iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.51503

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 V May 2023- Available at www.ijraset.com

Chronic Kidney Disease Prediction using Machine Learning Algorithms

Prasanna Kumar M J[1], Nethravathi K G[2]
[1, 2]*Assistant Professor, Dept of CSE, BGSIT, Adichunchanagiri University*

Abstract: ***Diabetes and high blood pressure are the primary causes of Chronic Kidney Disease (CKD). A person with CKD has a higher chance of dying young. Doctors face a difficult task in diagnosing the different diseases linked to CKD at an early stage to prevent the disease. Early discovery of CKD empowers sufferers to get the opportunity remedy to decorate the motion of this infection. CKD is among the top 20 causes of death worldwide and affects approximately 10% of the world's adult population. CKD is a disorder that disrupts normal kidney function. The novelty of this study lies in developing a diagnosis system to detect chronic kidney diseases. This study focused on evaluating a dataset collected from 400 patients containing 24 features. The mean and mode statistical analysis methods were used to replace the missing numerical and nominal values. To choose the most important features, Recursive Feature Elimination (RFE) was applied. Three classification algorithms applied in this study were k-nearest neighbors (KNN), Random Forest Classifier (RFC), and Ada Boost Classifier (ABC). All the classification algorithms achieved promising performance. The RFC and ABC Algorithm outperformed all other applied algorithms, reaching an accuracy, precision, recall, and F1-score of 100% for all measures. Therefore, Machine Learning techniques are of great importance in the early detection of CKD. These techniques are supportive of experts and doctors in early diagnosis to avoid developing kidney failure.***

I. INTRODUCTION

Chronic Kidney Disease (CKD) is a condition in which the kidneys are damaged and cannot filter blood as well as they should. Because of this, excess fluid and waste from blood remain in the body and may cause other health problems, such as heart disease and stroke. It is rapidly expanding and becoming one of the major causes of death all over the world. The research from 1990 to 2013 specifies that life loss caused globally by CKD increased by 90% per year and it is the 13th leading cause of death in the world. Worldwide approximately 850 million people are likely to have kidney diseases from different factors. According to the report of world kidney day of 2019, at least 2.4 million people die every year due to kidney-related disease. Currently, it is the 6th fastest-growing disease to cause death worldwide and becoming a challenging public health problem with increasing prevalence worldwide. It is even higher in low-income countries where detection, prevention, and treatment remain low. Mainly in Ethiopia, this is affecting hundreds of thousands of people irrespective of age, sex, pure water, starvation, and physical activities are believed to have contributed. Additionally, communities living in rural areas have less knowledge about CKD. According to the WHO report of 2017, the number of deaths in Ethiopia due to kidney disease was 4,875. It is 0.77% of total deaths that has ranked the country 138th in the world. The age-adjusted death rate is 8.46 per 100,000 of the population and the death rate increased to 12.70 per 100,000 that has ranked the country 109 in 2018. National kidney foundation classifies stages of CKD into five based on the abnormal kidney function and reduced Glomerular Filtration Rate (GFR), which measures a level of kidney function, The mildest stage (stage 1 and stage 2) is known with only a few symptoms and stage 5 is considered as end-stage or kidney failure. The Renal Replacement Therapy (RRT) cost for total kidney failure is very expensive. The treatment is not also available in most developing countries like Ethiopia. As a result, the management of kidney failure and its complications is very difficult in developing countries due to the shortage of facilities, physicians, and the high cost to get the treatment. About 37 million US adults are estimated to have CKD, and most are undiagnosed. 40% of people with severely reduced kidney function (not on dialysis) are not aware of having CKD. Every 24 hours, 360 people begin dialysis treatment for kidney failure. In the United States, diabetes and high blood pressure are the leading causes of kidney failure, accounting for 3 out of 4 new cases. In 2019, treating Medicare beneficiaries with CKD cost $87.2 billion, and treating people with ESRD cost an additional $37.3 billion. Hence, early detection of CKD is very essential to minimize the economic burden and maximize the effectiveness of treatments. Predictive analysis using machine learning techniques can be helpful through early detection of CKD for efficient and timely interventions. In this study, Random Forest (RF), (KNN) and Ada Boost classifier (ABC) have been used to detect CKD. Most of the previous research focused on two classes, which makes treatment recommendations difficult because the type of treatment to be given is based on the severity of CKD.

©IJRASET: All Rights are Reserved

II. LITERATURE SURVEY

Ajith Kumar, C. Hari Haran, D. Manu Vignesh has proposed in this paper that Chronic kidney disease (CKD) is an international fitness hassle with excessive morbidity and mortality rate, and it induces different diseases. Since there aren't any conspicuous aspect consequences for the duration of the start levels of CKD, sufferers frequently forget about to look the illness. Early discovery of CKD empowers sufferers to get an opportune remedy to decorate the motion of this infection. Machine getting to know fashions can efficiently assist clinicians accomplish this goal due to their short and specific acknowledgment execution. In this assessment, we advise a KNN and Logistic regression, Decision tree, Random forest, and machine for diagnosing CKD.

Vijendra Singh, Vijayan K. Asari, and Rajkumar Rajasekarana have proposed in this paper Diabetes and high blood pressure are the primary causes of Chronic Kidney Disease (CKD). Glomerular Filtration Rate (GFR) and kidney damage markers are used by researchers around the world to identify CKD as a condition that leads to reduced renal function over time. A person with CKD has a higher chance of dying young. Doctors face a difficult task in diagnosing the different diseases linked to CKD at an early stage to prevent the disease. This research presents a novel deep-learning model for the early detection and prediction of CKD. This research objective is to create a deep neural network and compare its performance to that of other contemporary machine learning techniques. In tests, the average of the associated features was used to replace all missing values in the database. After that, the neural network's optimum parameters were fixed by establishing the parameters and running multiple trials. The foremost important features were selected by Recursive Feature Elimination (RFE). Hemoglobin, Specific Gravity, Serum Creatinine, Red Blood Cell Count, Albumin, Packed Cell Volume, and Hypertension were found as key features in the RFE. Selected features were passed to machine learning models for classification purposes. The proposed Deep neural model outperformed the other four classifiers (Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Logistic regression, Random Forest, and Naive Bayes classifier) by achieving 100% accuracy. The proposed approach could be a useful tool for nephrologists in detecting CKD.

Bin Zhang, MD, Chun-song Cheng, Ph.D., Min-gang Ye, MD, Cheng-zheng Han, MD, Dai-yin Peng, Ph.D. has proposed in this paper Wuqinxi is a traditional medicinal exercise that is widely practiced in China now. Because of its obvious medical rehabilitation, Wuqinxi has been used in physical education for more than 1.2 million people in at least 24 Chinese Medicine university campuses in China for many years. This investigation aimed to evaluate whether Wuqinxi has a positive effect on physical improvements for female college students. Infrared scanners were used for real-time monitoring of body calorie dynamics; the electromyography (EMG) was used to detect the iEMG on biceps, brachioradialis, quadriceps, and gastrocnemius; besides, the physical health elements, heart rate, and cardiopulmonary function were also taken within the scope of our investigation and records. Wuqinxi exercise can improve body function by making the abdominal muscles, back muscles, and limbs strength exercise more effective; Wuqinxi exercise had also made the athletes better control their muscles to have a good way to contraction and keep balance; Moreover, the performance of speed of 800 m run, setting flexion, set-ups, and grip strength had a comprehensive promotion for each of the participants including long-time practitioners and short-term practitioners. Therefore, the new gymnastics derived from ancient Chinese Wuqinxi exercise can improve the physical health of female college students so that it can be used as part of the development of higher education's health quality in the future

Ebrahime Mohammed Senan, Mosleh Hmoud Al-Adhaileh, and more have proposed in this paper that Chronic kidney disease (CKD) is among the top 20 causes of death worldwide and affects approximately 10% of the world's adult population. CKD is a disorder that disrupts normal kidney function. Due to the increasing number of people with CKD, effective prediction measures for the early diagnosis of CKD are required. The novelty of this study lies in developing a diagnosis system to detect chronic kidney diseases. This study assists experts in exploring preventive measures for CKD through early diagnosis using machine learning techniques. This study focused on evaluating a dataset collected from 400 patients containing 24 features. The mean and mode statistical analysis methods were used to replace the missing numerical and nominal values. To choose the most important features, Recursive Feature Elimination (RFE) was applied. Four classification algorithms applied in this study were support vector machine (SVM), k-nearest neighbors (KNN), decision tree, and random forest. All the classification algorithms achieved promising performance. The random forest algorithm outperformed all other applied algorithms, reaching an accuracy, precision, recall, and F1-score of 100% for all measures. CKD is a serious life-threatening disease, with high rates of morbidity and mortality. Therefore, artificial intelligence techniques are of great importance in the early detection of CKD. These techniques are supportive of experts and doctors in early diagnosis to avoid developing kidney failure

Gabriel R. Vásquez-Morales, Sergio M. Martínez-Monterrubio, Pablo Moreno-Ger, and Juan A. Recio-García have proposed this paper This paper presents a neural network-based classifier to predict whether a person is at risk of developing chronic kidney disease (CKD). The model is trained with the demographic data and medical care information of two population groups: on the one hand, people diagnosed with CKD in Colombia in 2018, and on the other, a sample of people without a diagnosis of this disease.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Once the model is trained and evaluation metrics for classification algorithms are applied, the model achieves 95% accuracy in the test data set, making its application for disease prognosis feasible. However, despite the demonstrated efficiency of the neural networks to predict CKD, this machine-learning paradigm is opaque to the expert regarding the explanation of the outcome. Current research on explainable AI proposes the use of twin systems, where a black-box machine-learning method is complemented by another white-box method that provides explanations about the predicted values. Case-Based Reasoning (CBR) has proved to be an ideal complement as this paradigm can find explanatory cases for an explanation-by-example justification of a neural network's prediction. In this paper, we apply and validate an NN-CBR twin system for the explanation of CKD predictions. As a result of this research, 3,494,516 people were identified as being at risk of developing CKD in Colombia, or 7% of the total population

Ramesh Chandra Poonia, Mukesh Kumar Gupta, and more. Has proposed in this paper Kidney disease is a major public health concern that has only recently emerged. Toxins are removed from the body by the kidneys through urine. In the early stages of the condition, the patient has no problems, but recovery is difficult in the later stages. Doctors must be able to recognize this condition early to save the lives of their patients. To detect this illness early on, researchers have used a variety of methods. Prediction analysis based on machine learning is more accurate than other methodologies. This research can help us to better understand global disparities in kidney disease, as well as what we can do to address them and coordinate our efforts to achieve global kidney health equity. This study provides an excellent feature-based prediction model for detecting kidney disease. Various machine learning algorithms, including k-nearest neighbors algorithm (KNN), artificial neural networks (ANN), support vector machines (SVM), Naive Bayes (NB), and others, as well as Re-cursive Feature Elimination (RFE) and Chi-Square test feature-selection techniques, were used to build and analyze various prediction models on a publicly available dataset of healthy and kidney disease patients. The studies found that a logistic regression-based prediction model with optimal features chosen using the Chi-Square technique had the highest accuracy of 98.75 percent. White Blood Cell Count (Wbcc), Blood Glucose Random (bgr), Blood Urea (Bu), Serum Creatinine (Sc), Packed Cell Volume (Pcv), Albumin (Al), Hemoglobin (Hemo), Age, Sugar (Su), Hypertension (Htn), Diabetes Mellitus (Dm), and Blood Pressure (Bp) are examples of these traits.

III. PROPOSED SYSTEM

The chronic kidney disease is identified using the KNN Classification, Random Forest, and Ada Boost Classifier, all these algorithms are used to execute the high result with high accuracy where the random forest and Ada Boost Classifier provide the highest accuracy, precision, and f-measure. They Can be analyzed and improved through the given input dataset of the CKD. Where using the parameters the dataset of the ckd or no-ckd can be analyzed will be the result.

A. Data Processing

Each specific (nominal) variable is changed into code to facilitate the processing in a computer. For the values of RBC (red blood cells) and pc (p-cresyl), everyday and strange have been coded as 1 and zero, respectively. For the values of % and Ba (barium), gift and now no longer gift have been coded as 1 and zero, respectively. For the values of the (hypertension), dm, cad (coronary artery disease), pe (pulmonary embolism), and ane (acute necrotizing sncephalopathy), sure and none have been coded as 1 and zero, respectively. For the price of an applet, proper and negative have been coded as 1 and zero, respectively. Although the unique statistics description denes 3 variables sg (specific gravity), al (aluminum), and su (sulphur) as specific types, the values of those 3 variables are nevertheless numeric primarily based totally, therefore those variables have been dealt with as numeric variables. All the explicit variables have been converted into factors. Each pattern changed into given an impartial range that ranged from 1 to four hundred. There is a massive range of lacking values withinside the statistics set, and the range of whole times is 158. In general, the sufferers may pass over a few measurements for numerous motives earlier than creating a prognosis. Thus, lacking values will seem withinside the statistics whilst the diagnostic classes of samples are unknown, and a corresponding imputation approach is needed.

B. K-NN Classification

In sample recognition, the K-Nearest Neighbor set of rules (K-NN) is a non-parametric approach used for type and regression. In each case, the entry includes the K closest education examples withinside the characteristic space. KNN is a kind of instance-primarily based studying. In K-NN Classification, the output is a category membership. Classification is achieved through a majority vote of neighbors. If K = 1, then the magnificence is the unmarried nearest neighbor. In a not unusual place weighting scheme, a man or woman neighbor is assigned to a weight of 1/d if d is the gap to the neighbor. The shortest distance among any neighbors is constantly an immediate line and the gap is referred to as Euclidean distance. The hassle of the K-NN set of rules is its touchy with the nearby configuration of the statistics.

The technique of remodeling the enter statistics to a fixed of capabilities is referred to as Feature extraction. The chronic Kidney Disease dataset is taken from the UCI database which includes 25 variables with four hundred times. In that, we've got non-stop, nominal, and binary variables. Hence nominal variables attributes which include unique gravity, albumin, and sugar are taken. We convert all of the nominal variables to binary and we use KNN type okay values are chosen. In the education segment, a KNN set of rules is carried out, and withinside the take a look at segment outcomes are displayed. The most important aim of this algorithm is to assess the overall performance of ten distance formulae whilst KNN is used for binary statistics and additionally to discover the quality price of okay. Here, we assign okay values starting from one hundred seventy-five to one hundred ninety and discover the ensuing blunders rates.

C. Random Forest Classifier

Random Forest (RF), proposed for persistent kidney ailment type is a fast, exceedingly correct, noise-resistant type approach. Bagging and random characteristic choice is mixed. Every tree withinside the wooded area is encouraged through the values of random vectors sampled one after the other and has equal distribution as every other tree withinside the wooded area. RF includes an oversized range of selection timber wherein selection tree pick their keeping apart capabilities from bootstrap education set Si wherein I constitute ith inner node. Trees in RF are grown through Classification and Regression Tree (CART) approach and not using pruning. As a range of timber withinside the wooded area will become an oversized range, generalization blunders will even grow till it converges to a few boundary levels. We hired distinctive device studying strategies for the prognosis of persistent kidney ailment (CKD). Results represented with Confusion Matrix. Precision, F-degree, and Overall Classification Accuracy outcomes are shown. As it could be visible without problems from those tables, the best performances have been acquired through the usage of a random wooded area (RF) classifier. The overall accuracy of RF is quality evaluate with ordinary different sets of rules. The decision tree classifier additionally resulted next. K-NN with putting parameter c identical to a hundred and using normalized poly kernel has performed higher styles changed into successfully labeled. LR classifier resulted in an ordinary common accuracy rate

D. Ada Boost Classifier

It combines multiple classifiers to increase the accuracy of classifiers. AdaBoost is an iterative ensemble method. AdaBoost classifier builds a strong classifier by combining multiple poorly performing classifiers to get a high-accuracy strong classifier. The basic concept behind Adaboost is to set the weights of classifiers and train the data sample in each iteration to ensure accurate predictions of unusual observations. Any machine learning algorithm can be used as a base classifier if it accepts weights on the training set.

IV. RESULT & DISCUSSION

The Proposed model accuracy was calculated by making the CKD class value positive and the not-CKD class value negative. The confusion matrix was utilized to evaluate the performance by using True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). According to TP, CKD samples have been accurately categorized. The findings of the FN test show that CKD samples were misclassified. The not-CKD samples were not accurately identified, as indicated by a false-positive result (FP). True negative (TN) samples have been accurately categorized as not CKD. The dataset is randomly divided into 75% for training and 25% for testing and validation.

Accuracy: It refers to the proportion of correct guesses to total predictions. Accuracy can be described as the ability to accurately predict the outcome of a situation.

1) *Accuracy:* It refers to the proportion of correct guesses to total predictions. Accuracy can be described as the ability to accurately predict the outcome of a situation.

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

2) *Recall:* The recall calculates the proportion of accurately predicted positive observations to the total number of observations in the class, as shown in the following equation.

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

3) *Precision:* As stated in the equation below, this metric represents the proportion of accurately predicted positive observations to total predictive positive observations.

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

4) *F-Measure:* Precision and Recall are weighted averaged in the F-measure. False positives and false negatives are part of the process. F-measure is a term that is defined as

F-Measure = Two × (Precision × Recall) /(Precision + Recall) (9)

The F-Measure values lie from 0 to 1

Classifier	Precision	Accuracy	Recall	F1-Score
KNN	0.96	0.96	0.92	0.96
RFC	1.00	1.00	1.00	1.00
ABC	1.00	1.00	1.00	1.00

The Model Comparison is made between the respective opted three algorithms, to identify which algorithm gives the 100% Accuracy. Where the accuracy in the existing systems reached a ratio between 95.84% and 66.3%, while the accuracy of our systems reached between 100% by random forest & ABC and 96.0% by KNN

	Model	Score
1	Random Forest Classifier	1.00
2	Ada Boost Classifier	1.00
0	KNN	0.96

V. CONCLUSION & FUTURE WORK

Early prediction is very crucial for both experts and patients to prevent and slow down the progress of chronic kidney disease to kidney failure. In this study, three Machine Learning models KNN, RFC, and ABC were used to build proposed models. First, the three machine learning algorithms were applied to original datasets with all 24 features. Applying the original dataset's models, we have the highest accuracy with KNN, RFC & ABC. The accuracy was 100% for the binary class and 96% for the five-class. KNN produced the lowest performance compared to RFC. RFC & ABC produced the highest f1_score values i.e., 100% Accuracy. Hence we believe that multi-classification work was very important to know the stages of the disease and suggest needed treatments for the patients to save their lives.

This study used a supervised machine-learning algorithm, and feature selection methods to select the best subset features to develop the models. A limitation of the proposed model was that it had been tested on small data sets. It is better to see the difference in performance results using unsupervised or deep learning algorithms models and the model should also be tested with a large dataset. The proposed model supports the experts to give the fast decision, it is better to make it a mobile-based system that enables the experts to follow the status of the patients and help the patients to use the system to know their status.

REFERENCES

[1] https://doi.org/10.3390/diagnostics12010116.
[2] AJITH KUMAR C. HARI HARAN, D. MANU VIGNESH "Chronic Kidney Disease Prediction using Random Forest Algorithm in Machine Learning" International Journal of Innovative Science and Research Technology Volume 7, Issue 2, February – 2022, ISSN No:- 2456-2165.
[3] Bin Zhang, MD, Chun-song Cheng, Ph.D., Min-gang Ye, MD, Cheng-zheng Han, MD, Dai-yin Peng, Ph.D. "A preliminary study of the effects of medical exercise Wuqinxi on indicators of skin temperature, muscle coordination, and physical quality" http://www.mdjournal.com/ Zhang et al. Medicine (2018) 97:34.
[4] Ebrahime Mohammed Senan, Mosleh Hmoud Al-Adhaileh, Fawaz Waselallah Alsaade, Theyazn H. H. Aldhyani, Ahmed Abdullah Alqarni, Nizar Alsharif, M. Irfan Uddin, Ahmed H. Alahmadi, Mukti E Jadhav and Mohammed Y. Alzahrani "Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques" Journal of Healthcare Engineering Volume 2021, Article ID 1004767, 10 pages
[5] Gabriel R. Vásquez-morales1, Sergio M. Martínez-Monterrubio, Pablo Moreno-Ger and juan A. Recio-García "Explainable Prediction of Chronic Renal Disease in the Colombian Population Using Neural Networks and Case-Based Reasoning" EEE Access · October 2019.
[6] Ramesh Chandra Poonia, Mukesh Kumar Gupta, Ibrahim Abunadi, Amani Abdulrahman Albraikan, Fahd N. Al-Wesabi, Manar Ahmed Hamza and Tulasi B "Intelligent Diagnostic Prediction and Classification Models for Detection of Kidney Disease" Healthcare 2022, 10, 371.

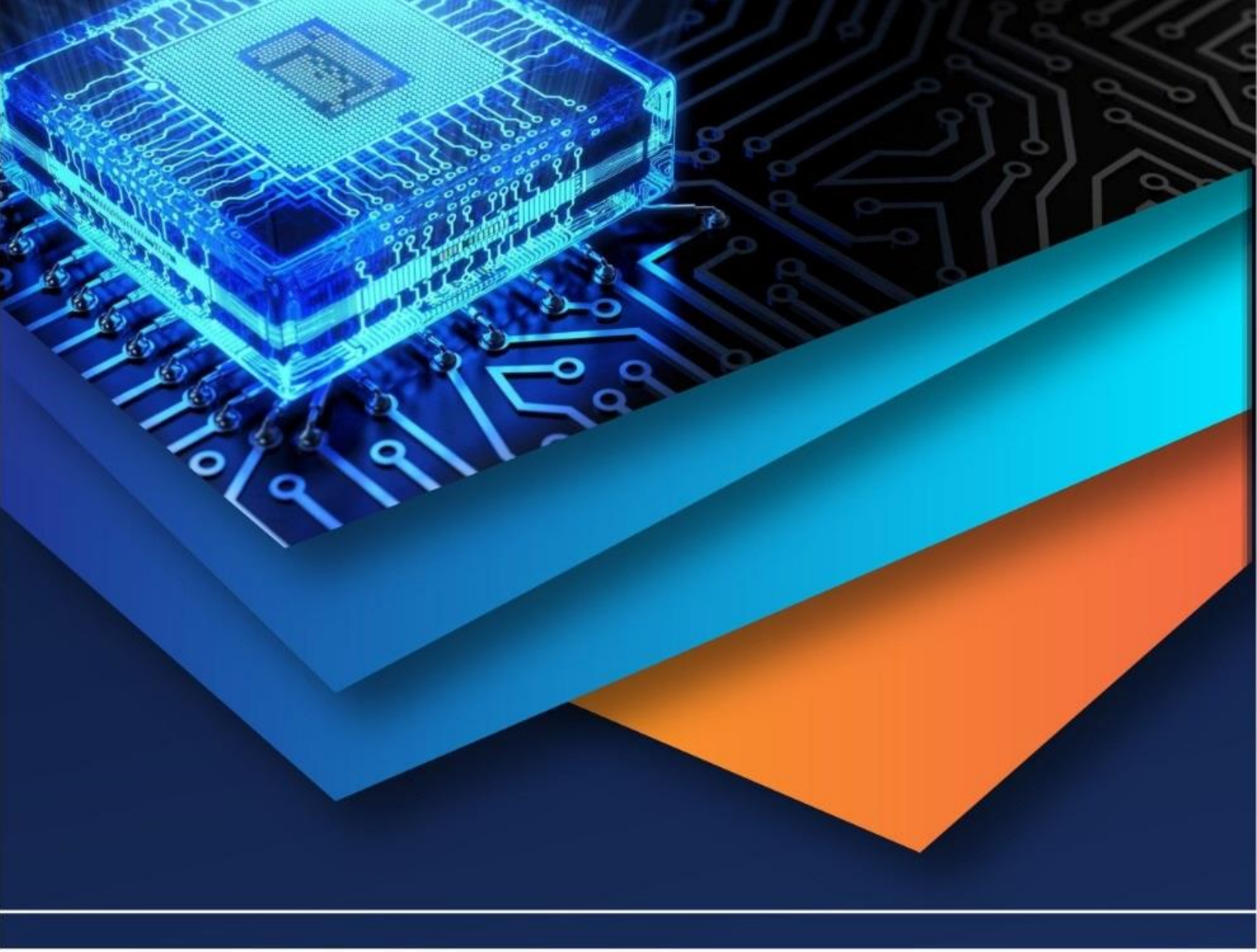
[7] Randal K. Detwiler, Emily H. Chang, Melissa C. Caughey, Melrose W. Fisher, Timothy C. Nichols, Elizabeth P. Merricks, Robin A. Raymer, Margaret Whitford, Dwight A. Bellinger, Lauren E. Wimsey, Caterina M. Gallippi [Member, IEEE] "Mechanical Anisotropy Assessment in Kidney Cortex Using ARFI Peak Displacement: Preclinical Validation and Pilot In Vivo Clinical Results in Kidney Allografts" IEEE Trans Ultrason Ferroelectr Freq Control. 2019 March ; 66(3): 551–562. doi:10.1109/ TUFFC.2018.2865203

[8] Dibaba Adeba Debal and Tilahun Melak Sitote "Chronic kidney disease prediction using machine learning techniques" Debal and Sitote Journal of Big Data (2022) 9:109.

[9] John W Stanifer, Bocheng Jing, Scott Tolan, Nicole Helmke, Romita Mukerjee, Saraladevi Naicker, Uptal Patel "The epidemiology of chronic kidney disease in sub-Saharan Africa: a systematic review and meta-analysis" Lancet Glob Health 2014; 2: e174–181.

[10] Diego Buenaño-Fernández, David Gil and Sergio Luján-Mora "Application of Machine Learning in Predicting Performance for Computer Engineering Students: A Case Study" Sustainability · May 2019

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



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)