Early Detection of Chronic Kidney Disease Using Machine Learning

Prof. P. Vijayalakshmi
M.E., Ph.d.,
Associate Professor,
Department of Computer Science and Engineering
Knowledge Institute of Technology,
Salem.

Lakshmi S,
Student, Knowledge Institute of Technology,
Salem.

Dhanush A,
Student, Knowledge Institute of Technology,
Salem.

Ilamathi M,
Student, Knowledge Institute of Technology,
Salem.

Karthiban M J,
Student, Knowledge Institute of Technology,
Salem.

Ramanan R S,
Student, Knowledge Institute of Technology,
Salem.

Abstract - Chronic kidney disease is the worldwide affected disease, people are unaware of the disease and know the seriousness in 3rd stage, nowadays people with various other medical report which are taken for different purposes can contain the valid information regarding the kidney disease with that information they can give the inputs like age, blood urea, blood glucose random, presence of pus cell, anemia, pedal edema, coronary artery disease, appetite, diabetes mellitus, blood pressure, hypertension, white blood cell count, red blood cell count, specific gravity, sodium, potassium, pus cell clumps, bacteria. We used machine learning technique and build a model with datasets and trained the model using random forest classifier algorithm where the predictions will be given with more accuracy. In our website we added information about the disease to get knowledge about chronic kidney disease (CKD), and we included BMI Calculator through which the user gives their height and weight and calculate the body mass index and know their normal health condition. The symptoms are the main parameters collected from user through website and result will generated for them. By this user can identify whether they are affected by the disease or not.

Keywords - Random Forest Classifier, BMI, Health status.

1. INTRODUCTION
Chronic kidney disease is a progressive loss of kidney function over a period of months or years. Our kidneys work to keep us healthy by cleaning wastes from our blood with millions of tiny filters, called nephrons. If these nephrons are damaged, they begin to shut down. However, by the time we notice the symptoms, CKD is usually at an advanced stage. In fact, a person can lose up to 90% of their kidney function before experiencing any symptoms at all. Someone with CKD is at increased risk of heart attack or stroke, especially if they smoke or are overweight. This is why one in ten people are living with CKD, but most of them don’t even know it.

In this project we developed the machine learning model where the user gets to know their result, whether the user is affected by the disease or not. Therefore, it becomes easy to know their health status and making it better platform to use.

We are collecting 18 parameters from the user. With the help of the details obtained from the user, the model predicts and gives output whether the user is affected by the disease or not. The model is trained using the random forest classifier algorithm. The accuracy of the model using random forest classifier is predicted to be 98.75%.
1. PURPOSE

The main purpose of the website is, the user no need to reach out doctor for checkups and to create awareness about the disease which has been spread worldwide and to reduce the death rate. Since may of us unaware of the disease. Based on the result the user can take further measures regarding their health.

2. PROBLEM STATEMENT

To know our heath status is very important and to take necessary actions based on the result. Many of us may be affected from the disease and still unaware of the disease and know the effect in last stage, sometimes it may lead to death. On the other hand, the user need to reach out doctor to take scan reports and should wait patiently for their results.

3. PROPOSED SOLUTION

If the disease is identified earlier, it can be cured. To overcome this problem, we had developed a website and trained a model with the previous records and for the prediction we used random forest algorithm which gives the highest accuracy in prediction. Based on the given input given by the user the model predicts and display the result.

4. DATA FLOW DIAGRAM

The user will navigate into profile page and fills their basic information then the user have to fill the input regarding the disease. By collecting the user health condition and the data gets processed by the ML model and the user can get their result. The user will also get basic knowledge about the disease. In this application we also provided BMI calculator where the user can check body fat and health condition by giving height and weight as the input.
5. SOLUTION ARCHITECTURE

![Figure 5.1 Solution Architecture](image)

6. TECHNICAL ARCHITECTURE

![Figure 6.1 Technical Architecture](image)
7. FEATURES

7.1 FEATURE 1 – BMI Calculator
BMI which means Body Mass Index is the value generator based on the height and weight given by the user. BMI range from 18 - 24 is healthy, below 18 is underweight, above 24 is overweight and above 30 is obese. People can know their BMI range by using this calculator and concentrate on the health condition. BMI is calculator as:

$$BMI = \frac{\text{weight}}{(\text{height} \times \text{height})}$$

Figure 7.1.1 BMI Calculator

7.2 FEATURE 2 - Information about the disease
In this web page the user can get the detailed information about the chronic kidney disease and get knowledge about it. This page contains symptoms of CKD, risk factors, causes, Doctors view port on the diseases. With help of this page the user can know what are the symptoms that matches them.

Figure 7.2.2 Information Page

8. ALGORITHM USED

8.1 RANDOM FOREST CLASSIFIER ALGORITHM
Random Forest is one of the most popular and commonly used algorithms by Data Scientists. Random forest is a supervised machine learning algorithm that is used widely in classification problems. It builds decision trees on different samples and takes their majority vote for classification.

One of the most important features of the Random Forest Algorithm is that it can handle the data set containing categorical variables in the case of classification.
9. PERFORMANCE METRICS

Performance metrics measure the behaviour, activities and performance of a business. It measures the data that is required within a range that is in the form of data. This measures the performance which is the key target to check. The random forest classifier is a classification problem in which the data is identified based on the category. The given dataset is used by the model for training.

An important evaluation metric for the classification problem is the confusion matrix which is in the form of a table that contains binary classifiers of the outcome from the model prediction.

Confusion Matrix - [53, 1, 0, 26]
Accuracy Score – 98.75%

1 CONCLUSION

By using this application Early detection of Chronic Kidney Disease developed using Machine Learning classification algorithm Random Forest Classifier the output is predicted whether the user is affected by the disease or not.

In our application, the user can give their input like age, blood pressure, blood urea, blood glucose random, appetite, anemia, coronary artery disease, hypertension, pus cell, pus cell clumps, specific gravity, diabetes mellitus, pedal edema, bacteria, RBC count, WBC count, sodium, potassium. The model predicts based on the given input. The model predicted using Random forest classifier has an accuracy of 98.75%.

The application also provides information regarding the disease to get basic knowledge about the disease. The application also contains BMI calculator where the user have to give their height and weight as input to know their body fat.

11. FUTURE ENHANCEMENT

In the future, the application can be enhanced with the chatbots where the user can clarify the doubt with admin regarding the disease. In addition to that, we have collected blood group of the user in the profile page, we can mail them in case of any emergencies need of particular blood group to hospital. We will collect the feedback form and ratings form the user to know their thoughts about our website and the betterments of the website will be included based on the feedback given in the form.

12. SAMPLE OUTPUT

![Figure 8.1 Model Training](image)

![Figure 12.1 Home Page](image)
**Figure 12.3 Collecting User Details**

**Figure 12.4 Collecting Input Regarding Disease**

**Figure 12.5 Output**

We are happy to say that you are not affected from chronic kidney disease now, and it may possible in future. So always aware of it and concentrate in your food habits that you consume and maintain your good health.
REFERENCES