

# Kidney Failure Due to Diabetics – Detection using Classification Algorithm in Data Mining

J.Vijayalakshmi

Assistant Professor, Dept.of. Comp. Science, K.C.S Kasi Nadar College of Arts and Science, Chennai.

Email ID:viji.jayaprakash@gmail.com

**Abstract:** In order to analyse the chosen data from various points of view, data mining is used as the effective process. This process is also used to sum-up all those views into useful information. There are several types of algorithms in data mining such as Classification algorithms, Regression, Segmentation algorithms, association algorithms, sequence analysis algorithms, etc.,. The classification algorithm can be used to bifurcate the data set from the given data set and foretell one or more discrete variables, based on the other attributes in the dataset. The ID3 (Iterative Dichotomiser 3) algorithm is an original data set  $S$  as the root node. An unutilised attribute of the data set  $S$  calculates the entropy  $H(S)$  (or Information gain  $IG(A)$ ) of the attribute. Upon its selection, the attribute should have the smallest entropy (or largest information gain) value. The prime objective of this paper is to analyze the data from a Kidney disorder due to diabetics by using classification technique to predict class accurately.

**Keywords:** Data mining, Classification algorithm, Decision tree, medical dataset.

## 1. INTRODUCTION

To discover patterns in large sets of data, this researcher uses data mining as a process for computation. It is a device for the intersection of intelligence which is not natural, machine learning algorithm, statistics and database systems. To cull information is the ultimate objective of the data mining process, from a data set and transform it into an intelligible mass for later use. For analysing "knowledge discovery in databases" process, or KDD, data mining is the ideal step. This involves six common classes of tasks, such as anomaly detection, association rule mining, clustering, classification, regression and summarization. The classification method techniques used for large medical data set mining, clustering, classification, regression and summarization.

The classification method techniques are used for large medical data set for the diagnosis and prognosis of kidney failure due to diabetics. To obtain large set of medical data, the classification method is extremely useful. Genetic based ID3 algorithm is the simplest and easiest type for diagnosis and prognosis of kidney failure due to diabetics from the given data set. Decision tree classifier does not require any domain knowledge or a setting of parameter. It can handle multidimensional data and is simple and fast. Various decision tree algorithms like CART, ID3, C4.5, SLIQ, and SPRINT are available which are not complex. They function fast and they can accurately deal with data of varied dimensions. There is no need of any basic knowledge or measurement limits.

The approximate prevalence of kidney failure is 800 per million population (pmp), and the incidence of end-stage renal disease (ESRD) is 150–200 pmp. The most common cause of kidney failure in population-based studies is diabetic nephropathy. Early detection of kidney failure is essential for reducing life losses. The estimation of the ultimate result of a disease and the analysis of the course it is likely to take is called prognosis. The hope for the successful treatment of the disease is caused by the prognosis of the affected patient. Prognostic information is generated by statement of prognosis. In the formulation of the prognosis certain bits of information are used. These pieces of information are related to the obvious yield of the disease. Such information can be called in other words prognostic factors. This paper is structured as follows: Section 2 contains the review concepts of pre processing method, ID3 and Diabetics nephropathy. Section 3 contains existing method. Section 4 explains our proposed method. In section 5 Results are discussed and section 6 has the conclusion.

## 2. BASIC CONCEPTS

The pre-processing methods using data mining techniques identify the target data from the large data set. The pre-processing method has been some tasks, such as Data cleaning, Data integration, Data transformation, Data reduction, Data discretization. Data cleaning: this could be defined as a process to eliminate noise and make the data coherent and consistent. By the process, values which are absent are incorporated, outliers are recognized and detected. Data integration: using many databases, data cubes or files. Data transformation: normalization and aggregation. Data reduction: Reducing the volume but producing the same and similar analytical results. Data discretization: Part of data reduction and replacing numerical attributes with nominal ones.

### Decision tree algorithm

A decision tree is a tool majorly used to support the decision making operations. It uses a tree like model which comprises its consequences like the event of the chance outcomes, costs incurred for the resources and utilities. This method uses the decision tree like method to predict which models get them closer to the target result. Two types of decision trees exist. They are 1) classification tree 2) Regression tree. The target element in the tree model takes a finite set of values commonly known as Classification Trees. If the target tree takes continuous values then the trees are called Regression trees.

### Algorithms for decision trees

ID3 (Iterative Dichotomiser 3)

We can expand ID3 as Iterative Dichotomiser 3 Mr.Ross Quinlan invented this algorithm. Inorder to take a dataset and

produce a decision tree from it, this ID3 is used. ID3 is used even before C4.5 algorithm. This typical algorithm enables learning through machines and is useful in the areas pertaining to the processing of natural languages. This algorithm known as ID3 starts with the root node which is the original set S. This algorithm iterates through each unutilised feature of the set S. This goes on in every iteration and leads to the calculation of the entropy H(S) of that particular attribute. The attribute with the smallest entropy or largest gain of information is chosen by it.

**Information gain**

Used by the ID3 tree generation algorithms. Information Gain is based on the concept of Entropy from Information Theory. In order to generate subsets of the data . the chosen attribute is used to split the set ‘S’.

$$I_E(f) = - \sum_{i=1}^m f_i \log_2 f_i$$

**Diabetic nephropathy**

Kidney disease that results from diabetes is the number one cause of kidney failure. Almost a third of people with diabetes develop diabetic nephropathy. People with diabetes and kidney disease suffer more overall than people with only kidney disease. This is because people with diabetes tend to have other long-standing medical conditions, like high blood pressure, high cholesterol, and blood vessel disease (atherosclerosis). People with diabetes are also more likely to have other kidney-related problems, such as bladder infections and nerve damage to the bladder.

Kidney disease in type 1 diabetes is slightly different than in type 2 diabetes. In type 1 diabetes, kidney disease rarely begins in the first 10 years after diagnosis of diabetes. In type 2 diabetes, some patients already have kidney disease by the time they are diagnosed with diabetes.

**Symptoms of Diabetic Nephropathy**

There are generally no symptoms of early diabetic nephropathy. As the kidney function worsens, symptoms may include:

- Swelling of the hands, feet, and face
- Trouble sleeping or concentrating
- Poor appetite
- Nausea
- Weakness
- Itching (end-stage kidney disease) and extremely dry skin
- Drowsiness (end-stage kidney disease)
- Abnormalities in the heart’s regular rhythm, because of increased potassium in the blood
- Muscle twitching

Table 1: Attributes of Diabetes Dataset

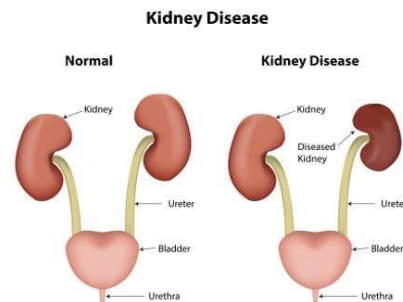
| Attribute No. | Attribute | Description  |
|---------------|-----------|--|
| 1             | Plasma    | Plasma glucose concentration a 2 hours in an oral glucose tolerance test |
| 2             | Pressure  | Diastolic blood pressure(mmHg)   |
| 3             | Skin      | Triceps skin fold thickness(mm)  |

|   |           |                                |
|---|-----------|--------------------------------|
| 4 | Insulin   | 2-Hour serum insulin (mu U/ml) |
| 5 | Pregnancy | Number of times pregnant       |
| 6 | Mass      | Body Mass Index(BMI)           |
| 7 | Pedigree  | Diabetes Pedigree function     |
| 8 | Age       | Age(in years)                  |
| 9 | Class     | Class variable(0 or 1)         |

**Treatment for kidney failure**

If your kidneys can't keep up with waste and fluid clearance on their own and you develop complete or near-complete kidney failure, you have end-stage kidney disease. At that point, you need dialysis or a kidney transplant.

- **Dialysis.** Dialysis artificially removes waste products and extra fluid from your blood when your kidneys can no longer do this. In hemodialysis, a machine filters waste and excess fluids from your blood. In peritoneal dialysis, a thin tube (catheter) inserted into your abdomen fills your abdominal cavity with a dialysis solution that absorbs waste and excess fluids. After a period of time, the dialysis solution drains from your body, carrying the waste with it.
- **Kidney transplant.** A kidney transplant involves surgically placing a healthy kidney from a donor into your body. Transplanted kidneys can come from deceased or living donors. You'll need to take medications for the rest of your life to keep your body from rejecting the new organ. You don't need to be on dialysis to have a kidney transplant.



For some who choose not to have dialysis or a kidney transplant, a third option is to treat kidney failure with conservative measures. However, once you have complete kidney failure, your life expectancy generally would be only a few months.

**3. EXISTING METHOD**

Existing method used six algorithms i.e. Naive Bayes, Multilayer Perception, SVM, J48, Conjunctive Rule and Decision Table for experiments. These algorithms are implemented using WEKA data mining tool to analyze accuracy which is obtained after running these algorithms in the output window. These algorithms have been compared with classification accuracy to each other on the basis of correctly classified instances, time taken to build model, time taken to test the model, mean absolute error, Kappa statistics and ROC Area. In the experiments Multilayer perceptron algorithm gives better classification accuracy and prediction performance to predict chronic kidney disease (CKD) using relevant dataset available at UCI machine learning repository.

### 3. PROPOSED METHOD

Our proposed approach uses Decision tree (ID3) to improve the classification accuracy of diabetic nephropathy data set. Applying algorithm is for the large data set collection from medical centre.

Step 1: Load the medical data set

Step 2: Apply pre-processing method on the data set and Identify related data set

Step 3: Applying ID3 with classified data set and getting diabetic nephropathy data set

Step 4: Classified diabetic data diagnosis and prognosis of diabetic nephropathy data set.

Accuracy of sample data has classifier computed as:

$$\text{Accuracy} = \frac{\text{Number of samples correctly classified in test data}}{\text{Total number of samples in the test data}}$$

### 4. RESULTS AND DISCUSSION

This method can be used in the algorithm in data mining techniques for improved accuracy level than existing method. The performance of our proposed method has been tested with 10 data sets from medical data set and 2 non medical data sets. We can easily detect kidney failure by this method

### 5. CONCLUSION

In this paper the researcher has presented classification of diabetic nephropathy using ID3 algorithm. Our proposed method is to improve accuracy level using the medical data set given. Experiment results carried out on 10 data sets show that our approach is a competitive method for classification. The proposed method is for identification of diabetic data set and diagnosis and prognosis of diabetic nephropathy.

#### Reference

- [1] Joshi J, Rinal D, Patel J, Diagnosis And Prognosis of Breast Cancer Using Classification Rules, International Journal of Engineering Research and General Science,2(6):315-323, October 2014.
- [2] Kusiak, A., Dixon, B., & Shah, S., 2005. Predicting survival time for kidney dialysis patients: a data mining approach. Computers in biology and medicine, 35(4), pp. 311-327.
- [3] Mukesh kumari and Dr. Rajan Vohra, "Prediction of Diabetes Using Bayesian Network," in proceeding of International Journal of Computer Science and Information Technologies, vol.5, 2014
- [4] Iyer Aiswarya, Jeyalatha S and Sumbaly Ronak. Diagnosis of diabetes using classification mining techniques. International Journal of Data Mining & Knowledge Management Process. 2015; 5:1-14.
- [5] Sanakal Ravi and Jayakumari T. Prognosis of Diabetes Using Data mining Approach-Fuzzy C Means Clustering and Support Vector Machine. International Journal of Computer Trends and Technology. 2014; 11:94-8.