# Forecasting Provisional Interval of Kidney Disease Stage 3 to 5with the help of Data Mining Methods

Ambresh Bhadrashetty Assistant Professor, Dept. of Studies in Computer Applications (MCA), Visvesvaraya Technological University, Centre for PG studies, Kalaburagi *ambresh.bhadrashetty@gmail.com*  Sandesh Halappagol Student, MCA VI Semester, Dept. of Studies in Computer Applications (MCA), Visvesvaraya Technological University, Centre for PG studies, Kalaburagi halappagol.sand@gmail.com

*Abstract* –In the medical fields there is huge data found each and every day that's not possible to handle manually. Data Mining is one of the technology used to extract useful information among huge data. Data mining provide the method for transform these data to useful for decision making. Kidney disease is one of the health problem which is very expensive some can afford this treatment. In this paper we finding out the Chronic Kidney Disease(CKD) stages. The main objective of the study is to create a model for predicting transitional interval of kidney.

\*\*\*\*\*

Keywords - data mining; kidney dialysis;

## I.INTRODUCTION

One of the public health problem is kidney disease, number of kidney disease patients are increased yearly because, in a short period of time high possibility of death. As such, World Health Organization (WHO) has reported that South East Asia and Americas witness the highest annual rate around 1.4% population with the CKD (Chronic Kidney Disease).

As stated by 2014 report KDIGO (Kidney Disease Improving Global Outcome), this Kidney Diseases is divided into 6 stages based on the defined range of GFR (Glomerular Filtration Rate): CKD1, CKD2, CKD3a, CKD3b, CKD4 and CKD5. In End stage either patient need to transplantation or dialysis. To avoid this patient need to take proper treatment from the beginning of the stage 3.

Hospital staffs and doctor need to deal with personalized treatment plan. Using decision making tool to find out the stages and data mining approach to develop a classification model that is capable of prediction transitional interval of kidney disease stages 3 to 5.

#### **II.LITERATURE SURVEY**

In the paper "Chronic kidney disease: a research and public health priority" the authorsN. Peri co and G.Remuzzi describe that the growing global burden of non-communicable diseases (NCDs) worldwide has been disregarded until recently by policy makers, major aid donors and academics. In 2008, there were 57 million deaths globally, of which 63% were due to NCDs. These chronic diseases are the largest cause of death, led by cardiovascular disease followed by cancer, chronic lung disease and diabetes mellitus [1].

In the paper titled "Using data mining techniques to predict hospitalization of hemodialysis users" the authorsJ. Y.

Yeha, T. H. Wub and C. W. Tsaoa describe that Hemodialysis users might suffer from unhealthy care behaviors or long-term dialysis treatments and need to be hospitalized. If the hospitalization rate of a hemodialysis center is high, its service quality will be low. Therefore, decreasing hospitalization rate is a crucial problem for health care centers. This study combines temporal abstraction with data mining techniques for analyzing dialysis users' biochemical data to develop a decision support system [2].

In the paper titled "Predicting survival time for kidney dialysis users: a data mining approach" the authors A. Kusiaka, B. Dixonb and S. Shaha describe that the cost for providing care for users on hemodialysis due to end stage kidney disease is high. Finding ways to improve user outcomes and reduce the cost of dialysis is important. Dialysis care is intricate and multiple factors may influence user survival. Individual user survival may depend on a complex interrelationship between multiple demographic and clinical parameters, medications, medical interventions, and the dialysis treatment prescription. In this research, data preprocessing, data transformations, and a data mining approach are used to elicit knowledge about the interaction between many of these measured parameters and user survival [3].

#### **III. PROBLEM STATEMENT**

The main problem in this project, kidney illness patients, one of the overall general medical issues, has been expanded yearly. Because of the high plausibility of death inside a brief timeframe a patient have to be hospitalized and effectively treated because the main day of investigated as the stage 3.



Fig.1: architecture

In the above fig doctor or hospital staff will register first and enter the all details of patients and also view the all patients details but Hospital A don't have permission see or view the other hospital patients details only the admin that's nothing but a segregation, in segregation all patients details a viewed at time and find out the CKD stages of all patients.

## **V. IMPLEMENTATION**

**Data Mining:** Usually, information mining (sometimes mentioned to as information or records discovery) is that the technique of analyzing data from totally altered views and summarizing it into obliging info - info which will be wont to growth income, cuts costs, or both. Data miming code is one in all change of analytical tools for analyzing information. It allows users to analyze information from many alternate dimensions of angles categorize it and recap the relationships recognized. Officially, data mining is that the technique of finding correlations or patterns among dozens of fields in huge comparative databases.

**Data-mining Classification using decision tree:** The stocks an electronic info as more as operation of enterprise are computerized using a list in the mining. We can accumulate a large amount of queries it is a knowledge based discovery. Classification is parting are arrangements of things into modules if classes continue generated without looking at statistics it developed deal with lots of entities. Decision tree is a top down of tress's just select an attribute for root node and create an outlet for each probable attributes digit, next split instances into subset's then rehash the methodology recursively for each branch, utilizing condition to achieve the branch stops the cases same class. Predicting class's no model, model-evaluation, clustering. Tree induction is given to determine how n when to stop splitting, determine how to split records.

## **Decision Tree Algorithm**

FUNCTION build\_dec\_tree(examples,atts) // Takes a set of classified examples and // a list of attributes, atts. Returns the // root node of a decision tree Create node N; IF examples are all in same class

IJFRCSCE | July 2017, Available @ http://www.ijfrcsce.org

THEN RETURN N labelled with that class: IF atts is empty THEN RETURN N labelled with modal example class; best\_att = choose\_best\_att(examples,atts); label N with best att; FOR each value ai of best att si = subset examples with best att = ai;IF si is not empty THEN new\_atts = atts - best\_att; subtree = build\_dec\_tree(si,new\_atts); attach subtree as child of N; ELSE Create leaf node L; Label L with modal example class; attach L as child of N; **RETURN N:** 

Choosing the Best Attribute

A reasonable answer would be the attribute that best discriminates the examples with respect to their classes. Still many possible answers. Many different criteria have many used. The most popular is information gain.

Number of times Attribute1 has value 0.5 as stage1

Number of times Attribute1 has value 0.5 as stage2 or stage3

Attribute1	Stage1	Stage2 or Stage3	Information Gain
0.5	4	5	22
0.6	5	5	23
0.7	5	7	34

After obtaining the gain of all attributes and pick the attribute with highest information gain as root node

Again from there create a subset for each values and determine the highest information gain like this repeat process until the stage is obtained

**Classification:**Classification is a supervised learning method of Data mining where there are few predefined classes of database objects, which are passed on the model in order to separate the records in the form of those pre define groups. We give new object and apply the 'learned model' to consign this object to one of the predefined class.

**K-nearest neighbors:** Thenearest neighbor algorithm(KNN) belongs to the class of pattern recognition statistical methods. The method does not impose a priori any assumptions about the distribution from which the modeling sample is drawn. It involves a training set with both positive and negative values. A new sample is classified by calculating the distance to the nearest neighbouring training case. The sign of that point will determine the classification of the sample. In the k-nearest

neighbor classifier, the k nearest points are considered and the sign of the majority is used to classify the sample

In original formats, ranges of some selected attributes can be wide and uneven, which have been considered as not appropriate for data mining development. Thus, these values were transformed into suitable forms using the nominalization technique. As such, sugar value (high and normal-sugar value), fat value (high and normal-fat value), waste value in blood (high, low and normal-waste value) were scaled into the [0, 1] range with the following formulas.

$FBS - high = \frac{f}{f}$	bsHigh bsTotal
FBS - normal =	$\frac{fbsNormal}{fbsTotal}$
$HDL - high = \frac{h}{h}$	dlHigh dlTotal
HDL - normal =	hdlNormal hdlTotal
$BUN - high = \frac{b}{bh}$	unHigh unTotal
$BUN - low = \frac{bu}{buo}$	nLow nTotal
BUN - normal =	bunNormal bunTotal

All formulas above use similar pattern in calculating. For example, fbsNormal means the number of times with normal sugar value in blood, fbsHigh means the number of times with high-sugar value in blood, and fbsTotal means the total number of times with blood sugar value.

#### TABEL 1. A LIST OF SELECTED PARAMETERS FOR CREATING A MODEL

Parameter / Attribute	Data type	Value Domain	
Gender	Nominal	{Male, Female}	
Age	Numeric	[0, 1]	
High-sugar value in blood (FBS-high)	Numeric	[0, 1]	
Normal-sugar value in blood (FBS-normal)	Numeric	[0, 1]	
High-fat value in blood (HDL-high)	Numeric	[0, 1]	
Normal-fat value in blood (HDL-normal)	Numeric	[0, 1]	
High-waste value in blood (BUN-high)	Numeric	[0, 1]	
Low-waste value in blood (BUN-low)	Numeric	[0, 1]	
Normal-waste value in blood (BUN- normal)	Numeric	[0, 1]	
BMIHigh	Nominal	{Yes, No}	
NSAIDs	Nominal	{Yes, No}	
Diabetes (DM)	Nominal	{Yes, No}	
Hypertension (HT)	Nominal	{Yes, No}	
Stone (N20)	Nominal	{Yes, No}	
Urinary disease (N30)	Nominal	{Yes, No}	
Ischemic heart disease (1200)	Nominal	{Yes, No}	
Disease related to heart failure (1500)	Nominal	{Yes, No}	
Gout or Rheumatoid (M109)	Nominal	{Yes, No}	
Albumin	Nominal	{Yes, No}	

After the nominalization stage, the next step is to manipulate the time period between the first day of stage 3 and stage 5 as the model's answer or known class label. These data was grouped into two classes: 108 records which is less than or equal to 3.5 years and the other 60 records that is more than 3.5 years. These two categories were regarded as the target classes of the model generation.

VI.RESULT

er Patent Data				
First Name		Last Name		
Age		Gender	Male	
FBS-high		FBS-normal		
HDL-high		HDL-normal		
BUN-high		BUN-low		
BUN-Normal		BMIHigh	Yes	
NSAIDs	Yes	Diabetes (OM)	Yes	
Hypertension (HI)	Yes	Stone (N20)	Yes	
Urinary disease (N30)	Yes	Ischemic heart disease (1200)	Yes	
Disease related to heart failure (1500)	Yes	Gout or Rheumatoid (M1)	09) <sup> Yes</sup>	
Albumin	Yes			



#### Here a doctor or hospital staff fill the details of the patients and save the details and they can view the details of patients

iew All patent	Data								
Name	Age	Highsugar	Normalsugar	Highfat	Normalfat	Highwaste	Lowwaste	Normalwaste	BMIHigh
xyz	40	12	90	1	1	20	85	14	1
mallu	45	10	10	10	10	10	10	10	1
sfa	45	1	1	1	1	1	1	1	1
qwe	45	1	0	0	0	0	D	0	1
qwe	45	36	35	38	35	40	30	0	1
apc	45	36	36	36	37	47	32	38	1
aqw	48	52	54	50	59	56	54	52	1
aqw	48	50	50	50	50	50	50	50	1
adw	48	360	360	360	360	360	360	360	1
adiv.	48	450	450	450	450	450	450	450	1
/5	23	500	500	500	400	450	200	100	1
adha	45	600	500	600	500	600	500	500	1
new	50	650	120	450	120	600	150	450	0

# Fig.3: segregated data

# Here admin can see the all hospital (hospital A and B) patient's details at a time

edication Stage 1 to 3		
Name	Age	Predication Value
72	40	31
allu	45	10
a	45	1
we	45	0
(v/e	45	30
ibc	45	37
qw	48	53
qw	48	50
qw	48	360
5	23	378
ew .	50	362
JSER1	44	354
user2	35	352

## Fig.4: Prediction stages

Here in this page it show the stages from 3 to 5, if prediction value is between 0 to 3.5 than it's between stage 1 to 3 that's is mild stage, if prediction value is between 3.6 to 4 than its stage 4 that's is severe stage, if prediction value is 4 or more than its stage 5 that's end stag



Fig.5: Analysis Form

#### VII. CONCLUSION

We collecting the all the details of patients from different hospitals and segregate the details and find out the stages and using data mining technology for prediction transitional interval of kidney states so it's helpful for the doctors to give a treatments. In Future patient also can see the stages and doctor can suggests the treatments to the patients and send the report to patient.

## REFERENCES

[1] N. Peri co and G.Remuzzi, "Chronic kidney disease: a research and public health priority," in Nephrology Dialysis Transplantation.

- [2] J. Y. Yeha, T. H. Wub and C. W. Tsaoa, "Using data mining techniques to predict hospitalization of hemodialysis patients," in Decision Support Systems, vol. 50,2011, pp. 439-448.
- [3] A.Kusiaka, B. Dixonb and S. Shaha, "Predicting survival time for kidney dialysis patients: a data mining approach," in Computers in Biology and Medicine, vol. 35, 2005, pp. 311-327.
- [4] Social Security Office. (2014). Policy for kidney dialysis
  [Online]. Available:http://www.sso.go.th/wpr/content.jsp?lang=th& %20cat=762%20&id=3700
- [5] Agency for Healthcare Research and Quality. (2012). KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease [Online]. Available: http://www.guideline.gov/content.aspx?id=465 I