iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** VIII **Month of publication:** August 2020

DOI: https://doi.org/10.22214/ijraset.2020.30927

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: 7.429
Volume 8 Issue VIII Aug 2020- Available at www.ijraset.com

A Data Mining Approach for Kidney Stages Diagnosis Implying Machine Learning

Ms. Pooja S. Sonar[1], Prof. Kishor N. Shedge[2]
[1]*PG Student, Dept. of Computer Engineering, SVIT Nashik, Maharashtra, India*
[2]*HOD, Dept. of Computer Engineering, SVIT Nashik, Maharashtra, India*

Abstract: ***Chronic kidney disease (CKD) is identified by persistent urine abnormalities, structural abnormalities or the impaired excretory renal function suggestive of a loss of functional nephrons. The majority of patients with CKD are at risk of accelerate dcardio vascular disease and death. For those patients who progress to end-stage renal disease, the limited accessibility to renal replacement therapy is a huge problem in many parts of the world. Risk factors for the development and progression of CKD includes low nephron number at the time of birth, nephron loss occurred due to increasing age and acute orchronic kidney injuries caused by toxic exposures or diseases (for example, obesity and type 2 diabetes mellitus). The management of patients suffering from CKD is highly focused on early detection or prevention of the CKD patients and providing treatment of the underlying cause to curb progression and attention to secondary processes that contributes to on going nephron loss. Blood pressure control, inhibition of the reninangiotens in system and disease-specific interventions are the cornerstones of therapy. CKD complications such as anaemia, metabolic acidosis and secondary hyper parathyroidism briefly affects cardiovascular health and quality of life, and requires diagnosis and treatment.***
Keywords: SQL Server, eGFR, Machine learning, CKD, SVM, PNN, RBF

I. INTRODUCTION

Chronic kidney disease also known as chronic kidney failure, describes the gradual loss of kidney functioning of a person. Every individual's kidneys filter wastes and excess fluids from the person's blood, which are then excreted in the form of urine. When chronic kidney disease reaches an advanced stage or the dangerous stage, hazardous levels of fluid, electrolytes and wastes can build up in a human body. In the early stages of chronic kidney disease, a person may have few signs or symptoms indicating kidney problems. Chronic kidney disease may not become apparent and come into notice until the person's kidney function is significantly impaired. Treatment for chronic kidney disease focuses on slowing the progression of the kidney damage, usually by controlling the underlying causes leading to dysfunctioning of the kidney. Chronic kidney disease can lead to end-stage of kidney failure, which is fatal without artificial filtering i.e. dialysis or a kidney transplant.

A. Motivation

We notice that clinical choices are impacted by the stage, regardless of whether a patient is advancing, and the pace of movement. Additionally, characterizing the ailment organize is very critical as it gives a few signs that help the assurance of required mediation and medications. In this manner, information mining can assume a significant job in separating concealed information from the huge patient medicinal and clinical dataset that doctors oftentimes gather from patients to get bits of knowledge about the demonstrative data, and to execute exact treatment plans. Information mining can be characterized as the way toward removing concealed information from a huge dataset. Information mining procedures are applied and utilized broadly in different settings and fields. With information "Engaging Data Mining Algorithm for Predicting Numerous Stages of Kidney Diseases" mining systems we could anticipate, order, channel and group information. The objective or expectation attributes to the calculation preparing of a preparation set containing a lot of traits and results. Chronic Kidney Disease (CKD) may arise due to a multiplying of different insults to a renal function. However despite the wide range of pathological processes that may induce renal injury, substantial loss of nephrons provokes a common syndrome characterized clinically by system ichy pertension, proteinuria and a progressive decline in glomerular filtration rate(GFR) and pathophysiologically by progressive interstitial fibrosis, peritubular capillary loss with hypoxia and destruction of functioning nephrons because of tubular atrophy .Extensive studies suggest that the rate of loss of GFR, that is the rate of progression of CKD, may be largely due to common secondary factors ,often unrelated to the initial disease Some of these factors such as age and race are not open to intervention. The majority of them however, provides at least a potential for intervention in order to slow or halt the progression of early stage CKD.

II. RELATED WORK

We proposed an approach for the AI calculations have been utilized to anticipate and characterize in the human services have utilized the Support Vector Machine Algorithm to arrange and anticipate diabetes and pre-diabetes patients, and the outcomes show that SVM is helpful to order patients with normal illnesses. Thus, Mining have grouped Alzheimer's illness by utilizing a Support Vector Machine (SVM) to analyze entire mind anatomical attractive reverberation imaging (MRI) for a lot of patients, and the outcomes shows that SVM is a promising methodology for Alzheimer's ailment early recognition. Similarly, they have done coronary illness expectation utilizing the Probabilistic Neural Network Algorithm, Decision tree Algorithm, and Naive Bayes Algorithm, and PRNN furnishes the best outcomes contrasted and different calculations for coronary illness forecast. Chronic kidney disease, also known as chronic kidney failure, describes the gradual loss of kidney functioning. A Persons kidney filters wastes and excess fluids from your blood, which are then excreted via urine. When chronic kidney disease reaches at an advanced stage, increase in dangerous levels of fluid, electrolytes and wastes can build up in a person body with CKD.

In the early stages of chronic kidney disease, a person may have few signs or symptoms. Chronic kidney disease may not become apparent or come under notice until the person's kidney function is significantly impaired. Treatment for chronic kidney disease focuses on slowing the progression of the kidney damage usually by controlling the underlying cause. Chronic kidney disease (CKD) means your kidneys are damaged and can't filter blood the way they should. The disease is known as "chronic" because the damage occurred to a person's kidney happens slowly over a long period of time. This damage can cause wastes to build up in the person's body having CKD. CKD can also cause several health problems. The kidneys' important function is to filter out extra water and wastes out of our blood to make urine. To keep our body working properly, the kidneys balance the level of salts and minerals such as calcium, phosphorus, sodium, and potassium that circulate in the blood. Our kidneys also make hormones that helps in controlling blood pressure, make red blood cells, and keep our bones strong and healthy.

Kidney disease often can get at the worst condition over a time period leading to kidney failure. If your kidneys fail, you will need dialysis or a kidney transplant to maintain your health. The sooner you know you have kidney disease, the sooner you can make changes to protect your kidneys. You may wonder how you can have CKD and feel fine. Our kidneys have a greater capacity to do their job than is needed to keep us healthy. For example, you can donate one kidney and remain healthy. You can also have kidney damage without showing any symptoms because despite of the damage caused, your kidneys are still functioning enough work to keep you feel better. For many people, the only way to know if you have kidney disease is to get your kidneys checked by getting blood and urine tests.

As kidney disease gets worse over a time period, a person may have few symptoms like swelling, called edema. Edema occurs when the kidneys can't flush out of extra fluid and salt. Edema can occur in the legs, feet, or ankles, and less often in the hands or face.

1. Firstly, Information mining is a procedure of separating valuable data from enormous measure of dataset. Information mining is found in various areas like picture mining, assessment mining, web mining, content mining, diagram mining and so on. The uses of information mining incorporate peculiarity location, a monetary information examination, medicinal information investigations, interpersonal organization examination and advertise investigation and so forth. And furthermore it has turned out to be prevalent in well being association as there is a prerequisite of expository strategy for foreseeing and discovering obscure examples. Information mining assumes an imperative part in finding out new patterns present in social insurance industry.

It's especially valuable in wellbeing field where there is no accessibility of affirmation supporting a specific type of treatment where there's no choice is found. Expansive measure of complex information is being produced by social insurance industry about patients, infections, healing facilities, medicinal equipment's, cases, treatment cost and so forth that requires handling and examination for data extraction. It contains an Arrangement of various devices and procedures which when can be connected to prepare information and provide valuable data to medicinal services experts for settling on right choices and that increases the execution of particular patient administration errands. Patients with similar medical issues can be grouped together and better treatment arrangements could be given in a view of information gathered from past patient history.

The worldwide medical issue which exist is confronted now a days is constant kidney infection (CKD) this is the major region of concern. Interminable kidney illness is a condition where kidneys end up plainly harmed and can't channel poisonous materials in our body. Our work highly concentrates and focuses on identifying life debilitating ailments like Chronic Kidney Disease (CKD) utilizing various Classification calculations like Naive Bayes and Artificial Neural Network(ANN) like C4.5 to predicts phases of Chronic kidney disease(CKD).

III. ANALYSIS AND FORMULATION

A worldwide medical issue which is consistently developing is chronic kidney sickness (CKD). It is an incessant condition related with expanded grimness and mortality, a high danger of numerous different ailments including cardiovascular sickness, and high human services costs. More than 2,000,000 individuals overall get dialysis or kidney transplant treatment to remain alive, yet this number may speak to just 10per of individuals who need treatment to live.

Most of the 2 million individuals who get treatment for kidney disappointment are in just moderately well of count attempts, which speak to 12 per of the worldwide populace. By examination, just 20 per of the total populace is treated in around 100 creating nations, and they speak to practically a large portion of the worldwide populace. Actually, more than one million individuals in 112 lower pay nations bite the dust from untreated kidney disappointment, because of the colossal budgetary weight of dialysis or kidney transplantation treatment. We propose a approach for the AI calculations have been utilized to anticipate and characterize in the human services have utilized the Support Vector Machine Algorithm to arrange and anticipate diabetes and pre-diabetes patients, and the outcomes show that SVM is helpful to order patients with normal illnesses. Thus, Mining have grouped Alzheimer's illness by utilizing a Support Vector Machine(SVM) to analyze entire mind anatomical attractive reverberation imaging (MRI) for a lot of patients, and the outcomes shows that SVM is a promising methodology for Alzheimer's ailment early recognition. Similarly, they have done coronary illness expectation utilizing the Probabilistic Neural Network Algorithm, Decision tree Algorithm, and Naive Bayes Algorithm, and PRNN furnishes the best out-comes contrasted and different calculations for coronary illness forecast.

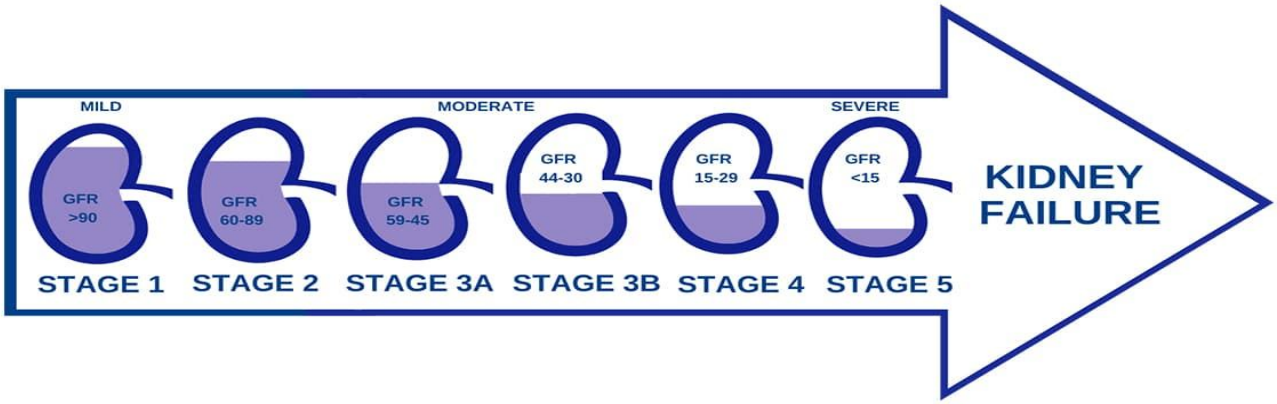


Fig.1 Stages of Kidney Diseases

IV. PROPOSED METHODOLOGY

A. *Architecture*

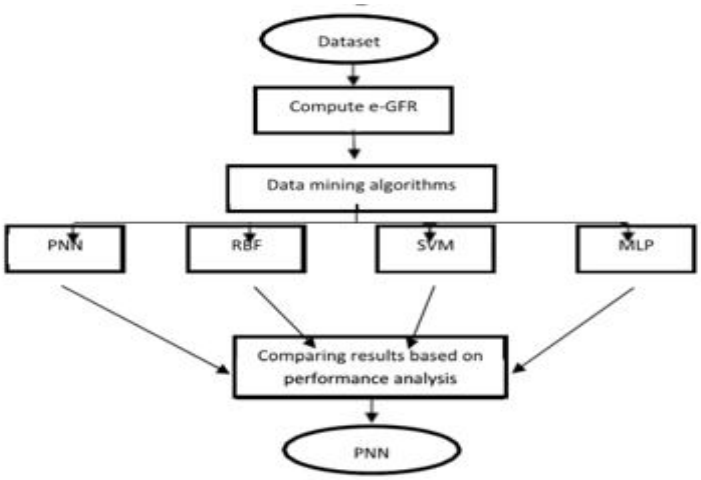


Fig.2 System Architecture

B. System Design

The system architecture of Chronic kidney Disease is shown in figure and its modules are mentioned as below

1) Data Mining Algorithm

a) Probabilistic Neural Networks: A probabilistic neural network is a feed forward neural network, which is widely used in classification and pattern recognition problems. In the PNN algorithm, the parent probability distribution function of each class is approximated by a Parzen window and a non-parametric function.

b) Radial basis function: Is a real-valued function whose value depends only on the distance between the input and some fixed point, either the origin, so that, or some other fixed point, called a center, so that any function that satisfies the property is a radial function.

c) Support Vector Machine: Is a supervised machine learning model that uses classification algorithms for two-group classification problems. After giving an SVM model sets of labelled training data for each category, they're able to categorize new text. So you're working on a text classification problem.

d) Multilayer Perceptron: Is a class of feed forward artificial neural network (ANN). The term learning, and is carried out through back propagation, a generalization of the least mean squares algorithm in the linear perceptron

2) Dataset: Data Mining was utilized in our study because it is a process of identifying novel, potentially useful, valid and ultimately understandable patterns in data. Supervised and unsupervised learning techniques are used for data mining classification. A “supervised" learning technique requires the building of a model based on previous performance analysis and is used in both medical and clinical research for classification statistical regression and association rules. On the other hand, the “unsupervised" learning technique is not guided by prior analysis and does not create a pre-analysis hypothesis. A model can be constructed based upon the results and is useful for clustering [6]. Three different types of the most commonly used artificial neural network algorithms and support vector machine algorithms have been used for this study, to determine which algorithm will give the best classification results, so as to identify the stage of chronic kidney disease.

DataSet Used

chronic_kidney_disease from UCI machine learning repository

The dataset contains:

- 400 instances
- 25 attributes
  - ✓ 14 are nominal
  - ✓ 11 are numeric

S.No	Attribute	Type	Values
1	Age	Numerical	in years
2	Blood Pressure	Numerical	in mm/Hg
3	Specific Gravity	Nominal	1.005,1.010,1.015,1.020,1.025
4	Albumin	Nominal	0,1,2,3,4,5
5	Sugar	Nominal	0,1,2,3,4,5
6	Red Blood Cells	Nominal	normal, abnormal
7	Pus Cell	Nominal	normal, abnormal
8	Pus Cell clumps	Nominal	present, notpresent
9	Bacteria	Nominal	present, notpresent
10	Blood Glucose	Numerical	in mgs/dl
11	Blood Urea	Numerical	in mgs/dl
12	Serum Creatinine	Numerical	in mgs/dl
13	Sodium	Numerical	in mEq/L
14	Potassium	Numerical	in mEq/L
15	Hemoglobin	Numerical	in gms
16	Packed Cell Volume	Numerical	-
17	White Blood Cell Count	Numerical	in cells/cumm
18	Red Blood Cell Count	Numerical	in millions/cmm
19	Hypertension	Nominal	Yes, No
20	Diabetes Mellitus	Nominal	Yes, No
21	Coronary Artery Disease	Nominal	Yes, No
22	Appetite	Nominal	Good, Poor
23	Pedal Edema	Nominal	Yes, No
24	Anemia	Nominal	Yes, No
25	Class	Nominal	ckd, notckd

Fig.3. Data Set

V. MATHEMATICAL MODEL

Mathematical Model: which describes the mathematical module of project which includes design and implementation constraints and set theory is shown below.

- vector v1 = (3, 5, 2) and vector v2 = (4, 1, 0).
- for kernel (k)

  K(v1, v2) = (3 * 4) + (5 * 1) + (2 * 0) = 17.0
- Many kernel functions have an optional gamma i.e. 0.5

  K(v1, v2) = 0.5 * [(3 * 4) + (5 * 1) + (2 * 0)] = 8.5
- polynomial kernel with degree = 2, gamma = 1.0 and constant = 0.

K(v1, v2) = [1.0 * ((3*4) + (5*1) + (2*0)) + 0]^2 = (1 * 17 + 0)^2 = 289.0

- public double PolyKernel(double[] v1, double[] v2)

```
{
 double sum = 0.0;
 for (int i = 0; i < v1.Length; ++i)
 sum += v1[i] * v2[i];
double z = this.gamma * sum + this.coef;
 return Math.Pow(z, this.degree);
 }
```

Fig.4.Mathematical Module

VI. RESULT

The stages of Chronic Kidney Disease (CKD) are mainly based on measured or estimated Glomerular Filtration Rate (eGFR). There are five stages of CKD but kidney function is normal in Stage 1 and minimally reduced in Stage 2. The KDOQI(Kidney Disease Outcomes Quality Initiative)stages of kidney disease are Definition of chronic :Labelling someone as having CKD requires two samples at least 90 days apart. Historical values can be used. The estimated Glomerular Filtration Rate (eGFR) depends on creatinine measurement, sex, race and age. One of the most accurate methods to calculate the eGFRisthe Modification of Dietin Renal Disease (MDRD)

eGFR calculation by applying the eGFR formula for detecting CKD

Modification of Diet in Renal Disease (MDRD) equation:

$eGFR = 186 \times (Creatinine/88.4)^{-1.154} \times (Age)^{-0.203} \times (0.742 \text{ if female}) \times (1.210 \text{ if black})$

Prognosis of CKD by GFR and Albuminuria Categories

GFR categories (ml/min/1.73 m²) Description and range				Albuminuria categories Description and range		
				A1	A2	A3
				Normal to mildly increased	Moderately increased	Severely increased
				<30 mg/g <3 mg/mmol	30-299 mg/g 3-29 mg/mmol	≥300 mg/g ≥30 mg/mmol
	G1	Normal or high	≥90			
	G2	Mildly decreased	60-90			
	G3a	Mildly to moderately decreased	45-59			
	G3b	Moderately to severely decreased	30-44			
	G4	Severely decreased	15-29			
	G5	Kidney failure	<15			

Green: low risk (if no other markers of kidney disease, no CKD); Yellow: moderately increased risk; Orange: high risk; Red, very high risk.
KDIGO 2012

Fig. 5 eGFR identification for CKD Risk

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VIII Aug 2020- Available at www.ijraset.com

VII.SCREENSHOT

Predicted the training and validating data used for algorithm i.e. Support vector machine and probabilistic neural network.

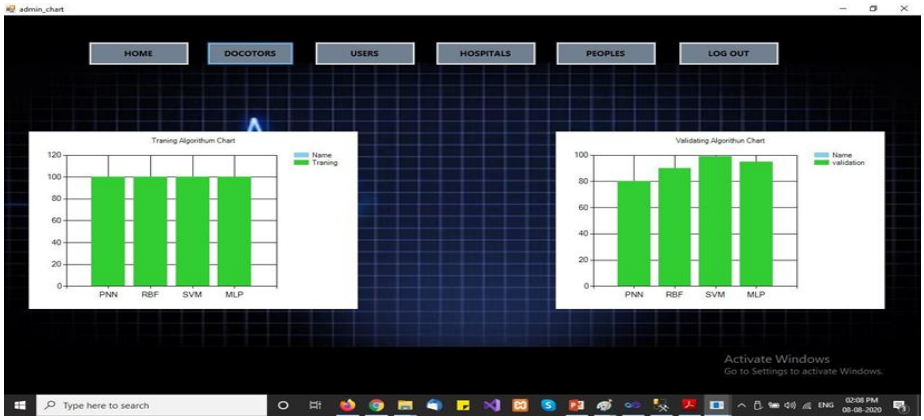


Fig. 6.Chart of Algorithm prediction

Here shows the all report of patient to doctor to check report of particular patient and send a suggestion to patient

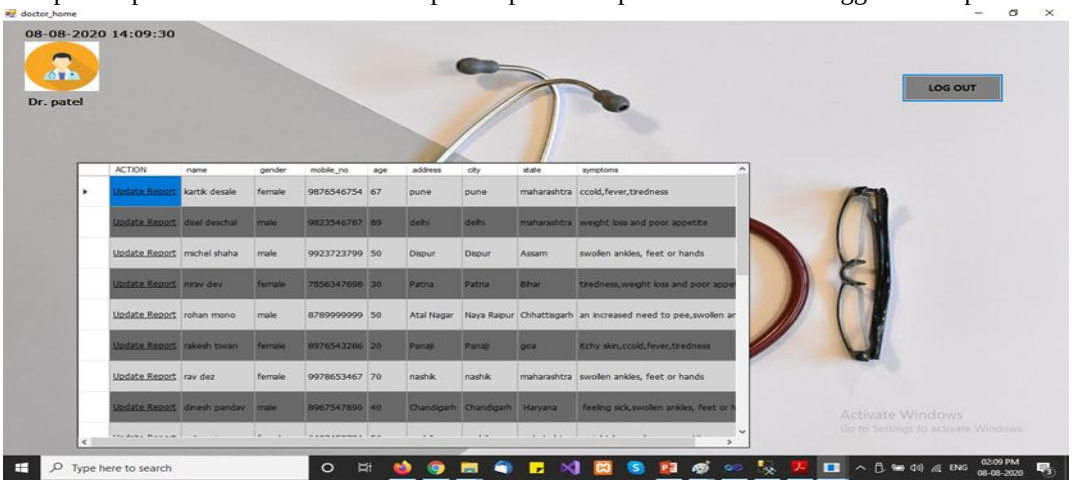


Fig. 7 Patient Details

Checks the possibility of CKD of patients:

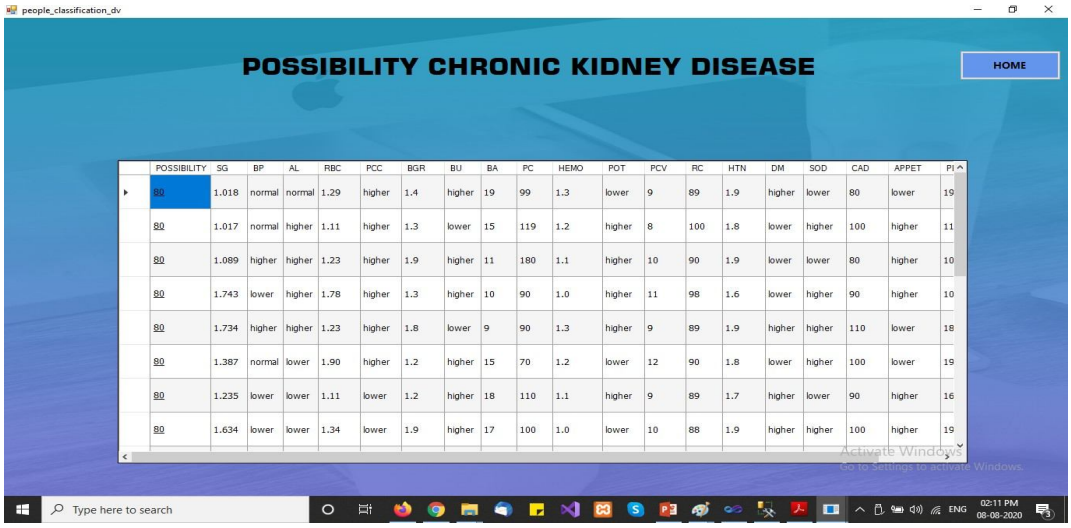


Fig. 8 Possibility of CKD

After checking report Doctor will send suggestion to particular patient.

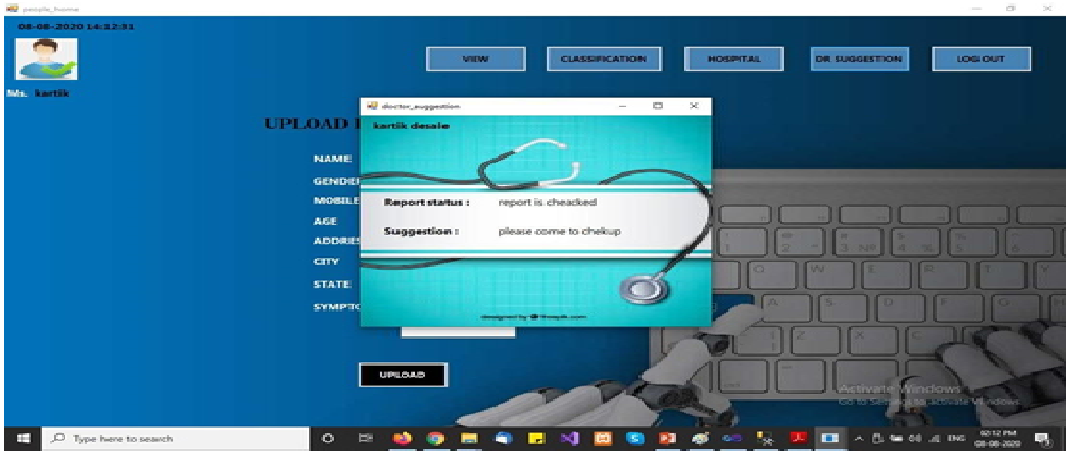


Fig. 9 Doctor Suggestion

VIII. CONCLUSION

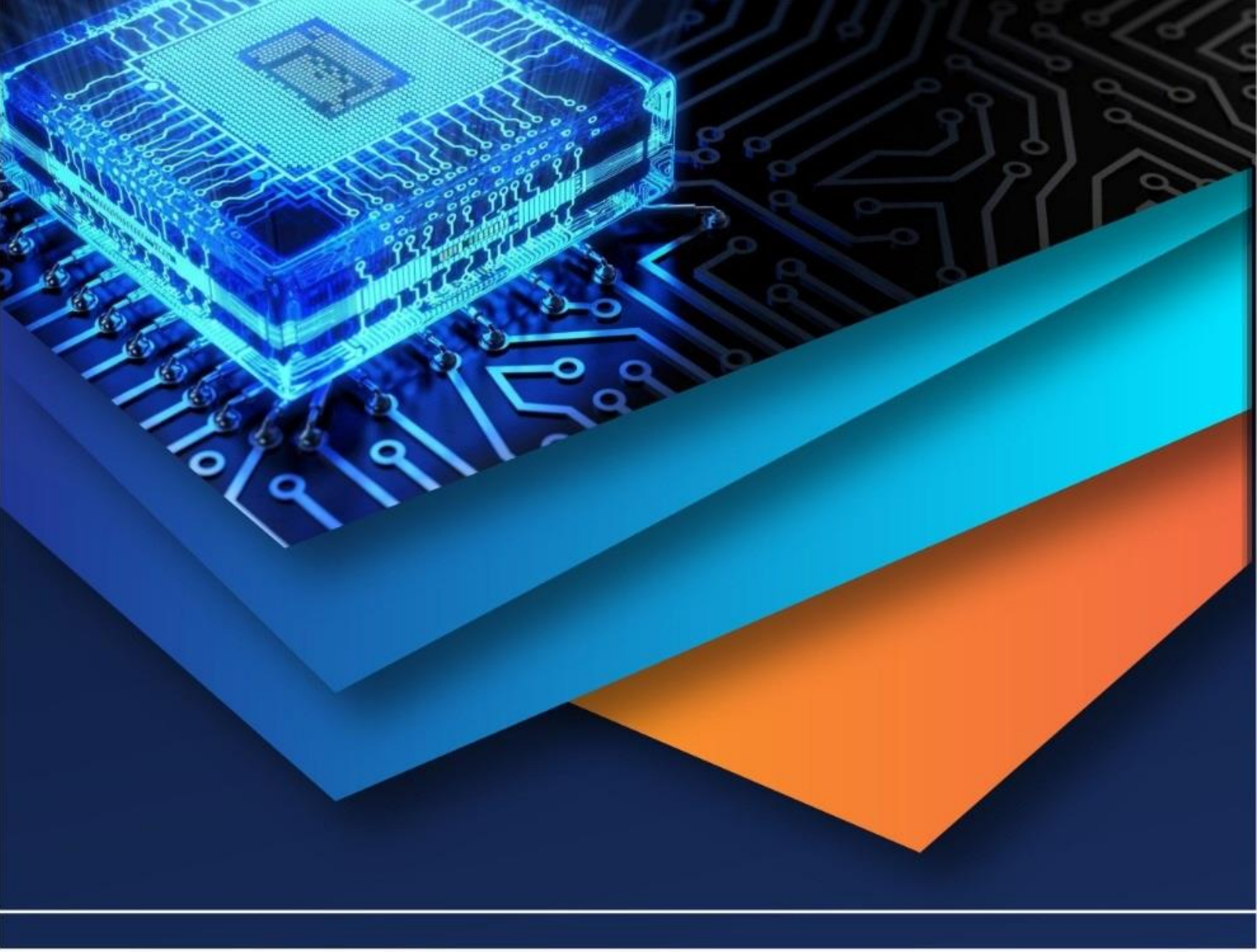
We have introduced an Engaging Data Mining Algorithm for Predicting Numerous Stages of Kidney Diseases We have performed an evaluation on a dataset of 400 patients, 250 among them have early stage of CKD. This dataset contains some noisy and missing values. Hence, we need a classification algorithm with the capability of handling missing and noisy values. We evaluated three classifiers: k-nearest neighbors, random forest, and neural networks to find a good solution for this application. To reduce over-fitting as well as to identify the most important predictive attributes for CKD, we have performed feature reduction

IX. ACKNOWLEDGEMENT

The satisfaction that accompanies the successful completion of any task would be incomplete without appreciating the people who made it possible. I am grateful to a number of individuals whose professional guidance along with encouragement have made it very pleasant endeavour to undertake this project. I have a great pleasure in presenting the project “Engaging Data Mining Algorithm for Predicting Numerous Stages of Kidney Diseases" under the guidance of Mr. Kishor N. Shedge and PG coordinator Prof. Devidas S. Thosar. I am truly indebted and grateful to Head of the Department Mr. Kishor N. Shedge for their valuable guidance and encouragement. I would also like to thank the Sir Visvesvaraya Institute of Technology (SVIT), Nashik for providing the required facilities, Internet access and important books. At last I must express my sincere heartfelt gratitude to all the Teaching and Non-teaching Staff members of Computer Engineering Department who helped me for their valuable time, support, comments, suggestions and persuasion.

REFERENCES

[1] Fadil Iqbal1, Aruna S. Pallewatte2, Janaka P. Wansapura1, —On Texture Analysis of Ultrasound Images of Chronic Kidney Disease,2018
[2] Helmie Arif Wibawa, Indra Malik, Indra Malik Evaluation of Kernel-Based Extreme Learning Machine Performance for Prediction of Chronic Kidney Disease 2017
[3] Dr. Uma N Dulhare, Mohammad Ayesha. Extraction of Action Rules for Chronic Kidney Disease using Na¨ıve Bayes Classifier. 2016
[4] M.P.N.M. Wick rama singhe, D.M. Per era, and K.A.D.C.P. Kahandawaa rachchi Dietary Prediction for Patients with Chronic Kidney Disease (CKD) by considering Blood Potassium Level using Machine Learning Algorithms. 2017
[5] Schaback R.Apractical guidet or adial basis functions.Electronic Resource 2007;11.
[6] LeveyAS, CoreshJ, BalkE, KauszAT, LevinA, SteffesMW, etal.National Kidney Foundation practice
[7] Guidelines for chronic kidney disease:evaluation,classification,and stratification.Ann Intern Med2003;139(2):[137]–47].
[8] Fukumi zuK. Active learning in multilayer perceptrons.Advance sin neural information processing systems.1996.p.295–301.
[9] YuW ,LiuT, ValdezR, GwinnM, KhouryMJ. Application of support vector machine modeling for prediction of common diseases :the case of diabetes andpre-diabetes. BMCM edInf DecisMak2010;10(1):16.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY