iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VI **Month of publication:** June 2022

DOI: https://doi.org/10.22214/ijraset.2022.44821

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 10 Issue VI June 2022- Available at www.ijraset.com

Stagewise Classification of Chronic Kidney Disease and Diet Plan Recommendation

Dr. Ayesha Taranum[1], Ms. Pragathi G[2], Ms. Rachana DM[3], Ms. Fathima Saireen[4]
[1]*Assitant Professor, Department of ISE, GSSSIETW, Mysuru, India*
[2, 3, 4]*Department of ISE, GSSSIETW, Mysuru, India*

Abstract: ***Chronic kidney disease means gradual loss in functioning of kidney. Undiagnosed Chronic Kidney Disease can be identified, predicting the patient condition and its prevention with machine Learning Techniques. Accurate analytical models can be formed by medical practices, which decreases consequences hence gradually progresses specifications. To estimate Chronic Kidney Disease and to identify the stages in it. Dietary plan will be recommended based on the stages of Chronic Kidney Disease. There are no symptoms of Chronic Kidney Disease at earlier stages and if the disease remains unrectified then it can lead to kidney failure. Manual examination of Chronic Kidney Disease is a possible solution however it is dependent on human errors. Detecting Chronic Kidney Disease at early stage can help doctor to give better medication and patient may recovers soon.***
Keywords: Chronic kidney disease, Stage classification, Naïve Bayes, Principal component analysis, Random forest

I. INTRODUCTION

The health service sector is producing mass information that should be extracted to discover stowed away data for viable expectation, determination, and direction.

Presently, kidney sickness has been a urgent issue. It is one of the main sources of death in India. Constant kidney sickness can be defined as the steady reduction in the functioning of kidney. Kidneys channel squanders and overabundance liquids from your blood, which are then discharged in urine. On the off chance that this illness deteriorates, squanders can amass in the blood and can cause troubles like hypertension, paleness, debilitating of bones, poor nourishing wellbeing, and nerve harm. Additionally, kidney illness builds the gamble of having heart and vein infection. Global Burden of Disease positions ongoing kidney illness as the 8th driving cause for death in India in 2018.

The introduced procedure helps in the expectation of quick and precise examination of the presence of the persistent kidney sickness. As we realize that it is one of the most life threating infection, we really want to anticipate the illness in the beginning phase so we can treat the patient by suggesting an appropriate dietary arrangement. Physically these things consume parcel of time for the specialist to dissect the sickness. we can expect the ailment taking into account the twenty-four parameters that are open in the UCI repository. Naive bayes is applied to estimate whether the patient is undergoing the chronic kidney disorder or not.This algorithm is fast and accurate when compared to other classification algorithm as it considers both discreate and continues values of dataset. For dimensionality reduction Principal component analysis is used. Glomerular filtration rate is a mathematical formula used to check the functioning of kidney. The numerical recipe decides the kidney work. Random Forest Algorithm is utilized to suggest the appropriate dietary arrangement for the patient.

II. LITERATURE SURVEY

In this paper, they have involved Machine Learning methods in expectation of Chronic Kidney Disease [1]. The means utilized are Data Collection, Data Pre-handling, Adding new characteristics, Prediction module, Diet proposal module. They have added two extra means which are Finding Missing Values and Data Transformation steps to find the missing characteristics in dataset and to change the characteristics into required structure independently. They have grouped the sickness in 5 phases, stage 1 to 5. Stage 1 is protected and involves an indulgent eating regimen intend to be observed. In stage 2, a probable CKD patient is provided controlled and extreme eating routine. Maintaining the equilibrium of nutrients, electrolytes, and fluids in the interior body is hard for stage 3 - 5 patient. Consequently, patients must take proper dietary direction. A healthy eating routine for a renal enhancement and avoid further damage is the key, which also benefits the balance of electrolytes and water inside the body which are liable for chronic kidney disease then machine learning process can systematize the sorting of the chronic kidney disease in diverse stages according to the severity.

In this research, they have estimated the test of renal diseases using the techniques of data mining which includes Support Vector Machine and Naive Bayes [2]. This kind of desk work especially designated on distinguishing the ideal grouping procedure in light of the characterization assurance and choking time point from the precise result. It has been viewed SVM achievements contributing more than the technique of Naive Bayes. Data mining in the medical field goes to the most recent thing in our precisely creating world for the finish of the disorder. By means of data mining moves, backpropagation and Random forest area estimation proposed system have broken down the sickness together with contemplated which the backpropagation computation is much more proper for examination of the renal disillusionment compared to the sporadic forest estimation. The attributes which proposed system have consider for the end are circulatory strain, potassium, irregularity of glucose in blood, sugar, hemoglobin, sodium, percentage of white platelet and red platelet, age, Serum Creatinine, squeezed cell size, release tissue packs, blood pressure, Heart Disease, Diabetes Mellitus, Pedal Edema, Appetite, Anemia.

In this research, the work arranged here utilizes three grouping procedures to foresee the existence of persistent problem of renal in people. The attributes used for classification utilizes KNN, Naive Bayes and Random Forest for classification. Dataset for persistent illness was accumulated and activated on each attribute of classification to forecasting illness [3] and accordingly, the presentation of attributes of classification is inspected considering their exactness, accuracy together with F measure. Working of plan is according to the following: The CKD patients' dataset is gathered as well as taken care into the Naive Bayes, Random Forest and the KNN. The expectation of Chronic Kidney Disease is finished within calculations acted in the language of C Sharp. The informational index, which is driven by the UCI AI archive, on account of contribution indeed for forecast. This dataset comprises of qualities along with values.

In this paper, their principal center was to dissect the datasets of Chronic kidney sickness. They have involved Machine Learning strategies in assumption for Chronic Kidney Disease. The means utilized are Data Collection, Support Vector Machine, Classifier's Performance Measures, Simulated Program [4]. In absolute it comprises four hundred cases, among which two hundred and fifty of the cases are patients suffering from CDK and the rest one hundred and fifty are not. The variable of objective shows regardless of patient suffering from CDK. Among twenty-five ascribes, twenty-four are elements of clinical and rest is an objective quality. The components which are confined into 3 areas medical record, genuine appraisal, and lab test verification. SVM results for a best hyperplane which separates between classes. That resulted best hyperplane is seen as the one which shrubbery the best edge among the 2 of the specific classes. The trial concentrates fundamentally have been finished by utilizing the Python along with Scikit-learn AI system have utilized in this review which contains countless calculations for information preprocessing, incorporate decision, portrayal, clustering, and finding the helpful standards.

In this paper, they have involved Deep Learning strategies in assumption for Chronic Kidney Disease. The means used are Sensing Module, Sample Collection and Testing, and Deep Learning Module [5]. The samples for this study are gathered from one hundred and two members, including forty sound workers and sixty-two people who are suffering from kidney sickness. Motivation behind the exploration task is to educate the members prior the investigation, composed assent are given to the members. Intended for the testing, one ml of unstimulated salivation is collected from each member in a test tube. The salivation tests are collected utilizing strategy of spitting along with the assistance of clinical staff. A profound knowledge CNN-SVM calculation is further carried out for processing and characterizing elements consequently from the result. The CNN calculation is a notable profound approach of knowledge which is regularly utilized in handling 2 Dimensional signs. They have executed the proposed method by varying the engineering of the traditional Convolution Neural Network organization to suite to 1 Dimensional sign.

In this paper they have applied six algorithms Navies bayes, Random forest, linear regression model, Decision strum, Simple logistic regression, simple linear regression for predicting the chronic kidney disease and gamble factors [6]. Datamining techniques uses the datasets with huge data strategies to diminish the time expected to foresee the sickness and it will assist with come by the specific outcome in view of the help of the mining arrangement method. Innocent Bayes calculation, Simple strategic relapse, Random backwoods are utilizing to anticipate the exactness of the model and straight relapse model, Decision stump, Simple Linear relapse model reutilizing to figure out most huge and non-critical gamble variable of Chronic Kidney Disease.

In this review paper, to predict the chronic kidney disease they have considered the glomerular filtration rate which varies from person to person based on the gender and also have considered only these parameters which are useful in foreseeing the presence of chronic kidney illness, for example, serum creatinine, albumin creatinine rate, cystatin are utilized to ascertain the Glomerular filtration rate which is easier and gives accurate results compared to other approaches such as radiolabeled tracer clearances which are intrusive, might include radiation, and need a few hours to perform [7].

The early expectation of Chronic Kidney Disease is a difficult undertaking for analysts and one which can be conceivable by carrying out cutting edge specialized devices, for example, AI, man-made consciousness methods, data mining, and so forth, some of which have as of now been carried out around here of Chronic Kidney disease. The creators proposed novel specialized apparatuses, for example, decision trees, strategic relapse, Naive Bayes, ANN, and data mining instruments which are used to create a robotized demonstrative framework which works on the extensive cycle in medical services. The action behind the motorized insightful structure is to deal with the data from the system informational collection and give early perceptive electronic results which is more careful appeared differently in relation to the conventional symptomatic framework. It is noticed that these specialized instruments will assume an essential part in the anticipation of chronic kidney disease.

This paper means to analyze and understand a piece of the systems that are used in expecting kidney affliction. Clinical industry mining information is used to recognize the educational records in kidney contamination using the MATLAB device [8]. The variables that are influencing the constant kidney sickness are hypertension, renal disappointment, cirrhosis, atherosclerosis, diabetes. It includes six unique classifier to be specific KNN, SVM, arbitrary trees, J48, and three social affair computation explicitly firing, adaboost, Random subspace were used in the survey. The outcome says that J48 premise calculation for use with random subspace and bagging ensemble calculations and random tree premise calculation for use with bagging ensemble calculation has given 100 percent arrangement victory. Data mining helps in removing stowed away data from large dataset. It similarly helps in arranging data, support them and induce excellent models in them. A few information mining strategies like arrangement, bunching, relapse, affiliation examination etc. The target of this paper is for the prediction of Chronic Kidney Disease which is compared with other characterization strategies like ANN and Naives bayes..

In this research paper to comprehend the machine leraning strategies to foresee kidney stones. They anticipated great precision in addition to C4.5, Classification tree and Random forest followed by Support Vector Machines (SVM) Logistic and KNN has likewise shown great exactness results with zero relative outright blunder and 100 percent accurately arranged outcomes [9]. ROC and Calibration bends utilizing Naive Bayes has likewise been developed for anticipating precision of the information. The Neural network algorithm along with other datamining techniques showed high potential in effective kidney sickness. This exploration depicts a proficient Diagnosis of Kidney Images Using Association Rules. Their methodology is separated into four major advances: pre-handling, include extraction and determination, affiliation rule age, and age of analysis ideas from classifier. This paper dissected that the most normally utilized Data Mining strategy like Decision Trees, ANN and Naïve Bayes, Logistic Regression, Genetic Algorithms (GAs) coming about also performing on clinical data sets. Additionally shows that DTs, ANNs and Naive Bayes are well performing calculations utilized for Kidney disease.

AUTHOR	YEAR	APPROACH	DESCRIPTION
“Ebrahime Senan, Mosleh Fawaz Alsaade, Ahmed Alqarni, Nizar ,Irfan, Mukti , and Alzahrani”	2021	Data mining techniques like Support vector machine, KNN, Random Forest	This paper presents the classification algorithm for predicting the chronic kidney disease.
“Hamida, Sajid, Mahvish Ponum, Muhammad Tahir, Iftikhar and Mubasher Malik ”	2021	Machine Learning approach for disease prediction	In this research they applied decision tree algorithm, GFR to classify the stages of CKD

In this paper, the survey is the meta-analysis of Chronic Kidney Disease pervasiveness all around the world and gives an extensive outline of literature. These evaluations demonstrate that CKD may be more normal than diabetes, which has expected commonness. The revealed commonness of CKD changed generally among the examinations and had high heterogeneity [10]. Assessment of GFR from serum creatinine is the clinical standard overall and the CKD the Kidney Outcomes Quality Initiative demonstrative models based on the meta regression of CKD Prevalence and mean example populace age Responsiveness examinations decided no distinction in the predominance gauge of CKD while utilizing just top notch investigations, concentrates on that pre-owned twofold proportions of creatinine just or concentrates on that had at least two variables for the estimation of creatinine.

In this paper, they had picked the elements, Recursive Point Elimination (RFE), and four section calculations are applied in research which K-closest neighbors, support vector machine, erratic lumber along with choice tree. The Recursive Point Elimination (RFE) algorithm was used in opting attributes in training datasets which are applicable to prognosticating the target variable value along with weak features [11]. Recursive Point Elimination system is accustomed to elect the important features at chancing elevated correlation among specific features together with target similar as albumin point, packed cell volume point, serum creatinine point. This SVM algorithm works on 2 different types that are direct and nonlinear. Chronic kidney disease is the crucial sources of death worldwide which affects ten of the world adult population. They concentrated on assessing a dataset collected from four hundred cases from the California University, Irvine ML Depository containing twenty-four features distributed into the numeric features and thirteen categorical features. Seventy-five training along with twenty-five testing datasets are divided in this proposed system.

They have applied 6 algorithms K-nearest neighbours, Logistic Regression, Decision Tree Classifier Algorithm, Random Forest, Support Vector Machine Algorithm, Stochastic Gradient Descent (SGD) Classifier for predicting the chronic kidney disease. By providing the prior data, the KNN segregates the coordinates into multiple groups categorized by a selected aspect. Decision Tree Classifier Procedure gives the arrangement of values based on the individual choice the chronic kidney disease output is predicted [12]. In this paper the decision tree had been classified into two stages: Induction and Pruning. The decision tree had been classified into two stages: Induction and Pruning. J48 algorithm is used based on decision tree algorithm. Decision Tree Classifier Algorithm gives the arrangement of values based on the individual choice of the chronic kidney disease for the output. Random Forest Algorithm reduction in over-fitting and gives much more precise results than a decision tree. The paper use Support Vector Machine process is a linear model for both the regression and classification, and which gives the perfect hyper-plane among the information of two classes in the training information. They have concluded that the Logistic algorithm gives an best accurate value than compared to other algorithms.

III. COMPARISON TABLE

"Ashiqul Islam, Sagar, Sadia, Shamima, Sadia Ahmed, Shahed**"**	2020	Data mining classification techniques KNN, Naives bayes, Support vector machine	This research uses Data mining techniques it consists of huge dataset to predict the chronic kidney disease.
"Snegha,Thara,Dhivya, Charanya, Bhavani **"**	2020	Data mining approach for classifying and forecasting of chronic kidney disease	Naive Bayes algorithm is used to predict chronic kidney disease. Principle component analysis is used to get the dataset with reduced features.
"Debnath, Bhanu Doppala, Thirupathi**"**	2020	Approach of Machine Learning practices in prognostication of chronic kidney disease	Logistic Regression procedure uses the training and testing data that are used to predict the chronic kidney disease.
"Navaneeth Bhaskar and Suchetha**"**	2019	Deep Learning Technique for the Automation of Chronic Kidney Disease	They have created recent detection approch designed for analysing the urea level. The predictable enzymatic urea conversion process is held by hydrolyzing the urea to ammonia. The CNN algorithm is one of the commonly used deep learning method that is used for processing 2-D signal
Yesubabu Kakitapallia,Janakiram Ampolua,Satya Dinesh Madasub, M.L.S. Sai Kumar	2019	Data mining technique to calculate the GFR	This approach helps in identifying the risk factor based on the Glomerular filtration rate which is further used for classifying the stages.

"Devika, Vaishnavi Avilala, Subramaniyaswamy"	2019	Techniques used are k -nearest neighbor, Naive Bayes	KNN is one of the datamining classification algorithms is used here to get dataset pattern and random forest approach gives cumulative values for the dataset
"Maurya, Rahul, Sebin, Dakshayani and Jhadav"	2019	Machine Learning approach for disease prediction	Classification algorithm for predicting CKD, GFR to identify the stages, Sodium content should be lass in diet plan.
"Champa M S and Rekha "	2019	Naive Bayes classification techniques and C5.0 algorithm for the chronic kidney disease prediction	In this research, different phases are identified with the help of Naïve Bayes Classifier. Data preprocessing is executed by applying C5.0 methods which is create by the spring boot applications (Micro-service)
"Jayalakshmi V, Lipsa Nayak and K.Dharmarajan"	2018	Artificial Neural Network, Naive Bayes, SVM, , ada boost classifier, KNN	In this research, clinical industry is utilized to find the datasets related to kidney disorder with the help of MATLAB tools.Naive bayes gives accurate result compared to KNN, SVM. ANN is used for kidney dialysis
"Yedilkhan Amirgaliyev, Shahriar Shamiluulu, Azamat Serek"	2018	Artificial neural network	The supervised learning algorithms approach is used for data sorting and regression. Artificial neural network algorithms are used as supporting algorithms in prediction

IV. METHODOLOGY

Proposed system is an automation for chronic kidney disease prediction using classification techniques. Based on the stage, dietary plan is recommended to the patient. System uses testing dataset from "UCI Repository" and uses tools such as "Visual Studio" as frontend and "SQL Server" as backend to develop the application. System is a real time application and a browser-based application accessed from different location. This system is a generic application meant for different hospitals. This system consists of four module which is Admin, Doctor, Receptionist, and the Patients.

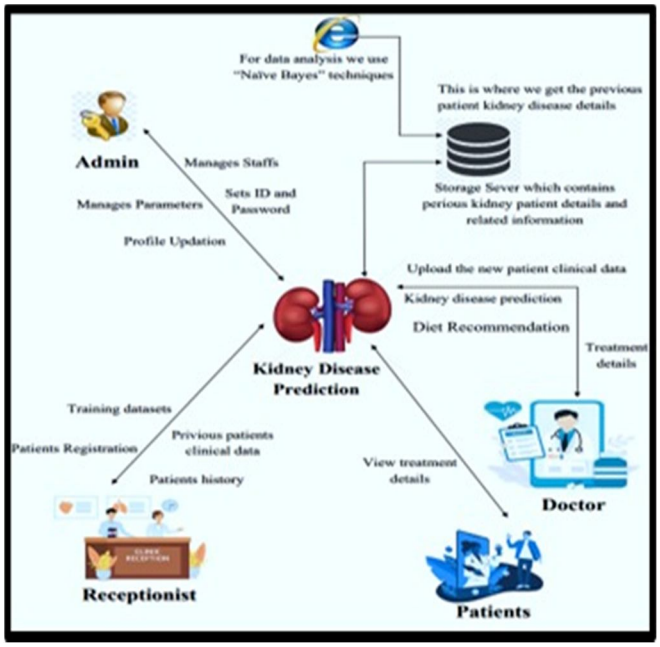


Figure 1: Methodology of Chronic Kidney Disease

A. Naive Bayes Algorithm

Proposed system is predominantly focuses on life threatening diseases using classification algorithm and implemented our project by using naive bayes algorithm. Naive bayes is a classification algorithm accustomed to divine whether patient is suffering from disease or not by using 24 parameters. This algorithm is fast and accurate when contrast to other classification algorithm. It requires training data in small amount to estimate the test data. So, the training period is comparatively less.

B. Glomerular Filtration Rate

Glomerular filtration rate is a formula used to examine how efficiently the kidneys are properly working. Specifically, it estimates the amount of blood which passes over the glomeruli apiece minute. Glomeruli are minute filters inside kidneys which filter waste of the blood. It is a mathematical formula that determines the kidney function.

GFR value = $A * (Scr/B)^{c} * (0.993)^{age}$

GFR	A	B	C
Female (Scr <= 0.7)	144	0.7	-0.329
Female (Scr > 0.7)	144	0.7	-1.209
Male (Scr <= 0.9)	141	0.9	-0.411
Male (Scr > 0.9)	141	0.9	-1.209

Figure 2: Constant values used in GFR calculation

STAGES	FUNCTIONALITY	GFR	% OF KIDNEY FUNCTION
STAGE 1	Kidney damage with normal kidney function	90 or higher	90-100
STAGE 2	Kidney damage with mild loss of kidney function	60-89	60-89
STAGE 3	Mild to severe loss of kidney function	30-59	30-59
STAGE 4	Severe loss of kidney function	15-29	15-29
STAGE 5	Kidney failure	<15	<15

Figure 3: Different Stages of Chronic Kidney Disease

C. Principal Component Analysis

Principal Component Analysis is a technique for the dimensionality reduction in Naive Bayes algorithm. Principal Component Analysis enhance the execution of the Machine Learning algorithm by means of it eradicates correlated variables that does not contribute in a part of decision making. PCA benefits in control the data overfitting distribution by reducing the number of features. PCA consequences in lofty variance and thus enhance visualization.

D. Random Forest Algorithm

It's accustomed in the recommendation of diet plan based on the stages predicted in the previous stage. Random forest assembles collaborative decision trees as well as merges them simultaneously to acquire a better accurate along with firm result. This algorithm exertion completely with together categorical as well as continuous variables in addition to it can automatically handle missing values.

V. CONCLUSION

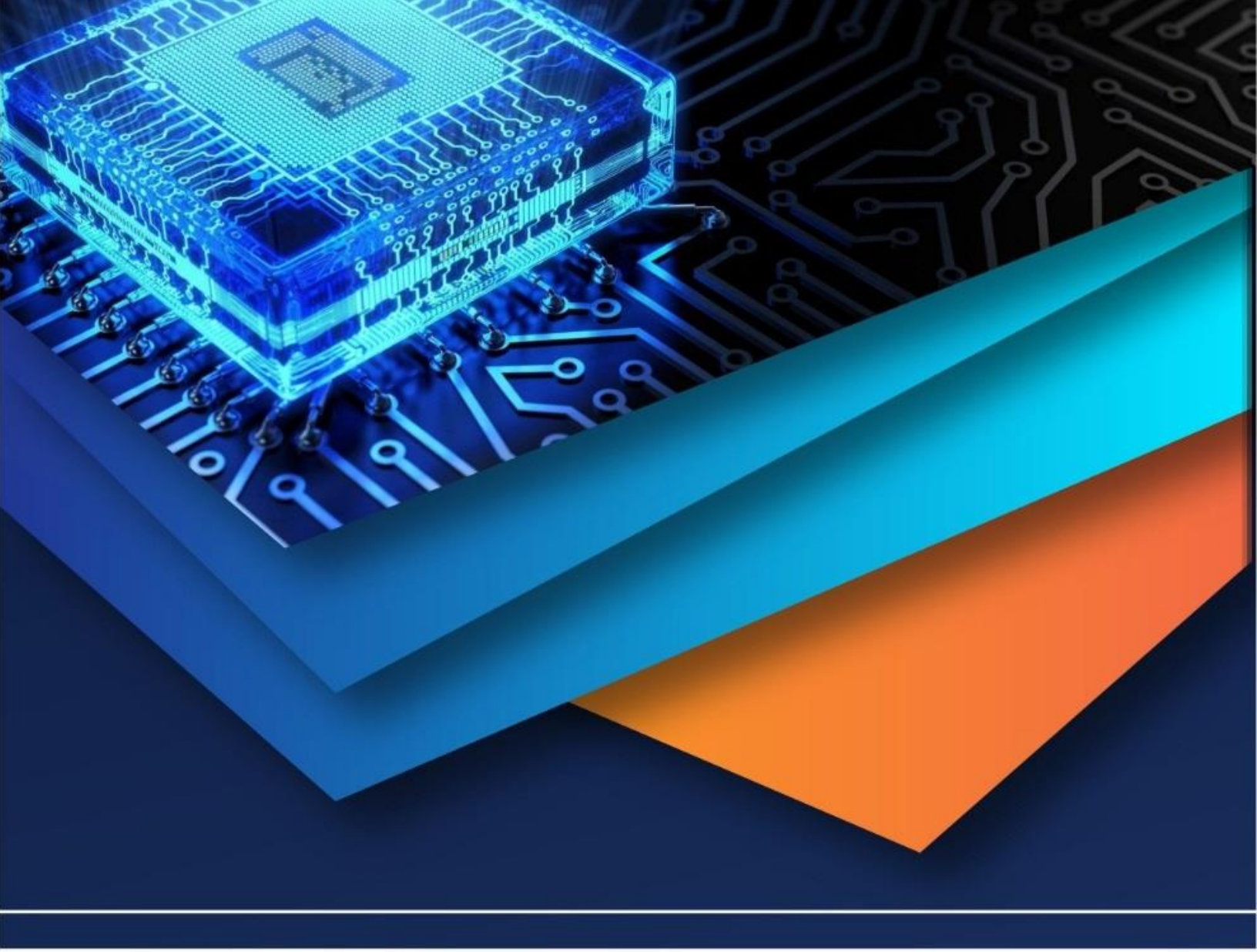
The proposed system is concerning a medicinal sector, it's an application which helps medical practitioners and doctor's in forecasting the chronic kidney disease based on the chronic kidney disease parameters. It is automation for chronic kidney disease prediction, and it classifies the disease stage and recommend the suitable diet plan based on the stage of prediction which, complications from the clinical database in a well-organized and an economically rapid approach. It is auspiciously satisfied by applying the Naïve Bayes algorithm for the technique of classification. This classification technique arises beneath data mining technology. This algorithm accepts chronic kidney disease parameters as input and predicts the disease based on earlier patient's data.

VI. FUTURE SCOPE

Chronic Kidney disease is one of the most life-threatening diseases in the world. As mentioned in the methodology all these algorithms will be implemented in the project using C# language and the proposed system is a real time application. Front end is developed using visual studio and SQL server as a backend tool. This browser-based application is used across the medical field which yields the good results. It is used by the dietician to recommend the healthy dietary plan. Kidney donors list can be added as one more functionality in patient's module which is very helpful for the patient to search for the same blood group kidney donor. This functionality has a wide range of scope in the field of research.

REFERENCES

[1] Akash, Rasika Shinde , Sebin John, Dakshayani.R, "Chronic Kidney Disease Prediction and Recommendation of Suitable Diet plan by using", 2019, International Conference on Nascent Technologies in Engineering (ICNTE 2019).
[2] J.Snegha, V.Tharani, Preetha, Charanya.R, Bhavani S, "Chronic Kidney Disease Prediction Using Data Mining", 2020 International Conference on Emerging Trends in Information.
[3] Devika , Vaishnavi, Subramaniyaswamy, "Comparative Study of Classifier for Chronic Kidney Disease prediction using Naive Bayes, KNN and Random Forest", Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC 2019), IEEE Xplore Part Number: CFP19K25-ART; ISBN: 978-1-5386-7808-4.
[4] "Yedilkhan, Shahriar, Azamat, "Analysis of Chronic Kidney Disease Dataset by Applying Machine Learning Methods", 2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT) "
[5] "Navaneeth and Suchetha, "A Deep Learning-based System for Automated Sensing of Chronic Kidney Disease", IEEE Sensors Letters (Volume: 3, Issue: 10, Oct. 2019) "
[6] "Ashiqul Islam, Sagar, Sadia Tisha, Shamima,Sadia Ahmed Keya,Shahed "Risk Factor Prediction of Chronic Kidney Disease Based on Machine Learning Algorithms" 2020, 3rd International Conference on Inteligent sustainable Systems (ICISS) "
[7] "Yesubabu, Janakiram Ampolua, Satya Madasub, Sai Kumar "Detailed Review of Chronic Kidney Disease" December 18, 2019, The Journal for Research Exchange between East and West"
[8] "Jayalakshmi, Lipsa and Dharmarajan "A Survey on Chronic Kidney Disease Detection Using Novel Methods" April 2018, International Journal of Pure and Applied Mathematics
[9] Suman, Krisha"A Literature Review on Kidney Disease Prediction using Data Mining Classification Technique" 7 July 2014, International Journal of Computer Science and Mobile Computing (IJCSMC), Vol. 3, Issue. 7, July 2014, pg.960 – 967
[10] "Nathan, Samuel, Jason Oke, Jennifer A, Christopher A. O'Callaghan, Daniel, F. D. Richard "Global Prevalence of Chronic Kidney Disease -A Systematic Review and Meta-Analysis" July 6, 2016 Applied Health Research and Care (CLAHRC)Oxford"
[11] "Ebrahime Senan, Mosleh Al-Adhaileh, Fawaz Alsaade, Theyazn Aldhyani, Ahmed Alqarni, Nizar Alsharif, M.Irfan, H.Alahmadi, Mukti, and Mohammed Alzahrani " Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques" 9 Jan 2021 Applied Hindawi Journal of Healthcare Engineering"
[12] Debnath, Bhanu Doppala, N. Thirupathi " Prediction and Forecasting of Persistent Kidney Problem Using Machine Learning Algorithms" International Journal of Current Research and Review (Volume:12, Issue 20. October 2020) "



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY