A Review on Predicting the Stages of Chronic Kidney Disease Using Machine Learning Techniques

Smita Patil¹, Dr. Savitha Chowdhary²

¹Department of CSE, Presidency University, Itagulpur, Rajanukunte, Bengaluru – 560064, India
²Department of CSE, Sir M Visvesvaraya Institute of Technology, Yelahanka, Bengaluru – 560064, India

Abstract: Kidney Disease refers to the loss of kidney function over the period, whose primary role is to filter the waste from the blood. The count of people suffering from the disease is increasing rapidly and if the person is not diagnosed at the earlier stage it may take the life of the patient. Otherwise, the patient has to undergo transplantation or dialysis. The stages of Chronic Kidney Disease are measured based on Glomerular Filtration Rate (GFR). Various machine learning algorithms are applied for predicting chronic kidney disease. The machine learning techniques namely SVM, Decision Tree, K-NN and Navie Bayes are analyzed.

Keywords: Chronic Kidney Disease, Machine Learning, Classification, Prediction.

1. Introduction

Machine Learning is a sub-domain of Artificial Intelligence. In Machine Learning we train our computer with the given dataset and when the new data comes, it tries to predict the output based on its previous experiences. Machine Learning algorithms build a model based on sample data known as training data to make predictions or decisions without being explicitly programmed to perform the operations.

The percentage of people suffering from kidney disease is increasing rapidly through the world, as there are no prior symptoms of the disease the number of people dies prematurely has also become high. If the disease is diagnosed to be at the higher stage, then the patient has to undergo dialysis or kidney transplant which may not be affordable to all in most of the developing countries. So it is necessary to consult the doctor and to be diagnosed at the early stage. Chronic Kidney disease is diagnosed through laboratory reports and e-GFR. If the disease is diagnosed at an early stage and if proper treatment is initiated at the right time, the patient can survive for a longer time. Once the disease is diagnosed to be in a higher stage, the condition becomes worse and the number of nephrons decreases dramatically.

Several Classification and feature selection methods are used for the detection and classification of CKD in the Literature. In General SVM, Decision Tree, Navie Bayes, K-NN, Random Forest and many other algorithms are used.

Factors for causing kidney disease are Diabetes, High Blood Pressure, Family History of kidney failure, high cholesterol, and kidney infections.

Chronic Kidney Disease are classified into five stages based on e-GFR number:
- Stage 1: Normal or high GFR (GFR > 90 ml/min)
- Stage 2: Mild CKD (GFR = 60-89 ml/min)
- Stage 3: Moderate CKD (GFR = 30-59 ml/min)
- Stage 4: Severe CKD (GFR = 15-29 ml/min)
- Stage 5: End Stage (GFR < 15 ml/min)

2. Literature Review

1) In [1], the author predicted the severity of chronic kidney disease using demographic and blood biochemical features. The Predictive models were built using machine learning techniques. The predictive models include Logistic Regression, Elastic Net, Lasso Regression, Ridge Regression, Support Vector Machine, Random Forest, XGBoost, Neural Networks, and K-NN. Each of these models was evaluated based on AU-ROC, Sensitivity, Specificity, Accuracy and Log-Loss and Precision.

2) In [2], the author used Deep Learning algorithms for automation of kidney function prediction and classification. The dataset i.e. ultrasound images have been collected from the hospital. Deep Learning approach is used for determining the estimated GFR and CKD status. Different augmentation schemes were applied to remove the peripheral region of the kidneys. Bootstrap aggregation was applied to avoid overfitting. The model developed has a classification accuracy of 85.6% higher from an experienced nephrologist.

3) In [3], the author proposed data mining techniques like boosting and rule extraction for analyzing kidney disease. Adaboost and LogitBoost algorithms are used for comparing the performance of classification and Ant-Miner is used to decline the rules. The proposed work has two aims: used a boosting algorithm for analyzing the performance to detect the CKD and deriving rules.

4) In [4], the author predicted chronic kidney disease using ultrasound imaging techniques. This technique is invasive, low-cost. The dataset was collected from Keelung Chang Memorial Hospital in Taiwan. The preprocessing technique called image inpainting and median filter was applied to remove noise. The feature extraction technique was applied to take out important
In [12], the author predicted chronic kidney disease using Deep Neural Networks. The dataset is collected from the UCI machine learning repository. The model showed better performance because it is implemented using the cross-validation technique. Different classification algorithms were applied and among all Deep Neural Network performed better.

6) In [6], the author downloaded the data from the UCI machine learning repository for diagnosing CKD. Correlation-based feature selection is used to select initial features. Different classifiers like Adaboost, SVM, K-NN and Navie Bayes were used. Four parameters are used to predict the accuracy of each classifier namely accuracy, precision, recall and f-measure. K-NN achieved the best result with a 98.1% accuracy rate.

7) In [7], the author presented computer-aided abnormality detection for a kidney. In this work, preprocessing was performed to remove noise and segmentation is applied to segment region of interest of kidney. Features like intensity histogram and Haralick are extracted from the segmented kidney region. In the first stage, a lookup table was used to classify normal and abnormal kidney images. In the second stage, SVM with Multilayer Perceptron used to classify the presence of stone or cyst in the kidney.

8) In [8], the author investigated the factors to find early detection of chronic kidney disease, the early process of treatment and to prevent complications resulting from the disease. The dataset is collected from the UCI machine learning repository. The proposed work deals with the diagnosis and prediction of disease. Machine Learning algorithms were built for classifying the training and testing results and to check the performance of each classifier.

9) In [9], the author discussed the prediction and classification of CKD using Random Subspace classification. In this work, six different classification methods are selected for the analysis and comparison of chronic renal failure. A Computer-based decision support system is used to predict chronic renal failure.

10) In [10], the author used two different machine learning tasks namely classification and clustering. In classification, a model is built to classify a patient with CKD or NONCKD. Various classification algorithms are used such as Logistic Regression, Support Vector Machine, Decision Tree, Adaboost, Random Forest, K-Means and Hierarchical clustering are used to analyze the data. The Random Forest and AdaBoost performs better than all other algorithms.

11) In [11], the author employed machine learning algorithms to extract decision rules from the patient’s data, which is used to diagnose chronic kidney disease. C4.5 decision tree algorithm is used to formulate a set of diagnosis rules for CKD. The primary indicator which needs to be monitored is serum creatinine levels and secondary indicators are pedaledema, hemoglobin, diabetes mellitus, and specific gravity.

12) In [12], the author performed the detection and classification of kidney disease using an automatic system. The system consists of five main steps. ROI segmentation, preprocessing, feature extraction, feature selection, and classification. A multilayered feed-forward neural network is used for classification. A classification rate of 97% was achieved.

13) In [13], the author collected the ultrasound kidney images from the hospitals. The images are manually cropped to get a region of interest in the kidney. The preprocessing technique is done on the cropped image. Feature extraction is done through run length texture and GLCM texture feature and K-NN is used to classify the image as normal or cystic kidney image.

### Table 1: Comparative Analysis of Different Techniques in terms of Accuracy

<table>
<thead>
<tr>
<th>S.No</th>
<th>Author Name</th>
<th>Techniques</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Jing Xiao, Ruifeng Ding, et al</td>
<td>Logistic Regression, Elastic Net, Lasso Regression, Ridge Regression, SVM, Random Forest, XGBoost, NN, K-NN</td>
<td>85.6%</td>
</tr>
<tr>
<td>2</td>
<td>Chin-Chi Kao, Chun-Min Chang, et al</td>
<td>Deep Neural Network</td>
<td>95%</td>
</tr>
<tr>
<td>3</td>
<td>Arif-Ul-Islam, Shamim H Ripon</td>
<td>AdaBoost, LogitBoost</td>
<td>99%, 99.75%</td>
</tr>
<tr>
<td>4</td>
<td>Chi-Jim Chew, Jun-Wen Pai, et al</td>
<td>SVM</td>
<td>83.74%</td>
</tr>
<tr>
<td>5</td>
<td>Sudipta Roy</td>
<td>IBK with Apriori Algorithm</td>
<td>99%</td>
</tr>
<tr>
<td>6</td>
<td>Zixian Wang, Joe Won Chung, et al</td>
<td>SVM</td>
<td>98.14%</td>
</tr>
<tr>
<td>7</td>
<td>K. Divya Krishna, V Akkala et al</td>
<td>SVM, Decision Tree Test-1 Dataset Test-2 Dataset</td>
<td>100%, 91.67%, 97.6%, 96.12%</td>
</tr>
<tr>
<td>8</td>
<td>Enes Celik, Muhammet Atalay and Adil Kondilogh</td>
<td>SVM, Decision Tree Test-1 Dataset Test-2 Dataset</td>
<td>98%, 96%, 95%, 91%, 90%, 60%</td>
</tr>
<tr>
<td>9</td>
<td>Ani R, Greeshma Sasi et al</td>
<td>ANN, KNN, Decision Tree, Random Subspace algo, LDA</td>
<td>81.5%, 78%, 76%, 90%, 93%</td>
</tr>
<tr>
<td>10</td>
<td>Manoj Reddy and John Cho</td>
<td>SVM with Linear Kernel, Random Forest, Adaboost, Logistic Regression, Decision Tree and SVM with RBF Kernel</td>
<td>98%, 96%, 95%, 91%, 90%, 60%</td>
</tr>
<tr>
<td>11</td>
<td>Alexander Arman Serpum</td>
<td>C4.5 Decision Tree</td>
<td>98.25%</td>
</tr>
<tr>
<td>12</td>
<td>Mariam Wagih Attia, Hossam El-Dim Moustafa</td>
<td>ANN</td>
<td>97%</td>
</tr>
<tr>
<td>13</td>
<td>Prema T Akkasaligar and Sunanda Biradar</td>
<td>K-NN</td>
<td>Normal - 92% Cystic - 85%</td>
</tr>
<tr>
<td>14</td>
<td>Made Sattria Wibawa, I made Dendi may Sanjaya</td>
<td>NB, KNN, SVM</td>
<td>95%, 95.8%, 95.8%</td>
</tr>
</tbody>
</table>

### References


