International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

PREDICTION OF CHRONIC KIDNEY DISEASE USING MACHINE LEARNING

J.Steephan Amalraj[1] , Sukirtha Sree V M[2] , Anish S[3] ,Nandhabalaa S[4]

[1], Assistant Professor, Dept. of Computer Science and Engineering,Bannari Amman Institute of Technology 2,3,4 UG Scholar, Dept. of Computer Science and Engineering,Bannari Amman Institute of Technology

---****---

Abstract - This project aims to predict chronic kidney disease and assess their reliability through the analysis of machine learning. By collecting data resources from various reports and also from existing datasets. We process and analyze the collected data to extract valuable information. Machine learning models were developed for Testing the collected data to forecast the probability of finding kidney diseases. The results of the project will benefit the patients who are facing kidney disease and help them to save their lives. A significant public health issue that affects millions of individuals globally is chronic kidney disease (CKD). To stop CKD from progressing to end-stage renal disease (ESRD), which necessitates dialysis or transplantation, early identification and management are crucial. Machine learning techniques may be used to discover the patterns connected to CKD by learning from big datasets of patient data. Machine learning can be used to predict CKD with great accuracy, according to a number of studies. First of all, it is a cheap and non-invasive technique to check for CKD. In order to implement early therapies, it can be utilized to identify individuals who are very susceptible to developing CKD. Thirdly, it may be used to track the development of CKD and forecast when patients could require dialysis or a transplant. One problem is that depending on the dataset used to train the algorithm, the prediction accuracy might change. Machine learning is a potential method for the early diagnosis and prediction of CKD despite these difficulties. The algorithms will get more precise and user-friendly as they advance. Better early CKD identification and management will result in better patient outcomes as a result of this.

Keywords :

Machine learning, Chronic kidney disease, prediction, early detection, management,Algorithms .

1. Introduction:

A brand-new and developing subject of study is chronic renal disease prediction and reliability analysis based on machine learning. By assessing patient data like age, blood pressure, and test results to forecast the chance of acquiring the condition, machine learning can play a significant role in CKD. The goal of this research is to develop a system that can identify sickness in people who arrive to hospitals with signs of renal disease.

The illness known as chronic kidney disease (CKD) causes the kidneys to progressively deteriorate over time. Early identification, individualized treatment regimens, and better patient outcomes can all be aided by this. Healthcare providers may make better management decisions for CKD

by using machine learning algorithms to find patterns and links in the data. The system would employ machine learning to keep track of past information from medical reports on kidney illness, such as the sugar content, potassium level, blood pressure, etc. Then, using this data, a machine learning model that can forecast chronic illness would be trained.A number of measures, including the accuracy, precision, and recall of the predictions, would be used to examine the predictability of the result. Additionally, the algorithm would be able to pinpoint the elements that influence how accurate the forecasts are. For doctors, the system would be a useful tool. By making sure that their patients may live happy lives, doctors can use the system to increase the effectiveness of their patients. Using patient information and the system, doctors might forecast disease more quickly and effectively.The caliber of the data used to train a machine learning model affects the model's accuracy. The accuracy of the model will increase with the amount of data utilized to train it.The accuracy of CKD prediction can be increased by using machine learning algorithms. It is crucial to remember that machine learning models are not flawless. They should not be utilized to make medical choices without contacting a doctor since they are still susceptible to errors.There are several benefits of using machine learning to predict CKD. First of all, it is a cheap and non-invasive technique to check for CKD. In order to implement early therapies, it can be utilized to identify individuals who are very susceptible to developing CKD. Thirdly, it may be used to track the development of CKD and forecast when patients could require dialysis or a transplant. There are several benefits of using machine learning to predict CKD. First of all, it is a cheap and non-invasive technique to check for CKD. In order to implement early therapies, it can be utilized to identify individuals who are very susceptible to developing CKD. Thirdly, it may be used to track the development of CKD and forecast when patients could require dialysis or a transplant.

1.1 Introduction to AI :

Artificial intelligence is a technique for teaching a computer, a robot operated by a computer, or software to think critically and creatively like a human mind. AI is achieved through examining the cognitive process and researching the patterns of the human brain. These research projects provide systems and software that are intelligent. In order to facilitate problem-solving, the discipline of artificial intelligence integrates computer science with substantial datasets. Additionally, it includes the branches of artificial intelligence known as deep learning and machine learning, which are commonly addressed together.Businesses and organizations may benefit from AI to better understand consumer behavior, market trends, and other crucial elements. Along with Herbert A Allen Newell, Marvin Minsky, and Alan Turing, he is one of the pioneers of artificial intelligence.

1.2 Introduction to Machine Learning :

Artificial intelligence (AI) known as "machine learning" focuses on using data and algorithms to mimic how humans learn, progressively increasing its accuracy. Artificial intelligence (AI) is a formidable field, and one that has exploded in prominence in recent years is machine learning. The art of creating statistical models and algorithms allows computers to learn and make predictions or choices without having to be explicitly programmed. Many technologies we use every day, such as recommendation systems on streaming platforms, driverless vehicles, and medical diagnostics, are built on machine learning.

Machine learning has the ability to transform industries and find solutions to challenging issues, but it also has obligations and constraints. For anyone interested in artificial intelligence, data science, and the future of technology, having a basic understanding of the field of machine learning is a valuable ability. In general, machine learning is a strong tool that can be used to address a wide range of issues. However, prior to employing machine learning, it is crucial to be aware of its drawbacks.

2. Literature Survey:

1. Glomerular filtration rate, or GFR, is a measuring characteristic for the prediction of kidney illness, according to research done in August 2015 by Dr. S. Vijayarani and Mr. S. Dhayanand. They have put into practice and contrasted the SVM (Support Vector Machine) and naive Bayes classification methods. According to their experimental findings, SVM are more accurate than Naive Bayes.
2. By using four categorization approaches on test results from patient medical reports, S. Ramya and Dr. N. Radha [4] created a methodology in January 2016 to predict renal function failure. 1000 entries with 15 qualities are available. They also contrasted these four methods: Random Forest, Radial Basis Function, and Back Propagation Neural Network. Their findings demonstrate that RBF (Radial Basis Function) has higher predictive value for chronic renal disease.
3. The chronic renal disease dataset was studied by Lambodar Jena and Narendra ku. Kamila [5] in November 2015 using a variety of classification approaches, including Nave Bayes, Multilayer Perceptron, Support Vector Machine, J48, Conjunctive Rule, and Decision Table. They made use of a weka program. For categorization, they have employed 25 distinct characteristics. According to their research, Multilayer Perceptron provides predictions for chronic renal disease with 99.75% more accuracy than previous methodologies.
4. Parul Sinha and Poonam Sinha [6] created a decision support system to forecast chronic renal disease in December 2015. Support Vector Machine and KNN (K Nearest Neighbor) results have been compared. Their research indicates that KNN is more accurate than SVM.
5. For the purpose of predicting renal illness, P Swathi Baby and Panduranga Vital [7] employed machine learning methods such as AD Trees, J48, KStar, and Naive Bayes Random Forest in July 2015. According to their analysis, Naive Bayes has the greatest rate of accuracy at 100%.
6. Abeer and Ahmad [8] constructed the SVM and Logistic Regression (LR) data mining classifiers in October 2014. Their findings demonstrated that SVM, with a 93.14 percent accuracy rate, is more accurate than other methods.
7. A new chronic renal disease dataset and three classifiers—radial basis function network, multilayer perceptron, and logistics regression—were implemented by Jurlin Rubini and Dr. P. Eswaran in July 2015 [9]. Finally, they discovered that the

multilayer perceptron outperforms the other two classifiers in terms of accuracy. Pushpa M. Patil, International Journal of Computer Science and Mobile Computing, Vol. 5, No. 5, May 2016, p. 135–141. IJCSMC 2016, Unless Otherwise Noted 137

8. Ruey Key [10] constructed three alternative neural network models in 2015, including the backpropagation neural network (BPN), the generalized feed forward neural network (GRNN), and the modular neural network (MNN), for the identification of chronic renal disease. He furthered the implementation of these models in his study by including the (GA) genetic algorithm into each associated neural element. All three models in the experiment are more accurate, or over 85%. According to observations, the back propagation neural network (BPN) outperforms the other two models in terms of accuracy.
9. Manish Kumar [12] used six distinct data mining approaches, including Random Forest Classifiers, Sequential Minimal Optimization, Naive Bays, Radial Basis Function, Multilayer Perceptron Classifier (MLPC), and Simple Logistic, to predict chronic renal disease in February 2016. He trained the prediction system using a total of 400 records. He has discovered that Random Forest has the highest accuracy among these methods.
10. Kora R.Various machine learning methods, including as Decision Trees, Random Forests, Support Vector Machines, and Neural Networks, used for CKD prediction are discussed in this paper. The study rates the effectiveness and contrasts several models according to their accuracy, sensitivity, and specificity.
11. For the identification of erythemato-squamous disorders, Ubeyli et. al[9] employed a hybrid approach that combines multiclass SVM with error-correcting output code (ECOC). The dataset for this erythematous squamous illness, which consists of 34 characteristics and six distinct classifications, has been retrieved from the UCI repository. The classification accuracy of this model is 98.32%.
12. In order to pick the best model for the diagnosis of erythematous squamous, Ubeyli [10] also utilized a mixed neural network and achieved a classification accuracy of 97.77%.

3. OBJECTIVE AND METHODOLOGY

3.1 OBJECTIVES:

The objective of the machine learning based Kidney Disease Prediction project is to address critical challenges in humans by providing advanced machine learning models technologies and artificial intelligence to significantly improve kidney disease detection and prediction. Machine learning can be used to analyze various features and patterns in medical data to predict chronic kidney disease.By training a model on a dataset of patient information, the model can learn to make predictions about whether a patient is likely to have chronic kidney disease. This project creates a user-friendly and accessible online platform that helps doctors,patients to reduce diseases effectively.Through the integration of AI, machine learning models, the project aims to provide users with rapidand accurate predictions of kidney diseases based on

International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

uploaded reports. By doing so, it not only prevents early disease detection but also enables data-driven decision-making for disease prevention and control strategies, leading to a higher life rate and reduced environmental impact. Furthermore, this project helps by encouraging medical sustainability by offering disease diagnosis ideas based on detected diseases. The goal is to empower medical professionals with a comprehensive tool that enhances human health, increases population productivity, and ensures food security while minimizing the economic and environmental costs associated with kidney diseases. Through the deployment of this cloud-based platform, the project endeavors to bring about a positive transformation in the medical landscape, promoting sustainable and data-driven practices that benefit both doctors and the humans. This can help in early detection and intervention,leading to better patient outcomes and improved healthcare management. To find patterns in user data and then anticipate outcomes based on these complex patterns in order to respond to business inquiries and resolve business issues. To examine patient symptoms in order to identify various diseases using various machine learning modeling approaches.

3.2 METHODOLOGY :

There are many crucial elements in the process for applying machine learning to predict chronic kidney disease (CKD). The first step is to gather and preprocess a large dataset that includes patient demographics, medical history, lifestyle variables, and pertinent laboratory test results. To prepare the data for machine learning algorithms, this preparation stage handles missing values, normalizes features, and encodes .There are many crucial elements in the process for applying machine learning to predict chronic kidney disease (CKD). The first step is to gather and preprocess a large dataset that includes patient demographics, medical history, lifestyle variables, and pertinent laboratory test results. To prepare the data for machine learning algorithms, this preparation stage handles missing values ,normalizes and encodes.

3.3 MACHINE LEARNING:

Computers can learn without being explicitly programmed thanks to a sort of artificial intelligence called machine learning. To identify those who are most at risk for developing CKD, forecast who will get it, and track CKD's progression, machine learning techniques can be utilized.Machine learning techniques come in a variety of forms and can be applied to predict CKD. Logistic regression, decision trees

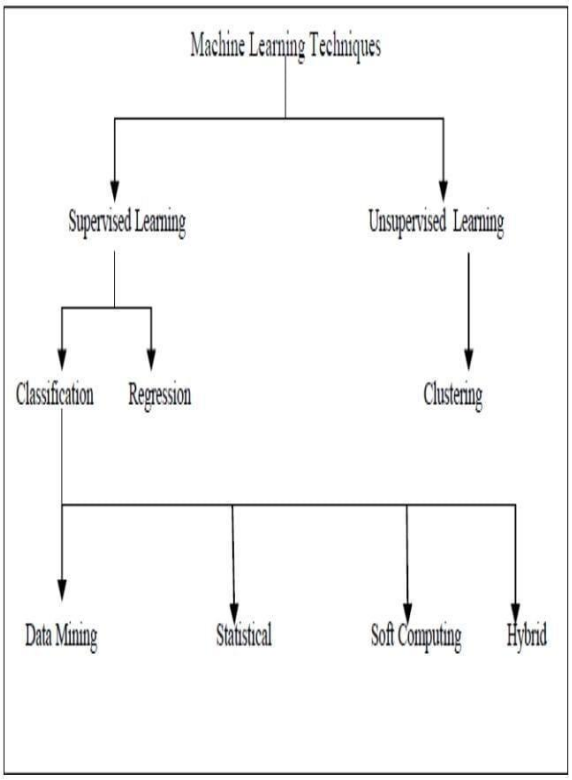

Fig 1. flowchart of ML techniques

Artificial intelligence (AI) in the form of machine learning (ML) enables software programs to forecast outcomes more accurately without having to be explicitly programmed to do so. In order to forecast new output values, machine learning algorithms use past data as input. Making models that can learn from data and make predictions or judgments without being explicitly programmed is the aim of machine learning. Algorithms for machine learning are trained on data, where they discover patterns in the data. Once they have figured out these patterns, they may apply them to forecast the outcome of fresh data.

3.3.1 LOGISTIC REGRESSION :

A sort of supervised learning technique that may be used to forecast the likelihood of an event occurring is logistic regression. Logistic regression may be used to determine the likelihood that someone will acquire CKD in the instance of CKD. A statistical model called logistic regression employs the logistic function, often known as the logit function in mathematics, as the relationship between x and y. Y is represented by the logit function as a sigmoid function of x. You will obtain an S-curve as shown below if you plot this logistic regression equation. The binary outcome and independent variables are included in the data set used to train the logistic regression model. The binary result and the independent variables are correlated by the model. Once trained, the model may be used to forecast the likelihood of a binary result given fresh data.

There are three types of logistic regression .

1. Binary.
2. Multinomial .
3. Ordinal .

3.3.2 DECISION TREE :

A supervised learning approach that may be used to categorize data is decision trees. Decision trees may be used to divide people into two categories in the case of CKD those who are likely to acquire CKD and those who are not. Decision tree regression trains a model in the form of a tree to predict data in the future and provide useful continuous output by observing the properties of an item. Continuous output denotes the absence of discrete output, i.e., the absence of representation by a discrete, well-known set of numbers or values. The non-parametric supervised learning approach used for classification and regression applications is the decision tree. It is organized hierarchically and has a root node, branches, internal nodes, and leaf nodes. Both classification and regression issues may be solved with it: Decision trees are effective in both classification and regression applications because they may be used to predict both continuous and discrete values.

3.3.3 RANDOM FOREST:

An ensemble learning approach known as random forests can be used to increase prediction accuracy. Making a lot of decision trees, then averaging their forecasts is how RF operate. As a form of ensemble learning method that mixes numerous decision trees to provide predictions, random forest is a supervised machine learning technique that may be used for both classification and regression applications. In order for the random forest technique to function, several decision trees must first be built. A random selection of the training data is used to train each decision tree. The forecast is then created by combining the trees. The majority vote of the trees is used to determine the outcome.

Some major Applications of Random Forest in different sectors:

- Banking Industry. Credit Card Fraud Detection. Customer Segmentation.
- Healthcare and Medicine. Cardiovascular Disease Prediction. Diabetes Prediction.
- Stock Market Prediction and Sentiment Analysis.
- E-Commerce.
- Product Recommendation.

3.3.4 SUPPORT VECTOR MACHINE :

Support vector machines (SVMs) are a form of supervised machine learning technique that may be used for regression and classification applications. SVMs are renowned for their capacity to manage complicated data and produce precise forecasts. Finding the ideal hyperplane between two classes of data is the aim of an SVM. A line or curve known as a hyperplane splits a space into two halves. In order to maximize the margin between the two classes, the SVM seeks out the hyperplane. The margin is the separation between the nearest data points from each class and the hyperplane. A versatile machine learning technique, SVMs are effective for a number of tasks. They are renowned for their precision and capacity for working with complicated data. SVMs can be challenging to comprehend and can be computationally costly to train. Both linear and non-linear classification issues may be solved with SVMs. The hyperplane can be a straight line for linear issues. The hyperplane for non-linear problems might take the form of a curve. SVMs may be applied to many different situations, such as:

1. Classification.
2. Regression.
3. Outlier detection.
4. Feature selection.

4. PROPOSED WORK MODULES:

These are only a few of the suggested task modules for the machine learning-based CKD prediction. Depending on the dataset and machine learning method being utilized, certain modules may be employed.The development of accurate and reliable machine learning models for CKD prediction faces a variety of problems in addition to the outlined task modules. These difficulties include:

1. The lack of important, fully documented patient data sets.
2. The CKD disease process depth.
3. CKD risk factor variation among various groups.
4. The requirement for the creation of interpretable models that may be utilized to direct clinical judgment.

International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

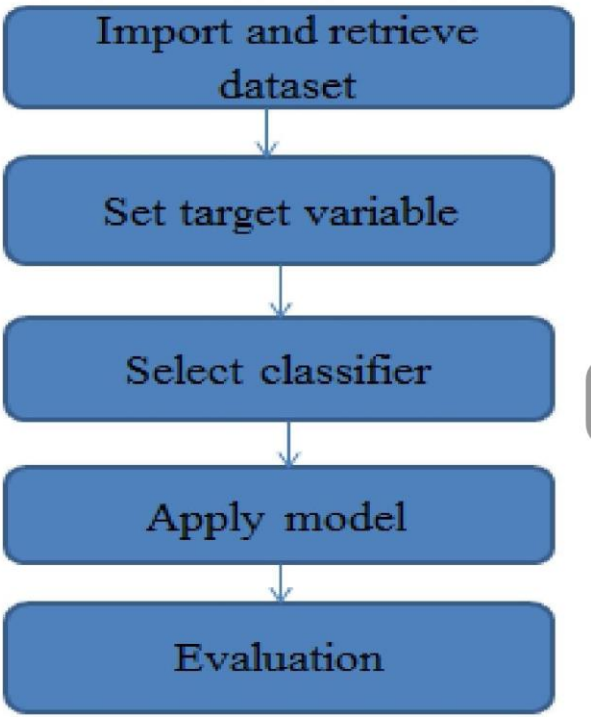

Fig 2. working modules

4.1 DATA COLLECTION:

Assemble a huge database with pertinent patient data, such as their demographics, medical history, test findings (such as serum creatinine levels), smoking status, and CKD diagnostic status (labeled data).Observation, interviews, questionnaires, scheduling, and surveys are the basic methods for acquiring data. The process of acquiring and analyzing data is called data collection. It is the initial stage of machine learning. In order to train the machine learning model, the data obtained will be used.

4.2 DATA PREPROCESSING:

Cleanse the data by managing outliers and missing values correctly (e.g., removal, imputation). To guarantee uniform scaling, normalize or standardize numerical properties. Using a technique like one-hot encoding, convert category variables to a numerical representation.

Preparing unprocessed data for machine learning is known as data preparation. It is a crucial stage in the machine learning pipeline since it has a big effect on the precision and efficiency of the models.Preprocessing data is to convert raw data into a format that machine learning algorithms can use. The data may need to be cleaned, outliers eliminated, missing values input, scaled, and converted into a new format.

4.3 FEATURE SELECTION:

To find the characteristics that are most important for CKD prediction, use exploratory data analysis. Think about developing new useful features using feature engineering, including ratios of pertinent lab data. To choose the most useful characteristics, use methods like feature importance ratings and correlation analysis.
The process of choosing the most pertinent features for a machine learning model from a dataset is known as feature selection. By lowering the noise and redundancy in the data, feature selection aims to enhance the performance of the model.

4.4 DATA SPLITTING :

To analyze the performance of the model, separate the dataset into training, validation, and testing sets.
If there are any class imbalances, make sure the data split is representative and takes them into account. A dataset is split into two or more subgroups through the process of data splitting. In machine learning, this is done to train and test the models. The model is trained using the training set, and its effectiveness is assessed using the test set.
A dataset can be divided in a variety of ways. The holdout method is the most typical approach. This approach divides the dataset

into two parts: a test set that is randomly chosen, and a training set that is used on the remaining data. Usually 20–30% of the entire dataset is the size of the test set.

4.5 MODEL SELECTION :

For CKD prediction, use suitable machine learning techniques. Logistic regression, random forests, support vector machines, gradient boosting, and neural networks ,decision tree are popular options. In machine learning, the process of selecting the best model for a given dataset is known as model selection. This is accomplished by comparing many models and selecting the one that excels on the training set.When choosing a model, there are many various things to take into account, including the model's precision, complexity, and interpretability.

4.6 MODEL TRAINING :

Use the specified algorithm to train the chosen machine learning model on the training dataset. To enhance model performance, adjust hyperparameters using strategies like grid search or random search. If CKD is a relatively uncommon result in the dataset, use approaches to manage class imbalance (e.g., oversampling, undersampling, synthetic data creation).
The process of fitting a machine learning model to a dataset is known as model training in machine learning. Finding the model parameters that reduce the difference between the model's predictions and the actual values in the dataset is the aim of model training.

4.7 MODEL EVALUATION:

Utilize the proper assessment measures, such as accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrices, to rate the model's performance on the validation set. To ensure robustness and lower the danger of overfitting, use cross-validation.
The particular problem that has to be solved will determine which measure is used. For instance, accuracy would be an excellent statistic to consider when trying to categorize photographs as either cats or dogs. However, accuracy and recall would be more useful measurements if the aim were to detect fraudulent transactions.

4.8 MODEL TESTING AND DEPLOYMENT:

Test the model on the held-out testing set after obtaining satisfactory performance on the validation set to get impartial performance measures.
Implement the model in a healthcare setting and, if necessary, connect it with electronic health records (EHR).

5. RESULT & DISCUSSION:

5.1 RESULT :

The effectiveness of the model and algorithm has been verified through simulation with a prediction of 97%.

The quality and quantity of the data, the machine learning algorithms chosen, and the preparation techniques can all affect the final outcomes. Additionally, healthcare applications like CKD prediction frequently emphasize interpretability and clinical relevance, thus working with medical professionals is crucial to validating and optimizing the model for practical application.

International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

Fig 3. Decision tree classifier

5.2 DISCUSSION:

- The impact of different data collection methods on the accuracy of the predictions.
- The challenges of dealing with missing data.
- The ethical implications of using machine learning to predict chronic kidney disease
- To train machine learning models for CKD prediction, extensive, well-annotated datasets are required.
- The significance of taking the clinical setting into account when analyzing the outcomes of machine learning models for CKD prediction.

6. CONCLUSION & SUGGESTION FOR FUTURE WORK:

6.1 CONCLUSION :

The project "Prediction of chronic kidney disease using machine learning" has made significant progress in the prediction of chronic kidney disease. The results of the project suggest that machine learning can be used to effectively predict chronic kidney diseases. This information can be used to predict diseases that help doctors to save lives more easily and efficiently. Chronic kidney disease is currently the largest cause of mortality, hence early detection is crucial. The performance of the classification system on the chronic renal disease dataset was 99.33% Specificity and 99.20% Sensitivity. Additionally, we are focusing on improving the performance of prediction systems using deep learning and neural networks.

6.2 SUGGESTION FOR FUTURE WORK:

Future research on the outcome of chronic kidney disease (CKD) may be useful in a number of areas. The creation of more precise and trustworthy prediction models is a crucial subject for future research. Although the existing models are helpful, they are not yet flawless and occasionally categorize patients incorrectly. The improvement of CKD early identification and the possibility of earlier intervention would result from the development of more accurate models. The discovery of novel CKD risk factors is a significant subject for future research. The models used today are based on a small number of risk variables, and it's probable that additional factors also play a role in the onset of CKD. Finding these additional risk variables would assist to increase the predictive model's accuracy.

□□□ □□□□ □□□□ □□□□□□□ : The more data you have, the more accurate your predictions will be. You can collect data from a variety of datasets.

□□□ □□□□ □□□□□□□□□□□□□ □□□□□□□ □□□□□□□□ □□□□□□□□□□ : There are a variety of machine learning algorithms that can be used for prediction of chronic kidney disease. Some of the more sophisticated algorithms, such as deep learning, can produce more accurate predictions.

□□□□ □□□□ □□□□□□□□□□□ □□ □ □□□□□□□ □□□□□□□ : A holdout dataset is a set of data that is not used to train the machine learning model. This data is used to test the accuracy of the model's predictions.

□□□□□□□□ □□□ □□□□□□□□□□□□ □□ □□□□ □□□□□□□□□□□□ : In addition to accuracy, it is also important to evaluate the reliability of your predictions. This means assessing how confident you can be in the predictions.

□□□□ □□□□ □□□□□□□□□□□□ □□□□□□□□□□□ : The goal of our project is to predict chronic kidney disease so that patients can live happily. our predictions will be easy to understand and use.

7. REFERENCE :

1. "A Survey On Implementation Of Machine Learning Techniques For Dermatology Diseases Classification", International Journal of Advances in Engineering & Technology, Madhura Rambhajani, Wyomesh Deepanker, and Neelam Pathak (2015).
2. Manish Kumar (2016), "Prediction Of Chronic Kidney Disease Using Random Forest Machine Learning Algorithm", International Journal of Computer Science and Mobile Computing, Vol. 5, Issue. 2, p. 24–33. G. Parthiban and K. R. Anantha Padmanaban (2016).
3. Indian Journal of Science and Technology, Vol. 9(29), "Applying Machine Learning Techniques For Predicting The Risk Of Chronic Kidney Disease."
4. Predictive analytics for chronic renal disease using machine learning approaches, Anusorn Charleonnan, Thipwan Fufaung, Tippawan Niyomwong, Wandee Chokchueypattanakit, Sathit Suwannawach, and Nitat Ninchawee (2016), The 2016 Management and Innovation Technology International Conference.
5. Using machine learning and data mining techniques, Jaymin Patel, Prof. Tejal Upadhyay, and Dr. Samir Patel (2016) published "Heart Disease Prediction Using Machine Learning and Data Mining Technique" in International Journal Of Computer Science & Communication, Vol. 7, No. 1, pp. 129–137.
6. Missing data imputation utilizing statistical and machine learning approaches in a real breast cancer problem, Jerez, J. M., Molina, I., Garcia-Laencina, P. J., Alba, E., Ribelles, N., Martin, M., and Franco, L. (2010). A.I., Med. 50, 2, 11–11.
7. (2007). A. Asuncion and D. J. Newman. [Online] UCI Machine Learning Repository. Available at: MLRepository.html (www.ics.uci.edu/mlearn).
8. Data mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann Series in Data Management Systems, Witten, Ian H. (2011).
9. P. B. Jensen, L. J. Jensen, and S. Brunak (2012) "Mining electronic health records: towards better research applications and clinical care," Nature Reviews Genetics, Vol. 13, no. 6, pp. 395–405.
10. "Septic shock prediction for patients with missing data," ACM Transactions on Management Information Systems (TMIS), Vol. 5, no. 1, p. 1, J. C. Ho, C. H. Lee, and J. Ghosh (2014).
11. "A relative similarity based method for interactive patient risk prediction," Data Mining and Knowledge Discovery, vol. 29, no. 4, pp. 1070–1093; B. Qian, X. Wang,

N. Cao, H. Li, and Y.–G. Jiang (2015).
12. The article "Types of Machine Learning Algorithms" by Taiwo Oladipupo Ayodele was published in 2010 in New Advances in Machine Learning, edited by Yagang Zhang, published by InTech.