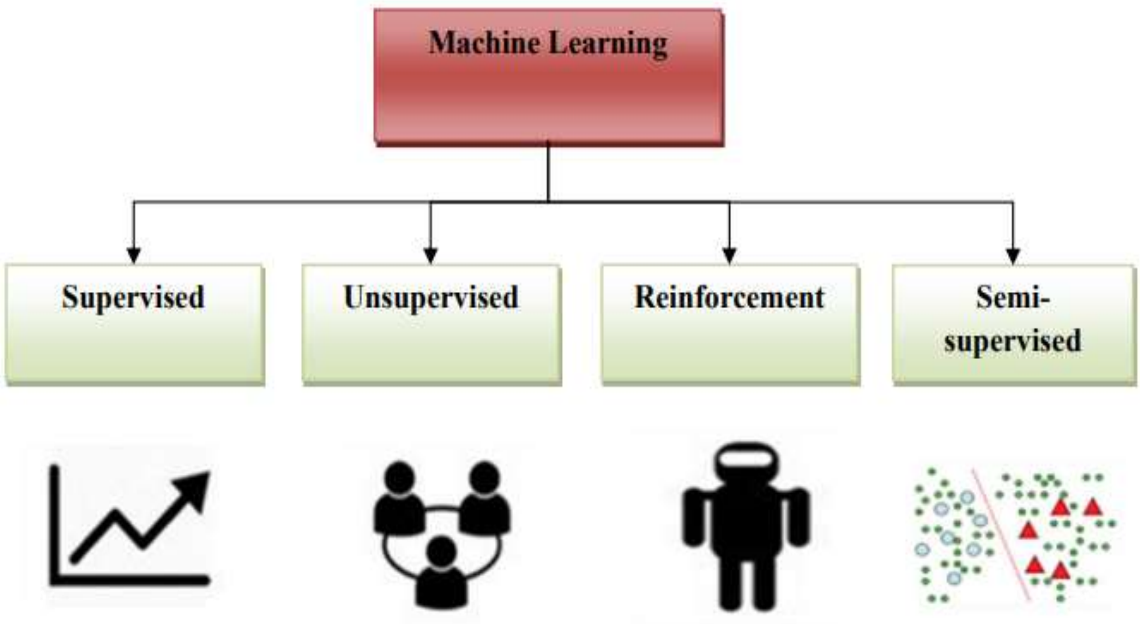

Figure.3: Types of Machine learning approaches

A type of machine learning called supervised learning involves feeding labeled data to the algorithm during the training stage. Each instance in the training data has both the appropriate output values or labels for the corresponding input features (often represented as vectors). The supervised learning algorithm's objective is to discover a mapping between input and output pairs such that it can forecast the right output for brand-new, unforeseen input data. The supervised learning methods Naive Bayes (NB), Random Forest (RF), Logistic

ISSN 2063-5346

Regression (LR), and Support Vector Machines (SVM) are frequently employed in medical research. Numerous medical issues, including disease classification, patient prognosis, and treatment prediction, have been successfully solved using these algorithms. Supervised learning does, however, have some drawbacks. The requirement for enough labeled training data is a fundamental restriction. It can be difficult and time-consuming to obtain high-quality labeled data, particularly in the medical field where expert annotations are necessary. In order to perform at their peak, supervised learning models may contain intricate optimal control parameters that require careful tuning. Unsupervised learning, on the other hand, is a subset of machine learning in which the algorithm is not given access to labeled data throughout the training process. Instead, it looks for structures or patterns in the data without any supervision or assistance. K-means clustering, a well-liked unsupervised learning technique, divides the samples into various clusters based on the traits of the training data. Unsupervised learning can be helpful in the context of medical research for problems when labeled data is scarce or nonexistent. It can aid in the discovery of inherently occurring patterns in histology data or medical imaging that may be vital for forecasting diseased diseases or generating fresh ideas. Unsupervised learning does present certain difficulties, though. It might not be able to capture complicated relationships or perform as effectively as supervised learning models in some situations since it lacks the advantage of labeled data. Furthermore, there is ongoing research to enhance the performance and capacities of unsupervised learning algorithms, and the difference between them and human intelligence is still quite small. In conclusion, both supervised and unsupervised learning have advantages and disadvantages, and the best learning strategy will rely on the particular medical research issue at hand as well as the accessibility of labeled data. Combining the advantages of these two learning approaches can result in medical AI solutions that are more thorough and efficient.

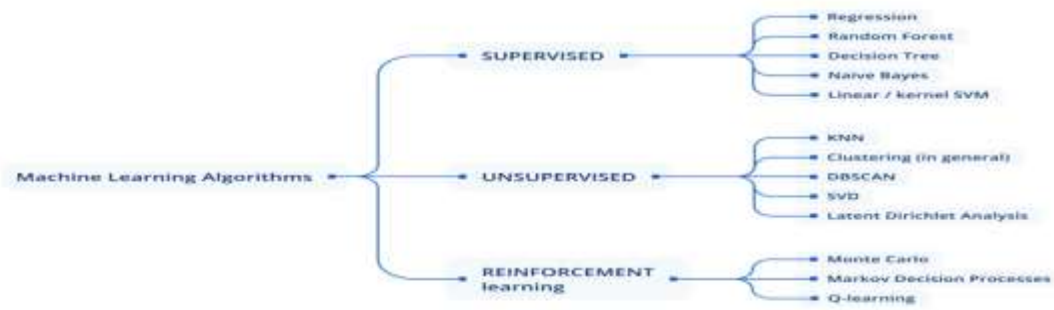

Figure.4: Machine learning techniques

A type of machine learning called reinforcement learning involves an agent interacting with its surroundings to learn how to make decisions that will help it accomplish specific objectives. Depending on the actions the agent does in the environment, the environment will either reward or punish the agent. The agent seeks to maximize the total rewards that it accrues throughout time. The Markov Decision Process (MDP), which encapsulates the uncertainty in the treatment process and the underlying random occurrences, is a typical model used in reinforcement learning. For sequential decision-making tasks like calculating dose for chronic conditions or anticipating dosage sequences, reinforcement learning is ideally suited. Transfer learning is a method for using information learned from one task or dataset to perform better on a related job or dataset, particularly when the latter contains less data. In transfer learning, a model is initialized or fine-tuned for a new task using the parameters and information learned from a pre-trained model. This strategy can shorten the time needed to complete the new work successfully and conserve computational resources. Transfer learning enables the model to build on the knowledge it has acquired from earlier tasks rather than having to train a new model from begin. Transfer learning is especially helpful when there aren't many labeled data points available for the new activity. The model can generalize more effectively to the new task with less data by starting with a pre-trained model that has previously acquired valuable features from a large dataset. This is particularly important in the field of medical research, since it can be costly and time-consuming to acquire labeled medical data. Overall, transfer learning and reinforcement learning are potent approaches that can be used in a variety of medical research contexts. Transfer learning helps in using previously learned skills for greater performance in tasks with limited data, whereas reinforcement learning supports decision-making processes. Both strategies help to advance medical AI research and enhance patient outcomes.

4. Kidney diseases

Electronic Health Records (EHRs) have significantly increased as a result of the introduction of the ECHIT (Electronic Health Care Information Technology) act. EHR data is a useful resource for disease prediction since it contains essential information on the development of diseases. For the purpose of illness prediction, machine learning (ML) techniques are widely used on EHR data, particularly when it comes to renal diseases. ML may be useful in identifying and forecasting the development of renal illnesses and evaluating renal function impairment. It is now possible to more effectively understand a patient's health condition and precisely identify risk factors for certain diseases by analyzing EHR data. Large-scale observational research has become popular and essential in the medical industry thanks to the big data era and improvements in machine learning techniques. Individualized predictions for kidney disorders can be made using machine learning models created from EHR data, which will improve the standard of medical care. With regard to reliably determining estimated Glomerular Filtration Rate (eGFR), a gauge of kidney function, ML has demonstrated encouraging results. The importance of AI in precisely forecasting the course of various phases of Chronic Kidney Disease (CKD) has been underlined in numerous research. For instance, scientists have developed effective models for predicting CKD with renal failure using laboratory data from common EHRs, such as albuminuria and eGFR. The capacity to forecast renal deterioration has been further improved by the integration of temporal information with medical data. To better forecast CKD and track changes in eGFR over time, researchers have investigated a number of cutting-edge ML methods, such as adaptive neuro-fuzzy interference systems and unsupervised learning models like latent Dirichlet allocation. There are still certain issues that need to be resolved despite the advances in creating risk prediction models. To make sure the model effects are reliable and generalizable, more investigation and validation are needed. Before using the models in clinical practice, the results must first be externally validated and calibrated. Figure.5's Model for Diagnosis of Chronic Kidney Disease serves as an illustration of how machine learning can be used to predict and diagnose CKD using EHR data. Improved patient outcomes and higher-quality healthcare are highly anticipated benefits of ongoing research and development in machine learning (ML) applications for illness prediction in nephrology.

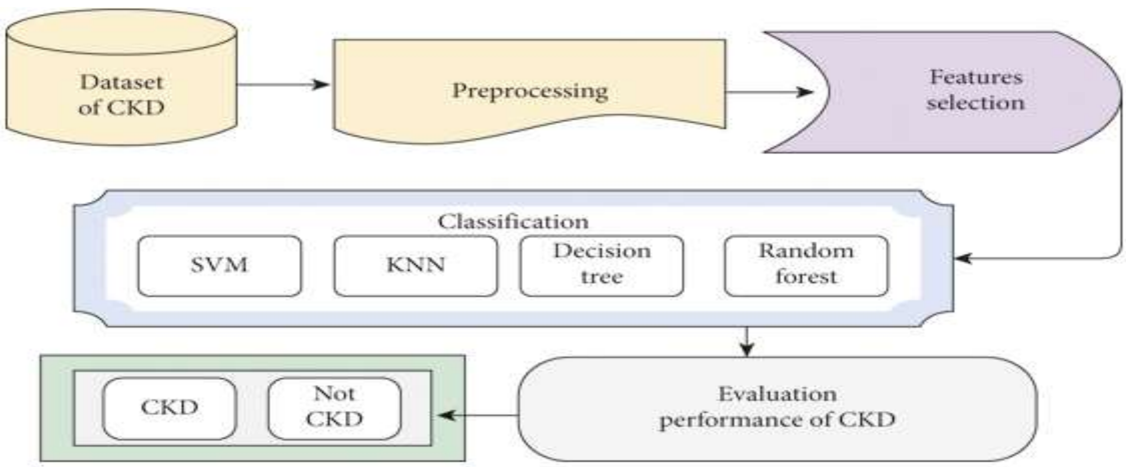

Figure.5: Model for Diagnosis of Chronic Kidney Disease

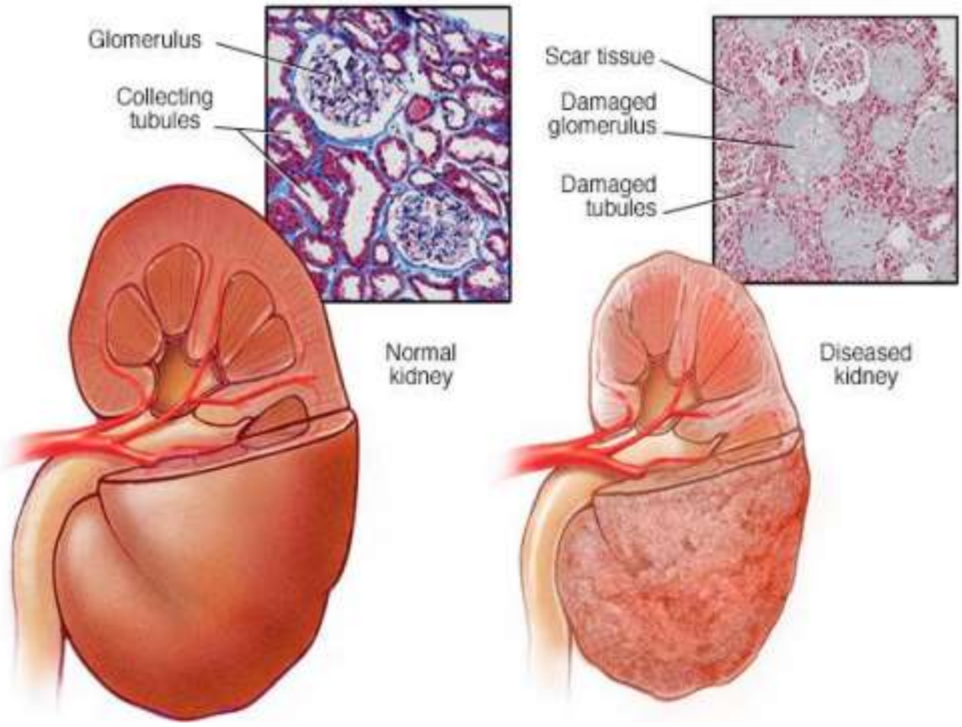

Figure.6:Generic view of Normal and diseased kidney

The paper's description of the framework of CKD diagnosis is illustrated in Figure 5. When addressing missing values, the preprocessing stage uses the mean approach to derive numerical values and the mode method to generate nominal values. The Recursive Feature Elimination (RFE) technique is used in the feature selection process to find critical features that are pertinent to the diagnosis of CKD. The disease classifiers are then fed with the chosen features as input. In this work, CKD is diagnosed using four classifiers: Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree, and Random Forest. All of these classifiers performed well in reliably classifying the dataset into either CKD or a normal kidney state, with promising outcomes. The mortality rate from acute kidney injury (AKI) is estimated to be between 10% and 12% worldwide. The clinical outcomes of AKI have been improved significantly, with a focus on early prognosis and individualized care. Reduced mortality rates and an improved prognosis for kidney function can result from early AKI evaluation. Numerous models have been created for the purpose of early AKI prediction utilizing real-data targets, which can help nephrologists save time and effort. Clinical research on AKI has benefited greatly from the use of Electronic Health Records (EHR) data in big data research. Future clinical practices should benefit even more from the development of Clinical Decision Support Systems (CDSS) based on self-learning prediction

Section: Research Paper
ISSN 2063-5346

models. These CDSS technologies will aid in the diagnosis and forecasting of AKI, improving patient outcomes. Figure 6 shows a generalized view of the damaged and healthy kidney. AKI and CKD can be diagnosed using this visual representation, which also explains the differences between the two ailments.

5. Research Challenges

To fully realize ML's potential, it is necessary to solve the drawbacks and difficulties associated with using it in conventional clinical practice. The "black box" nature of ML models, which makes it challenging for doctors to comprehend the underlying logic behind the model's predictions, is one of the key obstacles. To ensure the responsible and ethical usage of AI algorithms in healthcare, ML ethics must also be taken into account. Although there exist rules for AI governance in a number of business sectors, the development of precise clinical standards for ML applications is lagging behind the development of new technologies. Standardizing the use of ML in clinical settings requires the establishment of such guidelines. Another difficult aspect is gathering the necessary data. Electronic Medical Records (EMR) from different medical facilities may not be uniform and contain missing or partial data, which lowers the quality of the data and makes it more challenging to extract pertinent information. Due to the complexity and non-shared nature of data across several units, it is essential to ensure data authenticity and integrity, and external model verification is frequently inadequate. Any label inaccuracies can substantially impair the performance of the model because label accuracy is crucial for the accuracy of ML algorithms. It can be difficult to gather reliable data and make it available for modeling. Another important factor is the expense of gathering, storing, and processing data with a lot of computing power. The prediction of glomeruli stage is a key component of the study of renal disease prediction. The sample data size is frequently tiny, and the kidney sections used for modeling are typically diseased. Automatic pathological prediction of specific kidney diseases based on extensive patient data has not yet been fully attained despite the phenomenal rise of AI. Pathological prediction needs significant data and validation across many clinical practices and datasets before pathologists may be completely replaced. Overall, even if ML has promise applications in the medical area, overcoming these obstacles is essential if ML is to be fully incorporated into current clinical practice and reap the benefits of its use in the prevention and diagnosis of renal disorders and other illnesses.

Section: Research Paper
ISSN 2063-5346

6. ML Based Diagnosis System

The research issue and the research tools are identified as the first step in the technique given in this study. The data is subsequently analyzed, which is regarded as the core of the study, using data mining. The information is gathered from an internet data source and prepped for use in machine learning classification algorithms. Before using the algorithms, missing data are addressed, and feature scaling and classification are carried out. On the preprocessed data, a number of machine learning techniques are applied, including Logistic Regression, Decision Tree, Random Forest, and K-Nearest Neighbors. Each algorithm's performance is assessed, and the best-fit algorithm is suggested as the most effective tool for CKD prediction. The major goal of this part is to use learning techniques to develop a prediction framework for CKD. In order to attain an appropriate accuracy measure with hierarchical classification, the study focuses on employing Naive Bayes (NB) and Classification by Hierarchies (CbH). The evaluation metrics give the predictor model improved outcomes and reliability. The classification methods are selected after a thorough examination of the methods currently used in medical applications. With a recommendation to choose hierarchies in order to obtain adequate results with precision, the research seeks to compute the early prediction of CKD, specifically severe renal failure. The characteristics of the learning strategy are based on data gathering and outcomes that have stabilized over the course of model development. Information extraction is the process of acquiring knowledge from a repository, and this study employs information extraction from clinical data to forecast heart conditions. For determining the severity of an illness, feature or information extraction is an essential stage. For disease prediction, a combination of machine learning, medical databases, and statistical analysis is applied. Figure.7 shows the projected Naive Bayes classifier. Overall, this methodology seeks to forecast CKD using machine learning techniques and to offer an accurate and trustworthy prediction model for severe renal failure. In order to get the best outcomes when forecasting the severity of the condition, it makes use of data mining techniques and classification algorithms.

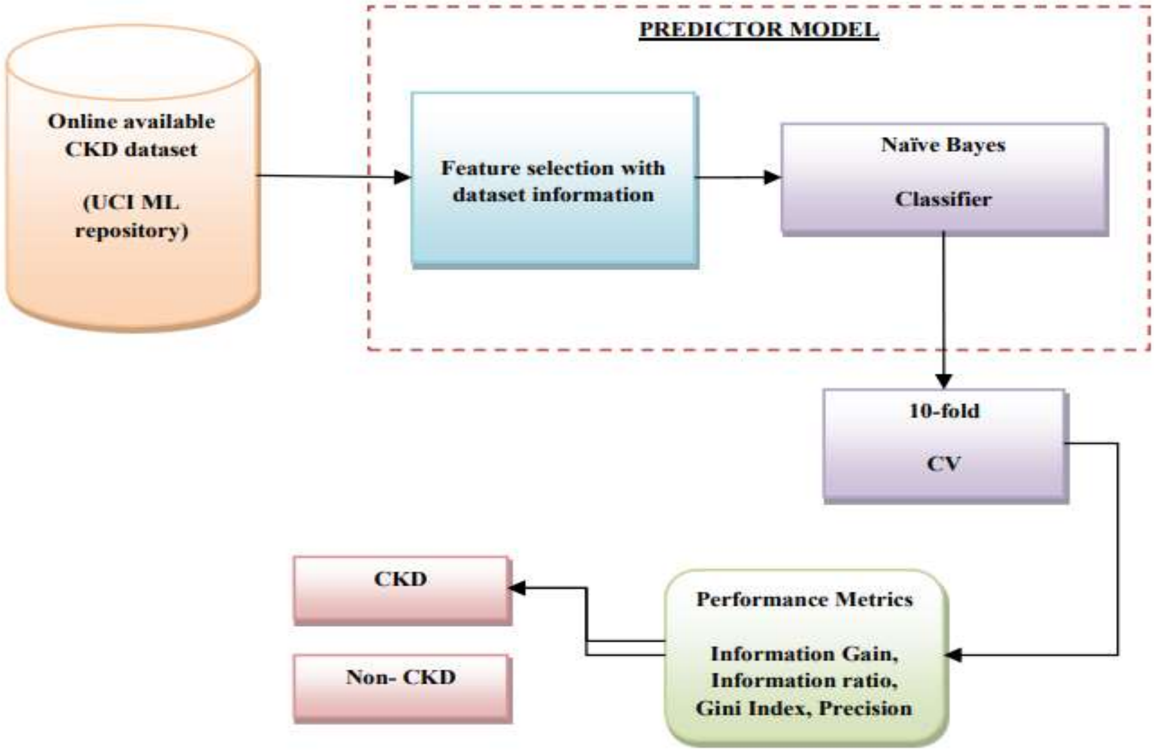

Figure.7: Block diagram of proposed NB predictor model

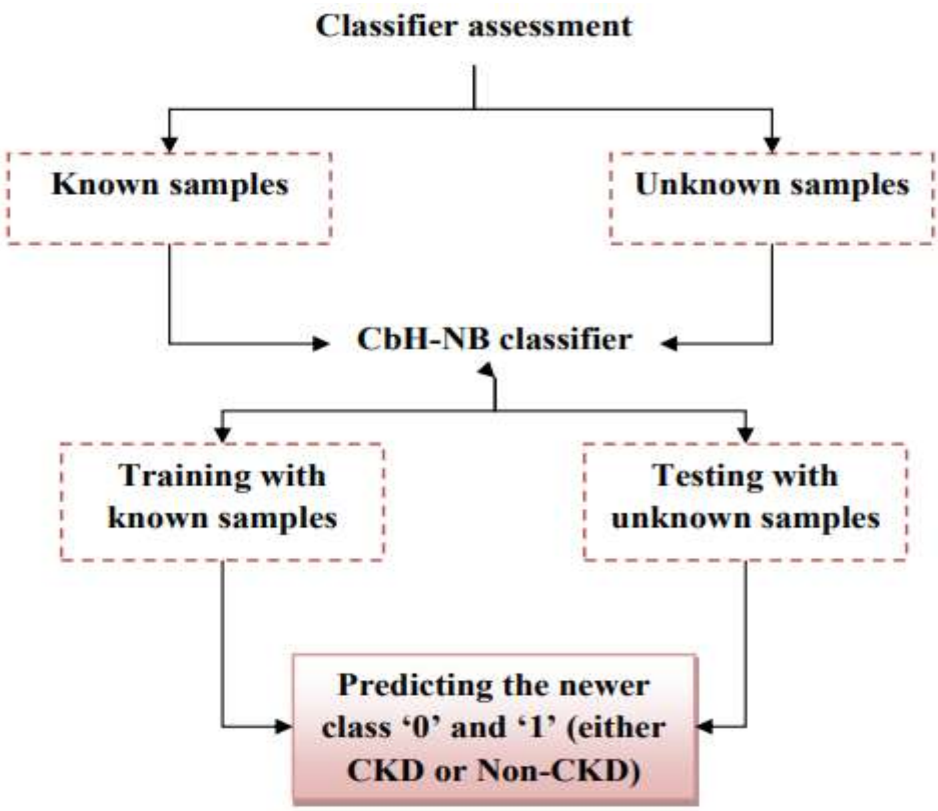

Figure.8: Prediction of newer classes using CbH-NB classifier

In this research, feature extraction strategies are applied in the existing approaches to classification and evaluation. To create classified analytical prototypes, guidelines, and evaluate the classification effectiveness, information extraction is done. Based on classification data, these activities are evaluated. Software like YALE and WEKA, which use Java for a wealth of options in knowledge base, business, scientific, and medical research, are used to extract information from a variety of sectors. For constructing a 10-fold cross-validation of machine learning algorithms to improve performance, faster information extraction is essential. The data is split into a training set and a testing set using a 90:10 ratio as part of the validation procedure. In order to create a set of data for instruction development that helps to overcome the shortcomings of existing models, cross-validation is used. The results are calculated using the supplied data, and the classification is carried out using the proper categorization, separating CKD from non-CKD. The prediction of more recent classes using the CbH-NB classifier is shown in Figure 8. To improve classification outcomes and accuracy in CKD prediction, this classifier combines the strength of Classification by Hierarchies (CbH) and Naive Bayes (NB). The CbH-NB classifier is used to forecast classes that have never been encountered before, making it an effective tool for CKD prediction. Overall, the research focuses on using information extraction and classification methods along with machine learning (ML) methodologies to predict CKD and produce accurate and trustworthy findings. The prediction model performs better and can predict newer classes more accurately when cross-validation and the CbH-NB classifier are included.

The major goal of this study is to examine significant elements of the diabetic dataset and to gather useful information from various medical records. Applying the CbH-NB classifier on the diabetes dataset reveals a precision of 90.2%. The diabetes dataset's classification procedure successfully separates classes and organizes the data into categories for evaluation. The study stresses how crucial it is to extract classes correctly and synchronize them in order to accurately assess the performance of the system. By applying methods like closest neighbors (NN) and clustering, the evaluation procedure gauges the classifier's accuracy. The objective is to improve system performance and predictive accuracy for diabetes situations. The data from the diabetic dataset that was retrieved is used to enhance the entire classification system and guarantee more accurate prediction results. In general, the project attempts to accurately categorize diabetic cases and extract useful insights from medical data using the CbH-NB classifier. Improving the classifier's performance is essential for increasing the system's predictive power and accuracy for diabetes cases. The research intends to provide useful information for diabetic diagnosis and therapy by assessing dominating aspects and applying appropriate classification techniques.

7. Results and Discussion

The major goal of this study is to use machine learning techniques to create an efficient Clinical Decision Support System (CDSS) for Chronic Kidney Disease (CKD). The researchers plan to accomplish this by taking a methodical approach. Data Gathering: The first step in predicting CKD is to compile a benchmark dataset. The CKD dataset was picked by the researchers in this instance from the UCI Machine Learning library. Feature Selection: The researchers concentrate on creating an effective feature selection approach in order to improve the classification accuracy while minimizing the amount of features. This process is essential for choosing the features that are most pertinent to CKD prediction. Model Design: The researchers selected suitable Machine Learning techniques for modeling the CKD predictor for medical applications based on a thorough literature review. The researchers' model of choice in this instance is the Naive Bayes (NB) classifier. Evaluation of Performance: The CKD dataset from the UCI repository is used to assess the performance of the proposed model. On a particular hardware setup, the simulation and testing are carried out using MATLAB R2018a. Analysis and Results: Based on the error/fault rate for all of the available samples in the CKD dataset, the researchers evaluate the performance of the NB classifier. Figures.9 and Figure.10 illustrate the analysis' findings. The development of a trustworthy and accurate CDSS for CKD prediction is the ultimate objective of this study. This tool will help healthcare providers make better decisions and improve patient outcomes. The performance of the NB classifier and the feature selection approach are crucial in reaching this objective and creating a reliable CDSS for CKD.

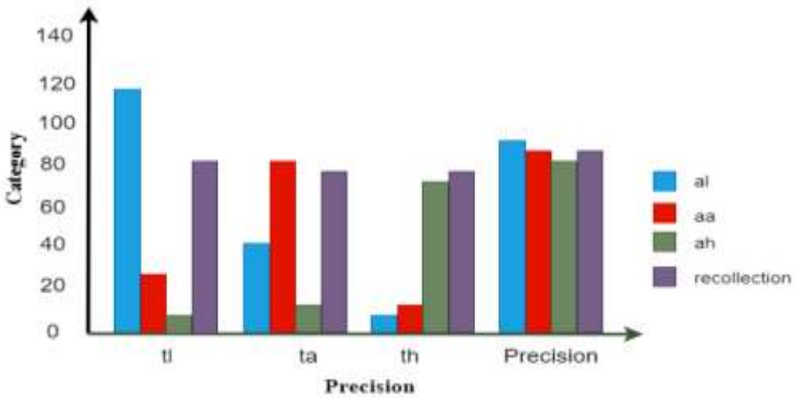

Fig ure.9: NB Scattering Precision

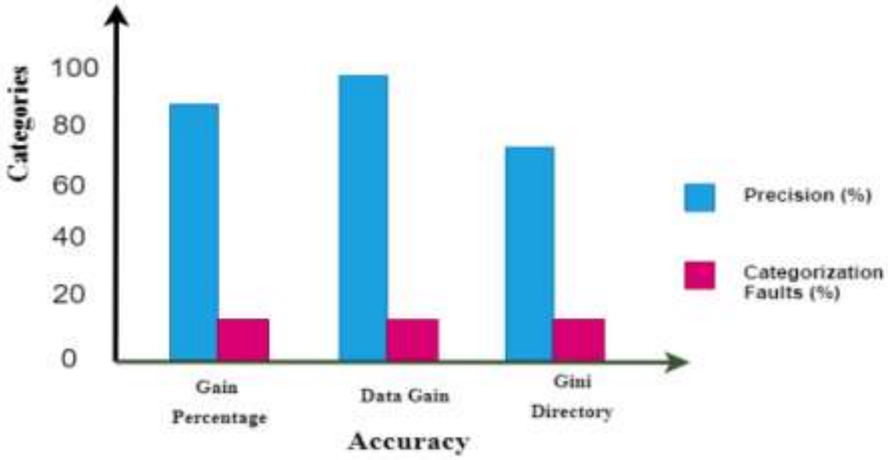

Figure.10: NB Scattering Accuracy

Figure.11 shows a comparison of the performance metrics for various CKD prediction models. The Adaptive Neuro-Fuzzy Inference System (ANFIS) model, the Naive Bayes (NB) model, and the Choice based hierarchy (CbH) model are the models that are being compared. Precision, sensitivity, and specificity are among the performance measures that are assessed. The findings reveal that in all three performance indicators, the ANFIS model performs better than the NB and CbH models. The ANFIS model's precision value is 9.8% and 11.3% greater than those of the NB and CbH models, respectively. This suggests that when compared to the other models, the ANFIS model has a higher accuracy in predicting true positive cases. The ANFIS model's sensitivity value is also 16% greater than that of the NB model and 6% higher than that of the CbH model. The ANFIS model's higher sensitivity suggests that it is more adept at identifying genuine positive cases. Sensitivity evaluates a model's capacity to accurately identify positive cases. Finally, the ANFIS model's specificity value is 15% greater than that of the NB model and 5% higher than that of the CbH model. The ANFIS model's higher specificity suggests that it is more adept at avoiding false positive cases. Specificity assesses a model's capacity to accurately detect negative cases. These findings clearly show that in terms of CKD prediction, the ANFIS model outperforms more traditional methods like NB and CbH. The ANFIS model has a high specificity value of 97%, 100% accuracy and sensitivity, and is a promising method for predicting CKD.

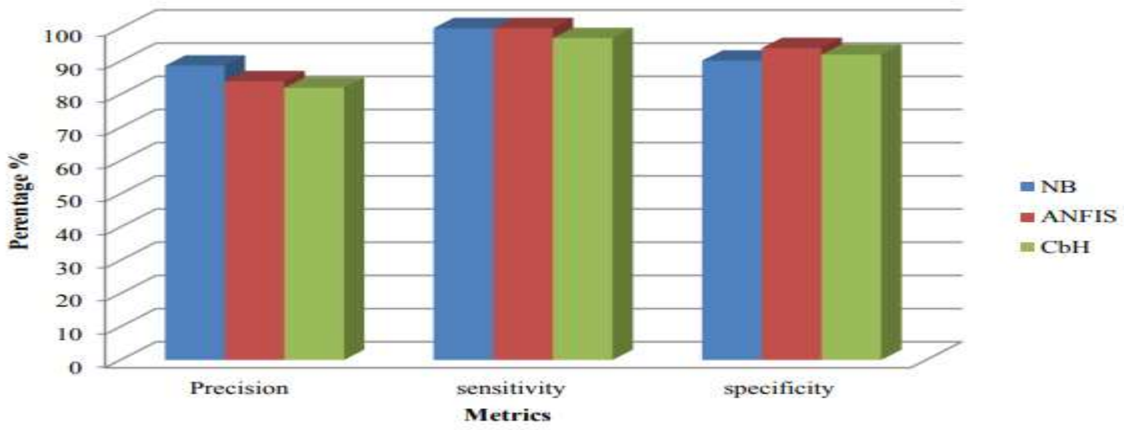

Fig ure.11: Comparison of performance metrics

The accuracy of the Adaptive Neuro-Fuzzy Inference System (ANFIS) model is 100%, compared to the 90.2% accuracy of the Naive Bayes (NB) and Choice based hierarchy (CbH) models. This shows that when compared to the NB and CbH models, the ANFIS model has a better accuracy in properly predicting situations. It has been discovered that the proposed ANFIS model is useful for early-stage prognosis of chronic kidney disease (CKD). Early detection of CKD is essential because it enables prompt medical intervention and decisions that can have a big impact on patient outcomes and quality of life. Nephrologists can better manage and care for CKD patients by using this predictive model to help them make decisions when they matter most. Early diagnosis and treatment may be able to prevent complications, limit the progression of the disease, and enhance patient outcomes. Predictive models may always need some improvement, though. Deep learning approaches can be researched and used in the future to improve the prediction model's accuracy and classification rate. Machine learning's subset, deep learning algorithms, have demonstrated considerable promise in a number of medical applications, including the prediction of disease. They can manage intricate linkages and patterns in data, which could increase CKD prediction performance. The suggested predictor model utilizing the ANFIS technique is, all things considered, a promising step towards early CKD prediction and enhanced patient treatment. The incorporation of deep learning techniques can improve the precision and potency of such models as technology and research advance, ultimately helping patients all across the world.

8. Conclusion

In comparison to individual models and conventional diagnostic techniques, the suggested ensemble-based methodology for predicting renal disorders using machine learning models shown important improvements. This study's main objective was to create a predictor model that could correctly forecast kidney disorders in their early stages. To do this, the researchers reduced the number of characteristics associated with chronic kidney disease (CKD) while still designing a powerful classifier with better classification and prediction accuracy. The most pertinent and instructive elements for the prediction model were chosen using feature engineering approaches. The important elements that are crucial for diagnosing renal illness were found by the researchers through an analysis of the literature and current research. The selected qualities were useful and pertinent for medical applications thanks to this thorough examination. Real-world information gathered from online sources was used to assess how well the projected model performed. The model's generalizability and suitability for use in practical situations were confirmed by include data from a variety of patient populations. The interpretability of the model's outcomes was one of this study's key contributions. The feature importance analysis offered insightful information about the variables affecting renal disease diagnoses. For healthcare practitioners, this interpretability can be quite helpful because it enables them to comprehend the underlying mechanisms and risk factors connected with renal illnesses. With this information at hand, healthcare providers can make better judgments and carry out targeted interventions to successfully monitor and cure renal illnesses. In terms of predicting renal illness, an ensemble-based technique combining feature engineering and interpretability analysis shows great promise. The suggested approach can help to earlier and more accurate diagnoses, improving patient outcomes and quality of treatment by combining the strengths of various machine learning models and offering insightful information for healthcare practitioners.

References

1. P. Chittora *et al.*, "Prediction of Chronic Kidney Disease - A Machine Learning Perspective," in *IEEE Access*, vol. 9, pp. 17312-17334, 2021, doi: 10.1109/ACCESS.2021.3053763.
2. W. D. Souza, L. C. D. Abreu, L. G. D. SilvaI and I. M. P. Bezerra, "Incidence of chronic kidney disease hospitalisations and mortality in Espírito Santo between 1996 to 2017", 2019.

3. Dhillon and A. Singh, "Machine learning in healthcare data analysis: A survey", *J. Biol. Today's World*, vol. 8, no. 2, pp. 1-10, Jan. 2018.
4. Y. Amirgaliyev, S. Shamiluulu and A. Serek, "Analysis of chronic kidney disease dataset by applying machine learning methods", *Proc. IEEE 12th Int. Conf. Appl. Inf. Commun. Technol. (AICT)*, pp. 1-4, Oct. 2018.
5. S. Drall, G. S. Drall, S. Singh and B. B. Naib, "Chronic kidney disease prediction using machine learning: A new approach", *Int. J. Manage. Technol. Eng.*, vol. 8, pp. 278-287, May 2018.
6. 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, pp. 4536-4541, May 2020
7. J. Xiao, R. Ding, X. Xu, H. Guan, X. Feng, T. Sun, et al., "Comparison and development of machine learning tools in the prediction of chronic kidney disease progression", *J. Transl. Med.*, vol. 17, pp. 119, Dec. 2019.
8. Nikhila, "Chronic Kidney Disease Prediction using Machine Learning Ensemble Algorithm," *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, Greater Noida, India, 2021, pp. 476-480, doi: 10.1109/ICCCIS51004.2021.9397144.
9. J. Aljaaf et al., "Early Prediction of Chronic Kidney Disease Using Machine Learning Supported by Predictive Analytics", *2018 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1-9, Jul. 2018.
10. E.-H. A. Rady and A. S. Anwar, "Prediction of kidney disease stages using data mining algorithms", *Inform. Med. Unlocked*, vol. 15, pp. 100178, Jan. 2019.
11. P. Arulanthu and E. Perumal, "Predicting the Chronic Kidney Disease using Various Classifiers", *2019 4th International Conference on Electrical Electronics Communication Computer Technologies and Optimization Techniques (ICEECCOT)*, pp. 70-75, Dec. 2019.
12. S. Y. Yashfi *et al.*, "Risk Prediction Of Chronic Kidney Disease Using Machine Learning Algorithms," *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kharagpur, India, 2020, pp. 1-5, doi: 10.1109/ICCCNT49239.2020.9225548.
13. Mamatha Bai B G Tabassum and Jharna Majumdar, "Analysis and Prediction of Chronic Kidney Disease using Data Mining Techniques", *International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)*, vol. 4, no. 9, pp. 25-31, 2017.
14. Sujata Drall, Gurdeep Singh Drall, Sugandha Singh and Bharat Bhushan Naib, "Chronic Kidney Disease Prediction Using Machine Learning: A New Approach", *International Journal of Management Technology And Engineering*, vol. 8, no. 5, pp. 278-287, 2018
15. Asif Salekin and Jhon Stankovic, "Detection of Chronic Kidney Disease and Selecting Important Predictive Attributes", *2016 IEEE International Conference on Healthcare Informatics (ICHI)*, pp. 262-270, 2016.

16. Sahil Sharma, Vinod Sharma and Atul Sharma, "Performance Based Evaluation of Various Machine Learning Classification Techniques for Chronic Kidney Disease Diagnosis", *International Journal of Modern Computer Science (IJMCS)*, vol. 4, no. 3, pp. 11-15, 2016.
17. Devishri S Pratibha, O R Ragin and G S Anisha, "Comparative Study of Classification Algorithms in Chronic Kidney Disease", *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 8, no. 1, pp. 180-184, 2019.
18. Gwozdzinski Krzysztof, Anna Pieniazek and Lukasz Gwozdzinski, "Reactive oxygen species and their involvement in red blood cell damage in chronic kidney disease", *Oxidative medicine and cellular longevity*, pp. 1-19, 2021.
19. F. M. Javed Mehedi Shamrat, P. Ghosh, M. H. Sadek, M. A. Kazi and S. Shultana, "Implementation of Machine Learning Algorithms to Detect the Prognosis Rate of Kidney Disease," 2020 IEEE International Conference for Innovation in Technology (INOCON), Bangluru, India, 2020, pp. 1-7, doi: 10.1109/INOCON50539.2020.9298026.
20. P. Ghosh, F. M. Javed Mehedi Shamrat, S. Shultana, S. Afrin, A. A. Anjum and A. A. Khan, "Optimization of Prediction Method of Chronic Kidney Disease Using Machine Learning Algorithm," 2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), Bangkok, Thailand, 2020, pp. 1-6, doi: 10.1109/iSAI-NLP51646.2020.9376787.
21. H. Zhang, C. Hung, W. C. Chu, P. Chiu and C. Y. Tang, "Chronic Kidney Disease Survival Prediction with Artificial Neural Networks", 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 1351-1356, 2018.
22. P. Yildirim, "Chronic Kidney Disease Prediction on Imbalanced Data by Multilayer Perceptron: Chronic Kidney Disease Prediction", 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC), pp. 193-198, 2017.