

Early Detection Of Chronic Kidney Disease Using

Machine Learning

Literature Survey

Team Members

- Muppala Mona Sree
- Monika M
- Manju Parkavi G
- Priyadharshini E M

IV year – Computer Science and Engineering, R.M.K Engineering College

Problem Statement:

To develop a web application that efficiently predict the Chronic Kidney Disease By early using Machine Learning Techniques.

Introduction:

Chronic Kidney Disease (CKD) is major medical problem and can be cured if treated in the early stages. Usually, people are not aware that medical tests we take for different purposes could contain valuable information concerning kidney diseases. Consequently, Attributes of various medical tests are investigated to distinguish which attributes may contain helpful information about the disease . The information says that it helps us to measure the severity of the problem and we

make use of such information to build a machine learning model that predicts Chronic Kidney Disease.

Survey 1:

Machine Learning methods used for Chronic Kidney Diseases:

According the 2010 global burden of disease study, Chronic Kidney Diseases (CKD) was ranked 18th in the list of causes of total no. of deaths worldwide. 10% of the population worldwide is affected by CKD. The prediction of CKD can become a boon for the population to predict the health. Various method and techniques are undergoing the research phase for developing the most accurate CKD prediction system. Using Machine Learning techniques is the most promising one in this area due to its computing function and Machine Learning rules. In recent time Neural network system has discovered its use in disease diagnoses, which is depended upon prediction from symptoms data set. After neural network is trained using back propagation algorithms, this trained neural network system is used for detection of kidney disease in the human body.

Some machine learning methods:

Supervised learning

An algorithm uses training data and feedback from humans to learn the relationship of given inputs to a given output. For instance, a practitioner can use marketing expense and weather forecast as input data to predict the sales of cans.

Unsupervised learning

In unsupervised learning, an algorithm explores input data without being given an explicit output variable (e.g., explores customer demographic data to identify patterns).

Reinforcement learning

It is about taking suitable action to maximize reward in a particular situation. It is employed by various software and machines to find the best possible behavior or path it should take in a specific situation.

Advantages:

- Supervised classification can be much more accurate than unsupervised classification
- Labeling of data demands a lot of manual work and expenses. Unsupervised learning solves the problem by learning the data and classifying it without any labels.
- Reinforcement learning doesn't require large labeled datasets. It's massive advantage because as the amount of data in the world grows it becomes more and more costly to label it for all required applications.

Disadvantage:

- Supervised learning cannot handle all complex tasks in machine learning.
- Less accuracy of the results is because the input data is not known and not labeled by people in advance. This means that the machine requires to do this itself.

- It does not take into account the workers inner feelings and reasons for their positive or negative behavior rather it focuses only on the factors.

Conclusion:

The proposed system will definitely help in improving the prediction of Chronic Kidney disease system by increasing its accuracy and prediction capability by reducing the error. The existing systems are focused on using single neural network. The mean square error will be reduced in subsequent neural training iterations.

Survey 2:

Dataset and methods:

Dataset

The Dataset here we use is the publically available CKD Dataset from UCI repository. It contains 400 samples of two different classes. Out of 25 attributes, 11 are numeric and 13 are nominal and one is class attribute. The data set contains number of missing values. Here the information of dataset uses the patient's data like age, blood pressure, specific gravity, albumin, sugar, red blood cells etc.

List of attributes present in the CKD dataset

Attributes	Type
Age	Numeric
Blood Pressure	Numeric
Specific Gravity	Numeric

Albumin	Numeric
Sugar	Numeric
Red Blood Cells	Numeric

CKD is caused due to diabetes and high blood pressure. Due to Diabetes our many organs get affected and it will be followed by high blood sugar.

Methods

a.) Pre-Processing

Data Pre-Processing is that stage where the data that is distorted, or encoded is brought to such a state that the machine can easily analyze it. A dataset can be observed as a group of data objects. Data objects are labeled by a number of features, that ensures the basic features of an object, such as the mass of a physical object or the time at which an event ensured. In the dataset there may be missing values, they can either eliminated or estimated. The most common method of dealing with missing values is filling them in with mean, median or mode value of respective feature. As object values cannot be used for the analysis we have to convert the numeric values with type as object to float64 type. Null values in the categorical attributes are changed with the most recurrent occurring value current in that attribute column. Label encoding is done to translate categorical attributes into numeric attribute by conveying each unique attribute value to an integer.

b.) Feature Selection

Feature Selection is the method where we computationally select the features which contribute most to our prediction variable or output. In this study we used Ant Colony Optimization (ACO) for selecting the best features from the dataset. It

is a technique for solving computational problems which can be condensed to finding good paths through graphs. Artificial Ants stand for multi-agent methods enthused by the behavior of real ants. The pheromone-based communication of biological ants is often the main paradigm used. Combinations of Artificial Ants and local search algorithms have become a method of choice for numerous optimization tasks involving some sort of graph. This algorithm evaluates the intensity of pheromone during each iteration rather than accumulating them. The proposed algorithm will change a small number of features in subsets which are selected by choosing the best ants. A classification algorithm has to be used to evaluate the performance of the subsets that is wrapper evaluation function.

c.) Classification

For classification we use Support Vector Machine(SVM) to predict the disease and its performance. As a first step we have to import the libraries for classification and prediction. We import SVM and datasets from the scikit-learn library. NumPy for carrying out efficient mathematical computations. Accuracy- score from sklearn metrics to predict the accuracy of the model. We have divided the data into training and testing sets. Now is the time to train our SVM on the training data. scikit-learn contains the SVM library, which contains built-in classes for various SVM algorithms. Since we are going to perform a classification task, we will use the support vector classifier class, which is written as SVC in the scikitlearn's SVM library.

Advantages:

Early detection of CKD with reduced glomerular filtration rate (GFR) may also facilitate appropriate dosing of medications and allow timely preparation for kidney replacement.

Disadvantages:

CKD includes conditions that damage kidneys and decrease their ability to keep healthy by filtering wastes from blood.

Survey 3:

In many cases, Chronic Kidney Disease is only found when a routine blood or urine test you have for another problem shows that your kidneys may not be working normally.

Symptoms

1. Weight loss or poor appetite.
2. Swollen ankles, feet or hands.
3. Shortness of breath.
4. Tiredness.
5. Blood in your pee (urine).
6. Peeing more than usual, particularly at night.

Tests for CKD (Chronic Kidney Disease) :**Blood test**

- The main test for kidney disease is a blood test. The test measures the levels of a waste product called creatinine in your blood.
- Your doctor uses your blood test results, plus your age, size, gender and ethnic group to calculate how many millilitres of waste your kidneys should be able to filter in a minute.

- This calculation is known as your estimated glomerular filtration rate (eGFR).
- Healthy kidneys should be able to filter more than 90ml/min. You may have CKD if your rate is lower than this.

Urine test

- A urine test is also done to:
- check the levels of substances called albumin and creatinine in your urine – known as the albumin:creatinine ratio, or ACR
- check for blood or protein in your urine
- Alongside your eGFR, urine tests can help give a more accurate picture of how well your kidneys are working.
- Other tests
- Sometimes other tests are also used to assess the level of damage to your kidneys.

These may include

- an ultrasound scan, MRI scan or CT scan – to see what the kidneys look like and check whether there are any blockages
- a kidney biopsy – a small sample of kidney tissue is removed using a needle and the cells are examined under a microscope for signs of damage
- Test results and stages of CKD
- Your test results can be used to determine how damaged your kidneys are, known as the stage of CKD.

This can help your doctor decide the best treatment for you and how often you should have tests to monitor your condition.

Your eGFR results is given as a stage from 1 of 5:

- stage 1 (G1) – a normal eGFR above 90ml/min, but other tests have detected signs of kidney damage
- stage 2 (G2) – a slightly reduced eGFR of 60 to 89ml/min, with other signs of kidney damage
- stage 3a (G3a) – an eGFR of 45 to 59ml/min
- stage 3b (G3b) – an eGFR of 30 to 44ml/min
- stage 4 (G4) – an eGFR of 15 to 29ml/min
- stage 5 (G5) – an eGFR below 15ml/min, meaning the kidneys have lost almost all of their function

Your ACR result is given as a stage from 1 to 3:

- A1 – an ACR of less than 3mg/mmol
- A2 – an ACR of 3 to 30mg/mmol
- A3 – an ACR of more than 30mg/mmol

For both eGFR and ACR, a higher stage indicates more severe kidney disease.

Survey 4:

Abstract

Chronic kidney disease (CKD) is a disorder in which the kidneys are weakened and become unable to filter blood. It lowers the human ability to remain healthy. The field of biosciences has progressed and produced vast volumes of knowledge from electronic health records. Heart disorders, anemia, bone diseases, elevated potassium, and calcium are the very prevalent complications that arise from kidney failure. Early identification of CKD can improve the quality of life greatly. To achieve this, various machine learning techniques have been introduced so far that use the data in electronic health record (EHR) to predict CKD. This chapter studies various machine learning algorithms like support vector machine, random forest,

probabilistic neural network, Apriori, ZeroR, OneR, naive Bayes, J48, IBk (k-nearest neighbor), ensemble method, etc. and compares their accuracy. The study aims in finding the best-suited technique from different methods of machine learning for the early detection of CKD by which medical professionals can interpret model predictions easily.

Background

The Chronic Kidney Disease (CKD) is a situation described by a steady decrease in kidney work in the long run. CKD is otherwise called chronic renal disease.

The Chronic Kidney Disease

Chronic kidney disease entails diseases, which affect the kidneys by doing the jobs mentioned, and reduce their ability to guard you. In the event that your kidney condition deteriorates, waste will develop in your blood to elevated levels and cause you to feel sick. Abnormal conditions, for example, hypertension, anaemia (decreased blood count), powerless bones, poor stomach related wellbeing and harm to the nerves may happen. Also, the possibility of heart and vein disappointment is brought up by kidney illness. For an extensive stretch of time, these issues can happen gradually. Diabetes, hypertension and different problems can cause constant kidney infection. Therapy and early determination will likewise keep constant kidney sickness from getting deteriorating.

Advantages

- If you can no longer work full time because of your kidney disease, the SSA could consider you disabled and you will be able to receive Social Security disability benefits.

- Managing Other Health Problems

Disadvantages

Chronic kidney disease can cause other problems throughout your body including: Heart and blood vessel problems. Anemia (low red blood cell count) Bone problems.