iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** II **Month of publication:** February 2023

DOI: https://doi.org/10.22214/ijraset.2023.49003

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue II Feb 2023- Available at www.ijraset.com

Chronic Kidney Disease Detection with Appropriate Diet Plan

Anubhav Raina[1], Nandini S S[2], Ritik Mrinal Purbey[3], Hemant Kumar [4], Dr. Mohd Tajuddin[5]
[1, 2, 3, 4]*Student,* [5]*Professor, Department of Computer Science, Dayananda Sagar College of Engineering, Bangalore, India*

Abstract: ***CKD doesn't have any distinct symptoms, it can be difficult to anticipate and prevent, which may result in long-term health issues. Here, it was intended to reduce diagnostic lead time and boost precision. The major goal of this work is to develop a predictive model for chronic kidney disease using data analysis and several machine learning techniques. Accuracy will be determined by contrasting various algorithms, including SVM, Random forest, and Naive Bayes.***
Based on information gathered from patient records, the algorithm determines whether a patient has chronic kidney disease and, if so, the severity of the condition based on blood potassium levels and a recommended diet.
Keywords: SVM, Naive Bayes, Glomerular Filtration Rate, Blood potassium level, Machine learning, renal illness.

I. INTRODUCTION

Overview
An introduction to the topic is provided, as well as a comparison of the present method to the proposed method. Motivation for selecting this application. Proposed algorithms, methodology/modules, data flow diagram, parameters table, and conclusion.
Machine learning is a part of artificial intelligence allowing machines to learn and enables software programmes or systems to get more and more accurate at predicting events without being explicitly programmed to do so. The aim is to analyse data structure and create models by fittting data so that it can be understood and used.
Chronic renal disease is defined as active kidney damage for more than three to four months. When the kidneys are unable to eliminate excess water or waste from the blood, it can lead to high blood pressure, anemia, bone weakness, poor nutritional health, and nerve damage. It also raises the risk of cardiovascular diseases. So it is critical to detect CKD at an early stage, but it is unpredictable because its symptoms are not specific to the disease, and some patients have no symptoms at all, so machine learning can help forecast whether the patient has CKD or not. This can be accomplished by leveraging previous data from CKD patients to train the algorithm and predict the result. Glomerular Filteration Rate (GFR) is a test used to assess renal function and stage of CKD. Based on a patient's blood potassium level, the model assesses existence or absence of chronic kidney disease utilising ML together with data mining, including finding as well as processing operations. Following this, the condition can be categorised into five phases, with stage one being safe and requiring only a lenient diet. Stage 2 necessitates a stringent and restricted diet. Mineral liquid equilibrium in the body can be challenging in stages 3 to 5, thus adequate food guidance is required. The basis to determine a proper food plan for a CKD patient is the level of potassium present in his or her blood. An essential diet is required for renal improvement and to prevent additional injury. Machine learning is more efficient in terms of achieving high accuracy in predicting CKD, as well as other diseases. The algorithms mentioned will be used to determine which provides the most accurate prediction of the condition. Clinics and hospitals can utilise this more efficient digital technology to predict chronic kidney disease.

II. LITERATURE SURVEY

A. Existing Approach Challenge

CKD is often diagnosed with the help of clinical data, imaging, various lab tests and biopsy. As a common diagnostic test, biopsies can cause infection, misdiagnosis, and are dangerous, expensive, and time consuming. Imaging studies have been utilized for a long time, although they have significant drawbacks, such as the influence of radiation. Many academics are currently involved with data mining algorithms in various kidney disease strategies in the existing system. There are few strategies available, and they are incapable of optimizing by boosting efficiency

B. Proposed Method

Methodology is a proposed method that works with the theoretical capability of researchbased work. It will provide clear information regarding the work concept. To make things easier, the study topic and instrumentation will be identified initially.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue II Feb 2023- Available at www.ijraset.com

To detect if chronic kidney diseases are present or not and determine performance/accuracy, various ML techniques such as SVM, Nave Bayes as well as Random forest are used.

C. Motivation

It takes several days and multiple tests to predict CKD. So, in order to prevent further harm, we came up with the concept of determining CKD using the main parameters without using an X-ray.

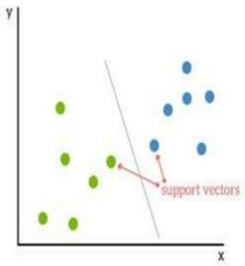
D. Content

Authors: AkashMaurya, Rahul Wable, Rasika Shinde, Sebin
John, Rahul Jadhav, R Dakshayani
Proposed Work: CHRONIC KIDNEY DISEASE PREDICTION AND RECOMMENDATION OF SUITABLE
DIET PLAN BY USING MACHINE LEARNING
Authors: Dr. S. Vijayarani, Mr.S.Dhayanand
Proposed Work: DATA MINING CLASSIFICATION ALGORITHMS FOR KIDNEY DISEASE PREDICTION

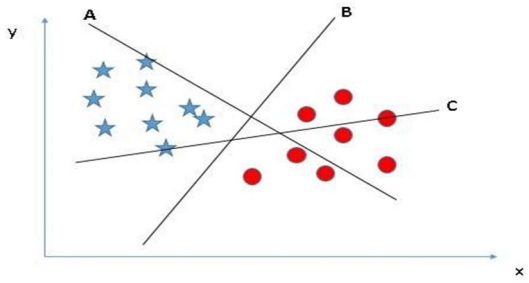
E. Algorithms

1) SVM Algorithm

The "Support Vector Machine" (SVM) is based on supervised machine learning that can be used both in classification and regression contexts. It is though generally used in categorization difficulties. Every data value is plotted as a point in space with n dimensions (n being the cardinality of features), where every feature's value pertains to some specific coordinates. Classification is accomplished as the best possible hyper plane that differentiates between the two classes is located(as shown in the graph below).



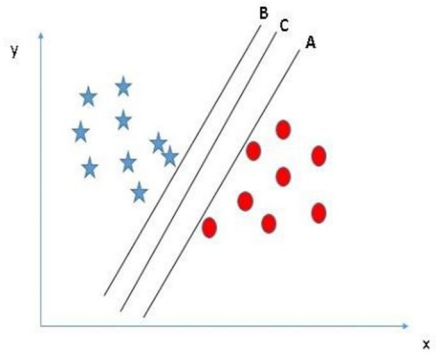
Co-ordinates that have distinct observations are used to calculate support vectors. Support Vector Machine is a divider which divides the classes in the best possible way (can be lines or hyper planes).
How does it function?
We became familiar with the method of dividing two classes using a hyper plane in the previous paragraph. The big question now is, "How do we find the proper hyper-plane?"
Let us try to comprehend :
Identifying the appropriate hyper plane (Case1): We have three hyper-planes here (A, B and C). Select the appropriate one to classify the stars and the circles.



©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue II Feb 2023- Available at www.ijraset.com

For finding the most suitable hyper-plane, we have to keep the following in memory: "Select the hyper-plane that best separates the two classes." Pertaining to the aforementioned circumstance, "B" did great work.
Identification of the most suitable hyper plane (Case-2): There are three of them here (A, B, and C) that can efficaciously segregate classes. But how is it that can the most suitable hyperplane can be found?



In this case, maximizing the separation of the closest data value (regardless of class) and the hyper-plane assists in determining the most suitable hyper-plane. The distance is called margin. Scikit-learn is a popular Python toolkit to implement ML algorithms. SVM is integrated with it and follows the similar framework (Importing library, creating object, fit the model and predicting).
Let's take a look at a real-world issue statement and dataset to see how SVM may be used for classification. SVM Advantages and disadvantages

a) Advantages

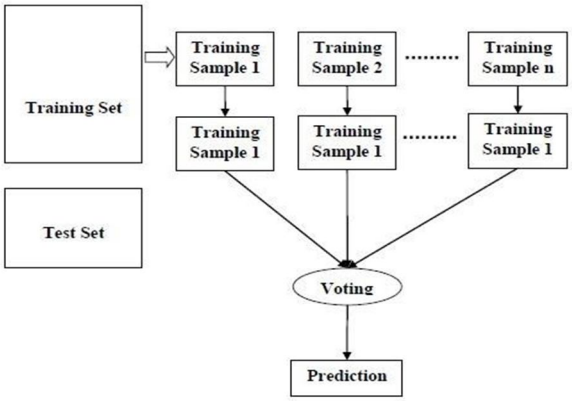
Its efficacy is high in the case we have a unambiguous separation margin.
It fares nicely especially in spaces involving three dimensions. Is useful in the case of the cardinality of dimensions exceeds that of the samples.
It is also memory efficient because it uses a subset of training points in the decision function (called support vectors).

b) Disadvantages

Won't function well in case of noisy data, i.e. it is difficult to differentiate the target classes .
Won't provide us the probabilistic approximations straightaway; those get aquired via an expensive five-fold cross-validation process. It is similar to the SVC method of the Python scikitlearn module.

2) Random Forest Algorithm

We define random forest as a meta estimator that creates and fits a number of decision tree classifierson various random sub-samples of the dataset . The ultimate class is obtained by taking into account the various votes of those classifiers for the test object. Random Forest improves the algorithm's prediction capability while simultaneously preventing over fitting. The most basic and extensively used algorithm is random forest. Used for classification as well as regression. It is a collection of randomised decision trees.



©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue II Feb 2023- Available at www.ijraset.com

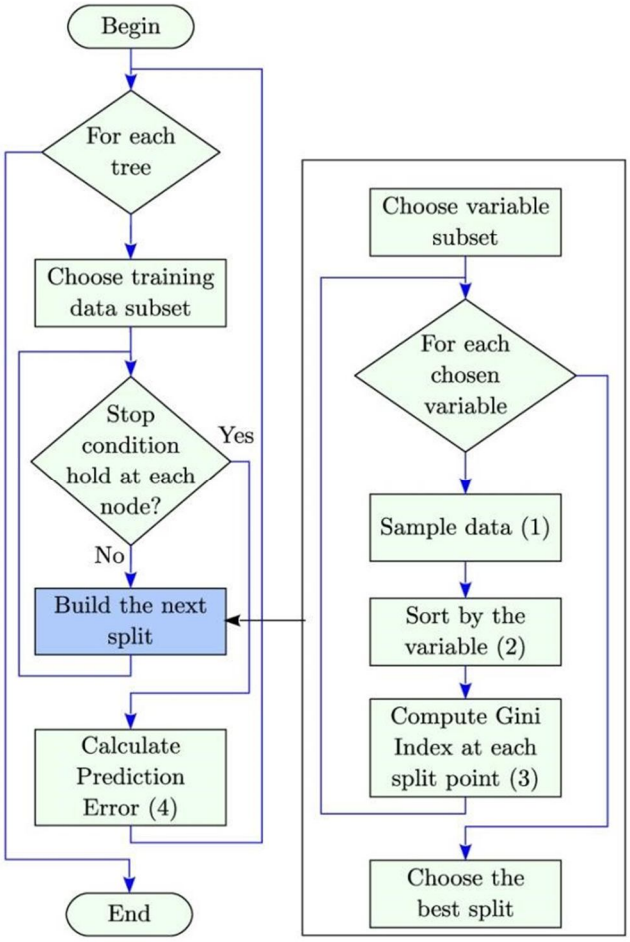
a) Advantages

- In case of a broad collection of data items, random forests generally tend to give better performance when compared to a sole decision tree.
- One of the biggest upsides is that a more thanhealthy as well as tolerable accuracy is maintained without the need to scale the data.
- Data generally needs to be scaled in other algorithms. In case of random forests, this does not pose any kind of problems whatsoever.
- When it comes to deviation or variance, random forests seem to have less of that in comparison to a sole decision tree, as well as being much more flexible and precise.
- Random forest algorithms are able to keep up a healthy performance metric in spite of a substantial chunk of data being amiss.
- Using the algorithm, it is easy to circumvent overfitting as it works by taking the average or combining the results of a number of decision trees.

b) Random Forest Algorithm Working

The various phases that enable anyone to comprehend the working of the Random Forest algorithm are as follows.

- *Phase 1:* In this phase, we start with the selection of arbitrary samples from a dataset of interest.
- *Phase 2:* In this phase, a decision tree is built for each sample using the algorithm. We then aquire the prediction outcome for each and every one of these decision trees.
- *Phase 3:* In this phase, taking into account each and every expected outcome, an election is held.
- *Phase 4:* In this phase, we elect the most accepted prediction outcome as the conclusive prediction outcome. The following figure shows how it works:



FLOW CHART OF RANDOM FOREST ALGORITHM

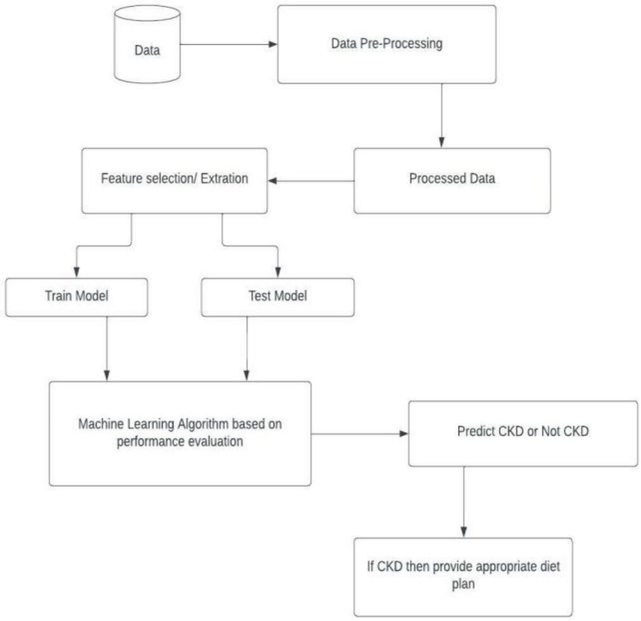
©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue II Feb 2023- Available at www.ijraset.com

F. Naïve Bayes

When we talk about the Naive Bayes classifier, it becomes important to remember that we define it as an uncomplicated and straightforward classifier based on probability with its bases being assuming solid (naive) individuality and the Bayes' theorem (which is an integral part of Bayesian statistics). An "independent feature model" is perhaps the best way to express the underlying model that is probabilistic in nature. When considering practicality and uses involving very realistic nature, however, this narrow individuality assumption seems to be incorrect more often than not. We thus use the keyword Naïve, and in spite of the aforementioned problem, the algorithm is a swift learner and is quite efficacious and is thus suited to cases that come under supervised classification. In other words, one of the biggest upsides to the Naïve Bayes classifier is that parameters , which generally are the mean as well as the variance, are approximated efficiently even without utilizing a substantial portion of the data used for training. The determination of the complete matrix of variance is not necessary but for every class just the variances pertaining to the variables due to the rudimentary assumption of individual variables.

Diagram



The Bayes theorem states that

1) The probability of the occurrence of an event C given the event X has already happened i.e. P (C|X) equals the expression P (X|C) •P(C) / P (X).
2) Across every class, P(X) remains fixed.
3) P(C) = relative frequency of class C samples c in such a way that p increases=c As a result, P (X|C) P(C) increases.
4) The problem is that computing P (X|C) is impossible.

G. Methodology/Modules

1) *Business Requirement:* In this section, we look at the business purpose and what its expectations are, and then create a model to fulfil the objective. Our primary goal is to determine a patient's CKD status as quickly as possible. This will allow doctors to predict the disease more accurately and in less time.
2) *Data Understanding And Collection:* This starts with collecting the data based on business requirements. We analyze the business plan and look for the type of data needed to reach our aim. Data is gathered from a variety of health centers, online platforms, and medical facilities.
3) *Data Preparation:* Once the data is available, exploratory data is analysed and preprocessed. Nominal values are converted to Real Numbers. Because the data is actual, there are missing values, which are then filled by the Mean value of the selected attribute. This makes no difference to the original data. **Parameters table** Correlation between the attributes is discovered in order to uncover disease-causing attributes that are highly connected.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue II Feb 2023- Available at www.ijraset.com

Table 1: Stages of CKD with estimated Glomerular Filteration Rate (GFR)

STAGES	DESCRIPTION	GFR (milliliters per minute)
G1	GFR normal or increased	>=90
G2	GFR decreased mildly	60-89
G3	GFR decreased moderately	30-59
G4	GFR significantly dropped	15-29
G5	kidney failure or dialysis	<15

4) Model Training: During this phase, just a few attributes are chosen. Highly Correlated attributes are chosen to help the machine learn more appropriately and precisely, resulting in greater accuracy. The classification model for prediction has to be created.

5) *Data Partitioning For Training And Testing:* The process involves partitioning a dataset into two subsets. Train Dataset: This dataset is used to fit the machine learning model. Test Dataset: Used to assess the fit of the machine learning model.

Table 2: Attributes of Data

ATTRIBUTES	USED VALUES
Age	Years
Blood Pressure	mm/Hg
Albumin	Nominal Values
Sugar	Nominal Values
Red blood cells	Normal or Abnormal
Pus cell	Normal or Abnormal
Pus cells clumps	Present or Not Present
Blood glucose random	mg/dl
Blood urea	mg/dl
Serum creatinine	mg/dl
Sodium	mEq/L
Potassium	mEq/L
Hemoglobin	g
Packed cell volume	Discrete Integer Values
White blood cell count	Cells/cumm
Red blood cell count	Millions/cmm
Anemia	Yes or No
Appetite	Good or Bad
Class	CKD or Not CKD

©IJRASET: All Rights are Reserved

III. CONCLUSION

Depending on how much user participation is necessary, different types of brain tumor segmentation techniques exist, they are as follows: manual, semi-automated, and automatic. In this surveyed paper, fully automated segmentation techniques are prioritized, since the manual segmentation techniques have significant drawbacks. While semiautomatic methods consumes less time than manual methods and can still produce effective outcomes, yet they are susceptible to user variability. Hence, the majority of current investigation on brain tumor segmentation is principally centered on automatic techniques. Exploitation of fully automated segmentation techniques will maximize the accuracy rate and minimize the error rate. In order to attain the best accuracy and quickest processing times, some pertinent publications of research on brain tumor identification were reviewed and how different machine learning and deep learning approaches were implemented. These techniques embrace SVM, K-Means, ANN, CNN When compared to other methods, our research demonstrates that the CNN algorithm achieved the highest accuracy and precision rate, which was approximately 97.9%. It was additionally brought to our observation that deep learning techniques have gained interest so as to enhance the accuracy and transparency of tumor prediction.

Hence, the project will be carried forward with CNN technique along with the usage of MATLAB, which includes extra capabilities such predicting the type of tumor using real time dataset.

REFERENCES

[1] GunarathneW.H.S.D,Perera K.D.M, Kahandawaarachchi K.A.D.C.P, “Performance Evaluation on Machine Learning Classification Techniques for Disease Classification and Forecasting through Data Analytics for Chronic Kidney Disease (CKD)”,2017 IEEE 17th International Conference on Bioinformatics and Bioengineering.

[2] S.Ramya, Dr.N.Radha, "Diagnosis of Chronic Kidney Disease Using Machine Learning Algorithms," Proc. International Journal of Innovative Research in Computer and Communication Engineering,Vol. 4, Issue 1, January 2016.

[3] S.DilliArasu and Dr.R.Thirumalaiselvi, “Review of Chronic Kidney Disease based on Data Mining Techniques”,International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 23 (2017) pp. 13498-13505

[4] L.Rubini,“Earlystage of chronic kidney disease UCI machine earning repository ,”2015. [Online]. Available :http://archive.i cs.uci.edu/ ml/datasets/Chronic Kidney Disease.

[5] S. A. Shinde and P. R. Rajeswari, “Intelligent health risk prediction systems using machine learning : a review,” IJET, vol. 7, no. 3, pp. 1019– 1023, 2018.

[6] HimanshuSharma,M A Rizvi,”Prediction of Heart Disease using Machine Learning Algorithms: A Survey”,International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169,Volume: 5 Issue: 8

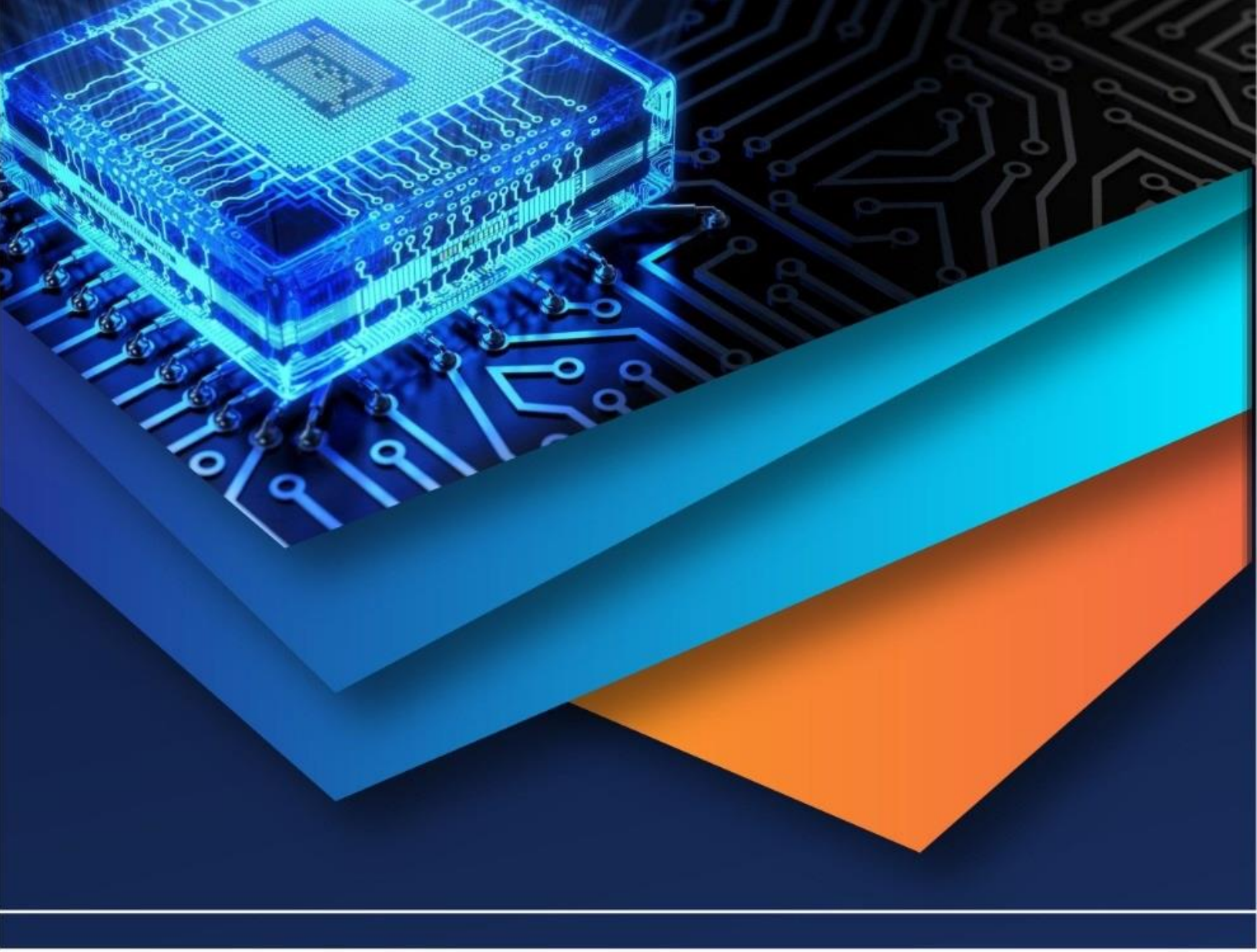
[7] Asif Salekin, John Stankovic, "Detection of Chronic Kidney Disease and Selecting Important Predictive Attributes," Proc. IEEE International Conference on Healthcare Informatics (ICHI), IEEE, Oct. 2016, doi:10.1109/ICHI.2016.36.

[8] Pinar Yildirim, "Chronic Kidney Disease Prediction on Imbalanced Data by Multilayer Perceptron: Chronic Kidney Disease Prediction," Proc. 41st IEEE International Conference on Computer Software and Applications (COMPSAC), IEEE, Jul. 2017, doi: 10.1109/COMPSAC.2017.84

[9] Sahil Sharma, Vinod Sharma, Atul Sharma, “Performance Based Evaluation of Various Machine Learning Classification Techniques for Chronic Kidney Disease Diagnosis,” July18, 2016.

[10] M. K. J. Ms.AsthaAmeta, "Data Mining Techniques for the Prediction of Kidney Diseases and Treatment: A Review," International Journal of Engineering and Computer Science, vol. 4, no. 2, p. 20376, February 2017.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)