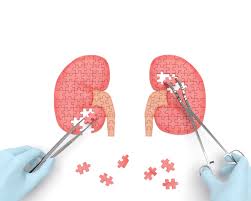
**Chronic Kidney Disease Analysis Using Machine Learning**

****

**DEV SINGH**

**YASH REDDY**

**AKSHAT BHARDWAJ**

**RISHITA SRIVASTAVA**

**INDEX**

1. INTRODUCTION
   1. OVERVIEW
   2. PURPOSE
2. LITERATURE SURVEY
   1. Existing Problem
   2. Proposed solution
3. THEORITICAL ANALYISIS
   1. Block Diagram
   2. Hardware/Software Designing
4. EXPERIMENTAL INVESTIGATION
5. FLOWCHART
6. RESULT
7. ADVANTAGES AND DISADVANTAGES
8. APPLICATIONS
9. CONCLUSION
10. FUTURE SCOPE

**INTRODUCTION TO PROJECT**

Chronic kidney disease, also called chronic kidney failure, describes the gradual loss of kidney function. Your kidneys filter wastes and excess fluids from your blood, which are then excreted in your urine. When chronic kidney disease reaches an advanced stage, dangerous levels of fluid, electrolytes and wastes can build up in your body. In the early stages of chronic kidney disease, you may have few signs or symptoms. Chronic kidney disease may not become apparent until your 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 can progress to end-stage kidney failure, which is fatal without artificial filtering (dialysis) or a kidney transplant.

Chronic kidney disease occurs when a disease or condition impairs kidney function, causing kidney damage to worsen over several months or years.

Diseases and conditions that cause chronic kidney disease include:

* Type 1 or type 2 diabetes
* High blood pressure
* Glomerulonephritis (gloe-mer-u-low-nuh-FRY-tis), an inflammation of the kidney's filtering units (glomeruli)
* Interstitial nephritis (in-tur-STISH-ul nuh-FRY-tis), an inflammation of the kidney's tubules and surrounding structures
* Polycystic kidney disease

**LITERATURE REVIEW**

Defined as a persistent abnormality in kidney structure or function (eg, glomerular filtration rate [GFR] <60 mL/min/1.73 m2 or albuminuria ≥30 mg per 24 hours) for more than 3 months, CKD affects 8% to 16% of the population worldwide. In developed countries, CKD is most commonly attributed to diabetes and hypertension. However, less than 5% of patients with early CKD report awareness of their disease. Among individuals diagnosed as having CKD, staging and new risk assessment tools that incorporate GFR and albuminuria can help guide treatment, monitoring, and referral strategies. Optimal management of CKD includes cardiovascular risk reduction (eg, statins and blood pressure management), treatment of albuminuria (eg, angiotensin-converting enzyme inhibitors or angiotensin II receptor blockers), avoidance of potential nephrotoxins (eg, nonsteroidal anti-inflammatory drugs), and adjustments to drug dosing (eg, many antibiotics and oral hypoglycemic agents). Patients also require monitoring for complications of CKD, such as hyperkalemia, metabolic acidosis, hyperphosphatemia, vitamin D deficiency, secondary hyperparathyroidism, and anemia. Those at high risk of CKD progression (eg, estimated GFR <30 mL/min/1.73 m2, albuminuria ≥300 mg per 24 hours, or rapid decline in estimated GFR) should be promptly referred to a nephrologist.

Only if providers recognize that their patients have CKD will the appropriate targeted management be initiated. Several investigators have demonstrated considerable under-recognition by primary care practitioners. De Lusignan and colleagues demonstrated that less than 4% of patients with CKD had been coded as having renal disease. Studies conducted by manual chart review (bypassing the known *International Classification of Diseases* (ICD)-9 coding sensitivity issues) demonstrated that over three-quarters of patients with CKD were not recognized as having CKD.

A first step in creating a tool to prompt early recognition of CKD is to determine if the provider has recognized the patient's CKD. The tool could search for appropriate documentation of CKD in the patient's notes as a proxy for recognition.

**THEORITICAL ANALYSIS**

Chronic Kidney Disease (CKD) is a 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

Following are the steps which are to be completed to complete this project

* Download the dataset
* Preprocess or clean the data
  + Import the libraries
  + Read the dataset
  + Analyze the dataset
  + Drop unnecessary columns
  + Change the column names
  + Remove the randomness in the columns
  + Find the missing values
  + Handle the missing values
  + Split the data into independent and dependent variables
  + Split the data to train and test
* Train  the machine with preprocessed data with an Appropriate Machine learning algorithm to build a model
* Save the model and its dependencies
* Build a Web application using flask that integrates with Model built.

Model is based on logistic regression, and it obtains the weight of each predictor and a bias. If the sum of the effects of all predictors exceeds a threshold, the category of the sample will be classified as ckd or notckd.

**EXPERIMENTAL INVESTIGATION**

There are several Machine learning algorithms to be used depending on the data you are going to process such as images, sound, text, and numerical values. The algorithms that you can choose according to the objective that you might have it may be Classification algorithms or Regression algorithms.

Example: 1. Linear Regression.

2. Logistic Regression.

3. Random Forest Regression / Classification.

4. Decision Tree Regression / Classification.

As the prediction for model is classification type, we apply a logistic regression algorithm on our dataset**.**

Logistic regression is the appropriate regression analysis to conduct when the dependent variable is binary.  Like all regression analyses, the logistic regression is a predictive analysis.  Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables.

Once the model is trained, it’s ready to make predictions.

Finally, there is a need to check to see how well our model is performing on the test data. There are many evaluation techniques are there. For this, we evaluate the accuracy score produced by the model. Confustion Matrix for the model is been used.

**FLOW CHART**

Diagram

Description automatically generated

**RESULT**

The sensitivities of the classifier and word-count methods were 95.4% and 99.8%, respectively. The specificity of both was 99.8%. Categorization of

individual patients as appropriately documented was96.9% accurate. Of 107 patients with manually verified moderate CKD, 32 (22%) lacked appropriate

documentation. Patients whose CKD had not been appropriately documented were significantly less likely to be on renin-angiotensin system inhibitors or have urine protein quantified, and had the illness for half as long (15.1 vs 30.7 months; p<0.01) compared to patients with documentation. The result of each classifier has been evaluated using Confustion Matrix and saved in the form of pickle form in order to use it in web application which is been made with the help of FLASK.

**ADVANTAGES AND DISADVANTAGES**

**Advantages:**

Logistic regression analysis is a statistical technique that describes the relationship between an independent variable (either continuous or not) and a dichotomy dependent variable (or dummy variable) (that is, a variable with only two possible values: 0=outcome absent and 1=outcome present). Hence, it eases the process.

The only limitation for this project is the dataset which is considered is quiet smaller in size.

**APPLICATIONS**

* Chronic Kidney Disease (CKD) is a major medical problem which has similar ramification as cancer.
* It can be cured if treated in the early stages.
* Attributes of various medical tests are investigated to distinguish which attributes may contain helpful information about the disease.
* This machine learning model is applicable in the field of medical science.

Machine learning refers to a computer program, which calculates and deduces the information related to the task and obtains the characteristics of the corresponding pattern. This technology can achieve accurate and economical diagnoses of diseases; hence, it might be a promising method for diagnosing CKD. It has become a new kind of medical tool with the development of information technology and has a broad application prospect because of the rapid development of electronic health record.

**CONCLUSION AND FUTURE SCOPE**

This work examines the ability to detect CKD using machine learning algorithms while considering the least number of tests or features. To approach this aim we have used logistic regression which is a machine learning classifier. In order to reduce the number of features and remove redundancy, the association between variables has been studied. Through this guided project I learnt how to implement Logistic regression in the real world problem and also understood the problem to classify if it is a regression or a classification kind of problem. Apart from this, it also provided in-depth knowledge about how to handle missing values and enlightened me with the information regarding Label encoding. I also learnt about Pickle library with which I was not familiar before.

After completing this, we can add more categories in this work, can make this more efficient. Using more classifiers on this dataset can get a better understanding on which classifier can be the best for this work.

**Video link:**