Chronic Kidney Disease Classification using Machine Learning Methods

Bhumi P Patel[1], Dr. Ankit K Shah[2]

[1] *Department of Applied Instrumentation, L.D. College of Engineering, Ahmedabad, Gujrat, India*
[2] *Department of Instrumentation and Control, L.D. College of Engineering, Ahmedabad, Gujrat, India*

ABSTRACT

The most significant health concern now is chronic kidney disease. In India and other nations of low- and middle-income, chronic kidney disorders are a major source of morbidity and an increase in mortality rates. worldwide all deaths of around 60 are caused by chronic diseases. In low- and middle-income nations, chronic illness deaths account for 80% of all fatalities worldwide. Chronic kidney disease is a potentially fatal condition that can be avoided with prompt, accurate prognosis and sensible safety measures. The purpose of the essay is to examine and comprehend a few methods for foretelling kidney disease. SVM and other classification approaches are used to categories chronic kidney disease (CKD). The MATLAB tool is utilised to detect the datasets in kidney illness in order to mine medical industry data.

Keyword: - *CKD, Feature selection, SVM, ANN*

---

1. INTRODUCTION

Around the world, kidney infection poses a serious threat to general health. Particularly in our nation, living is so expensive and people's monthly or daily wages are so low that patients didn't go to hospitals and clinics; instead, they could only take analgesics, and the majority of hospitals and clinics weren't prepared to manage Kidney problems early on [9].

This typically occurs to someone with healthy kidneys or to someone who has kidney issues. High blood pressure, diabetes, cardiovascular (heart and blood vessel) illness, and hereditary kidney failure histories are the key factors linked to the disease. When the kidneys are damaged and unable to filter blood, it is known as chronic kidney disease (CKD) [12].

In developed and developing nations alike, CKD is currently a severe issue. People are adopting a bad lifestyle in developing nations because of rising urbanization, which leads to diabetes and high blood pressure. While 15–25% of patients with diabetes pass away from renal disease, the rate of chronic kidney disease is rising in emerging nations due to urbanization [14].

The salts and minerals that circulate in the blood, including calcium, phosphorus, sodium, and potassium, are balanced by the kidneys. Additionally, they produce hormones that maintain bone density, help regulate blood pressure, and produce red blood cells. Advanced stages of chronic renal disease can cause our bodies to accumulate hazardous amounts of fluid, electrolytes, and wastes. It is challenging to diagnose chronic kidney disease in its early stages because there aren't many telltale signs or symptoms [13].

2. LITERATURE REVIEW

Zahid Ullah, and Mona Jamjoom, "Early Detection and Diagnosis of Chronic Kidney Disease Based on Selected Predominant Features", Hindawi, 2022 [1], In this research, a prediction-based methodology is used to detect and diagnose CKD patients, allowing for quick and accurate decision-making at an early stage. Numerous prediction models, including K-nearest neighbour (KNN), support vector machine (SVM), random forest (RF), and bagging, were trained using the preprocessed dataset in addition to a combination of feature selection and preprocessing techniques. In terms of accuracy, precision, sensitivity, F-measure, specificity, and area under the curve (AUC) score, the performance evaluation reveals greater dependability for all models.

Pankaj Chittora, Sandeep Chaurasia, Prasun Chakrabarti, Gaurav Kumawat, Tulika Chakrabarti, Zbigniew Leonowicz, Michał Jasiński, Łukasz Jasiński, Radomir Gono, Elżbieta Jasińska, And Vadim Bolshev, "Prediction of Chronic Kidney Disease - A Machine Learning Perspective", IEEE Access, 2021 [2], This article has examined the prediction of chronic kidney disease from this perspective. The UCI repository's chronic kidney disease dataset was used. In this study, seven classifier techniques were used, including artificial neural networks, C5.0, logistic regression, chi-square automatic interaction detectors, linear support vector machines with penalty L1 and L2 penalties, and random trees. The dataset also used the significant feature selection method. The results have been calculated for each classifier using the following methods: (i) full features; (ii) correlation-based feature selection; (iii) Wrapper method feature selection; (iv) least absolute shrinkage and selection operator regression; (v) synthetic minority over-sampling technique with least absolute shrinkage and selection operator regression selected features; and (vi) synthetic minority over sampling technique with full features.

Nikitha Saurabh, Tanzila Nargis, "Chronic Disease Prediction using Effective Feature Selection", IJRTE, 2019 [3], In this paper, Chronic renal disease is a potentially fatal condition that can be avoided with prompt, accurate prognosis and sensible safety precautions. In this study, a medical dataset is used to apply a variety of machine learning classifiers to create a prediction model that can determine if a person's current medical state will progress to the chronic stage of their disease in the future. By combining the Best First and Greedy stepwise algorithm with different classification techniques like Naive Bayes, Support vector machine (SVM), J48, Random Forest, and K Nearest Neighbour (KNN), a higher prediction accuracy and shorter build time are obtained with reduced feature set attributes.

Pawan Agarawal, Sanjay Kumar, "Chronic Kidney Disease Prediction using Classification techniques", JOURNAL OF CRITICAL REVIEWS, 2020 [4], In this study, the use of big data analytics and data mining in healthcare overcomes the difficulties of analyzing hidden information and extracting the useful information from the vast amount of data, such as patients' electronic health records (EHRs), which provides an intuition of predicting chronic kidney disease at the early stage. The purpose of the study is to use a classification algorithm on a structured dataset to predict chronic kidney disease. The prediction model uses a variety of classification techniques, including Logistic Regression, Random Forest, Naive Bayes, Decision Tree, Support Vector Machine, and K-Nearest Neighbours, to assess performance based on accuracy, precision, and f1-measure on the provided dataset. The classification algorithm's use in the prediction model is examined in this paper to assist doctors in making the best choices for patients' extended lifespans.

Reshma S, Salma Shaji, S R Ajina, Vishnu Priya S R, Janisha A, "Chronic Kidney Disease Prediction using Machine Learning", IJERT, 2020 [5], In this paper discussed about of Chronic renal Disease (CKD) is discovered through screening of individuals who are known to be at risk for renal issues, such as those with high blood pressure, diabetes, or a blood relative who has CKD. Therefore, in order to treat the disease effectively and combat it, early prediction is essential. The Ant Colony Optimization (ACO) method and the Support Vector Machine (SVM) classifier are two machine learning techniques that this study suggests be used for CKD.

Elias Dritsas and Maria Trigka, "Machine Learning Techniques for Chronic Kidney Disease Risk Prediction", MDPI, 2022 [6], In this paper discussed about of machine learning (ML) techniques are becoming more and more significant in the field of medical science and how they might be used to forecast diseases. Using an approach that makes use of ML techniques, the goal of the current research is to develop effective tools for forecasting the onset of CKD. To be more precise, we first use class balancing to address the non-uniform distribution of the instances in the two classes, after which features are ranked and analyzed, and lastly, a number of ML models are trained and assessed using a variety of performance measures.

Dibaba Adeba Debal and Tilahun Melak Sitote, "Chronic kidney disease prediction using machine learning techniques", Debal and Sitote Journal of Big Data, 2022 [7], In this paper discussed One of the goals is to minimize non-communicable disease-related premature mortality by one-third by 2030. Various studies on the early diagnosis of CKD have been conducted utilizing machine learning approaches. They weren't primarily concerned with predicting the individual stages. Both binary and multiple classifications for stage prediction have been done in this work. Three prediction models are employed: Decision Tree (DT), Support Vector Machine (SVM), and Random Forest (RF). For feature selection, analysis of variance and recursive feature removal with cross validation have been used.

3. RESEARCH METHODOLOGY

3.1 Feature selection

The process of choosing the most significant predictive features for use as model input is known as feature selection. To deal with the high dimensionality issue, preprocessing is a crucial step. Therefore, the primary goal of feature selection is to choose the subset of features that are pertinent and independent of one another for training the model. Similarly to this, choosing the right features is essential when building a model to predict chronic renal disease. This makes the model quicker, more efficient, and more accurate while reducing the dimensionality and complexity of the data. As a result, when the dataset was created, a feature selection method was utilised to choose pertinent features [7].

3.2 Artificial neural network

An ANN, or artificial neural network, is essentially a computational model. These computer models consist of a complicated network of fundamental elements or a collection of neurons-named nodes. There are connections between the nodes. Each link between the nodes has a particular weight attached to it. Any neural network's basic architecture consists of three layers. The first layer is referred to as the input layer, the second as the hidden layer, and the third as the output layer. These three layers with one or more nodes will undoubtedly be present in every neural network. One hidden layer will be present in a simple and fundamental neural network. However, a sophisticated neural network may contain numerous hidden layers. The initial input layers are where the inputs to the network are fed. The calculations are carried out in the hidden layers, and the corresponding output is revealed through the output layers. Neural networks come in a wide variety of varieties. Multiple hidden layers in neural networks are common today and are essential for many significant computational tasks [13].

3.3 Support Vector Machine (SVM)

A machine learning method based on the statistical learning concept is the support vector machine. The training data's descriptor space is transformed into a discrete hyperplane, and compounds are categorized according to where the hyperplane is placed [11].

The benefit of the SVM is that without explicitly transforming the original descriptors, the distance between a molecule and the hyperplane can be estimated in a changed (nonlinear) feature space using the so-called "kernel trick." For this investigation, the most popular radial basis function kernel (Gaussian kernel) was chosen. The following is how the kernel function is expressed [11]:

$$K(\bar{x}, \bar{x}_i) = \exp\left(-\frac{\|\bar{x}, \bar{x}_i\|^2}{2a^2}\right)$$

As mentioned equation the kernel width parameters control the amplitude of the Gaussian function reflecting the generalization SVM ability [11].

This approach, which works on classification and regression issues but is primarily used for classification problems, is the most well-known and significant supervised machine-learning algorithm. To separate marked information, SVM used a kernel function. One advantage of using parts in SVM is that it applies kernel dentitions to non-vector sources of information and allows kernels to be characterized based on a combination of different information kinds. Data are divided into two data points by the SVM algorithm when a hyper-plane is located between two branches. In the signifier space of the preparation information, SVM creates a discrete hyperplane, and mixes are organized according to the side of the hyperplane discovered [9].

4. FLOW OF SYSTEM

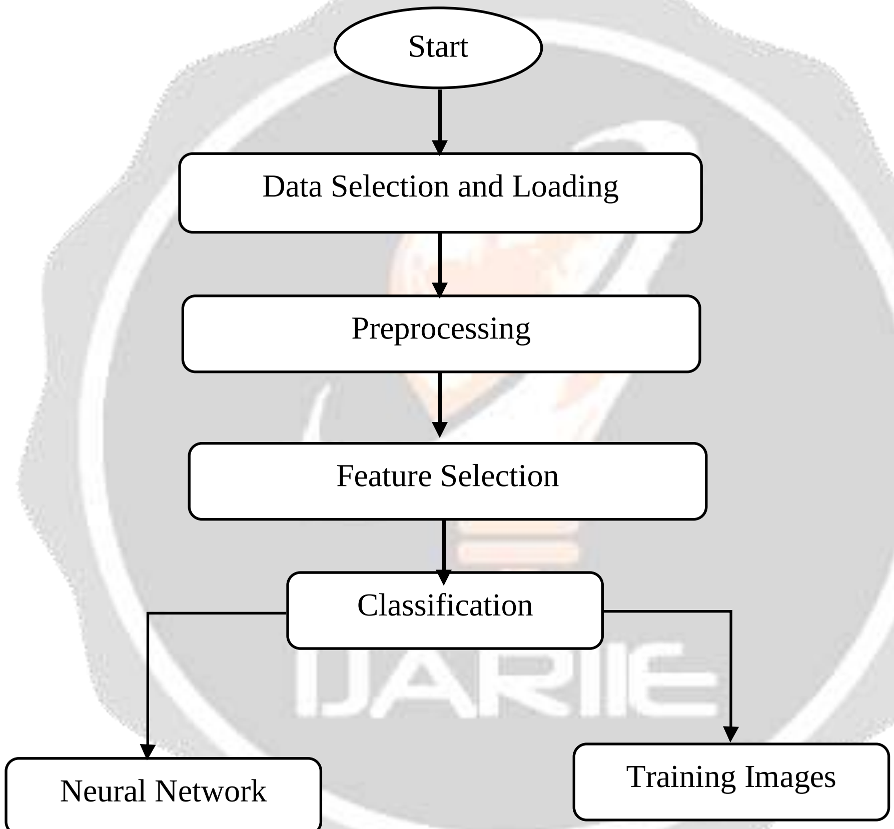
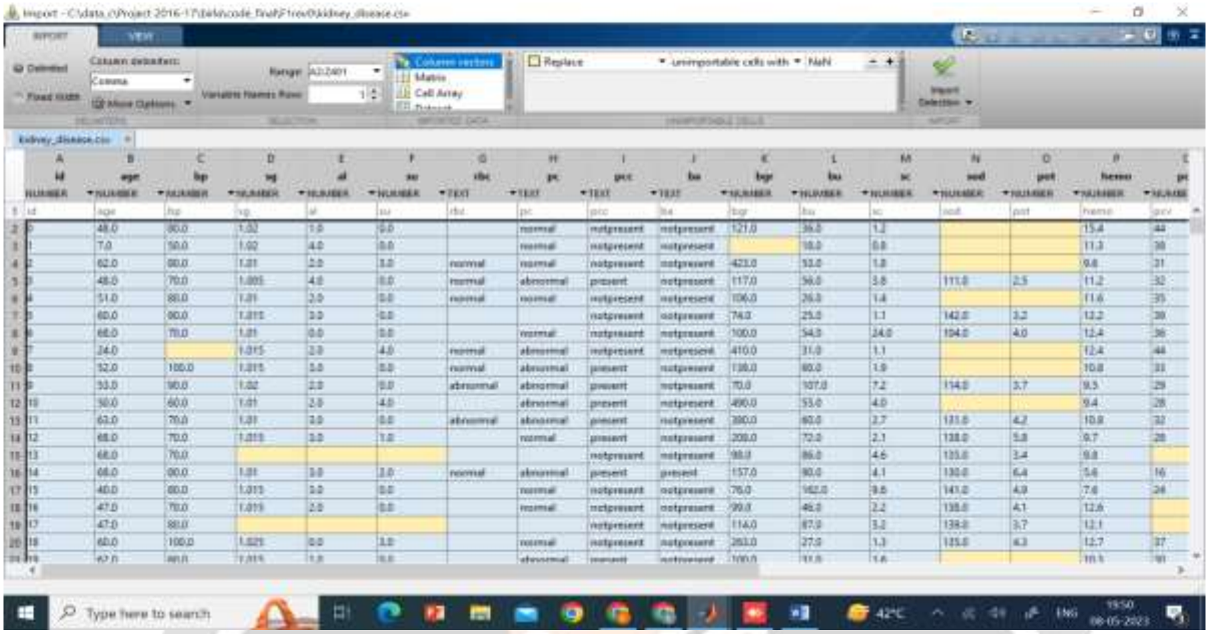
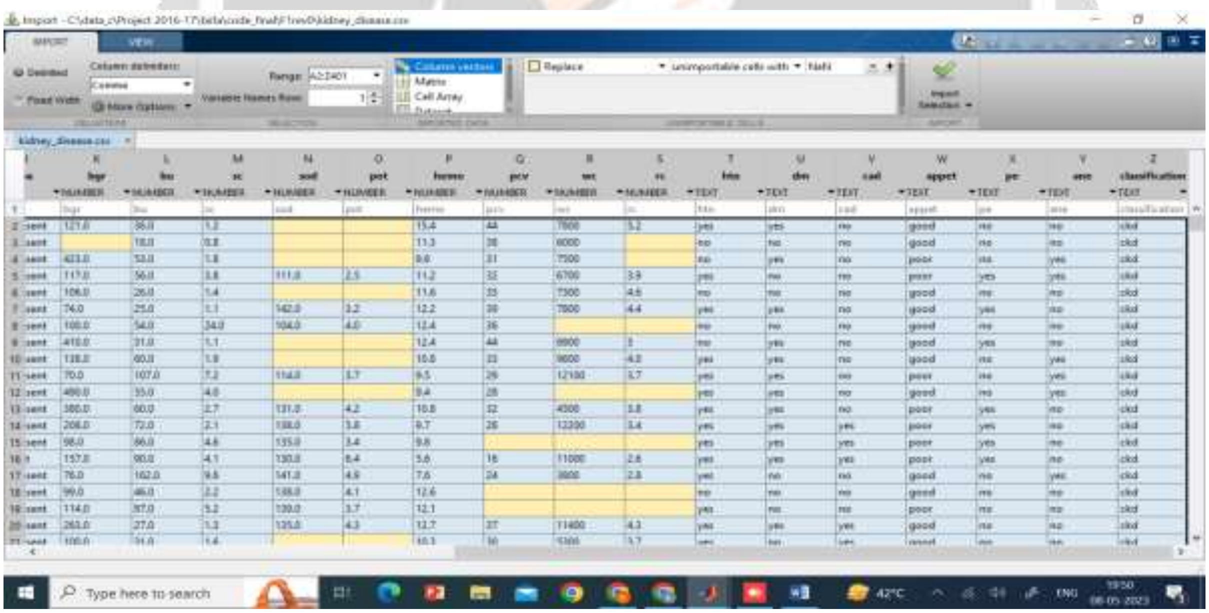


Figure 1: Flow of system

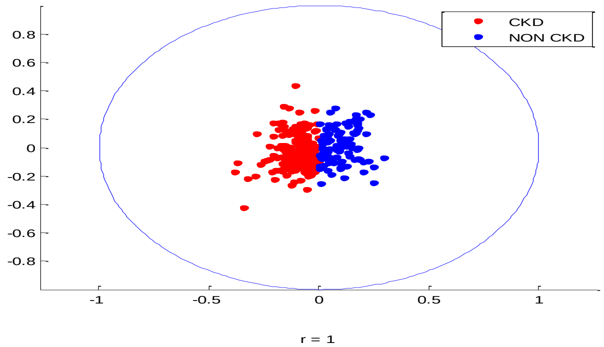
5. IMPLEMENTATION

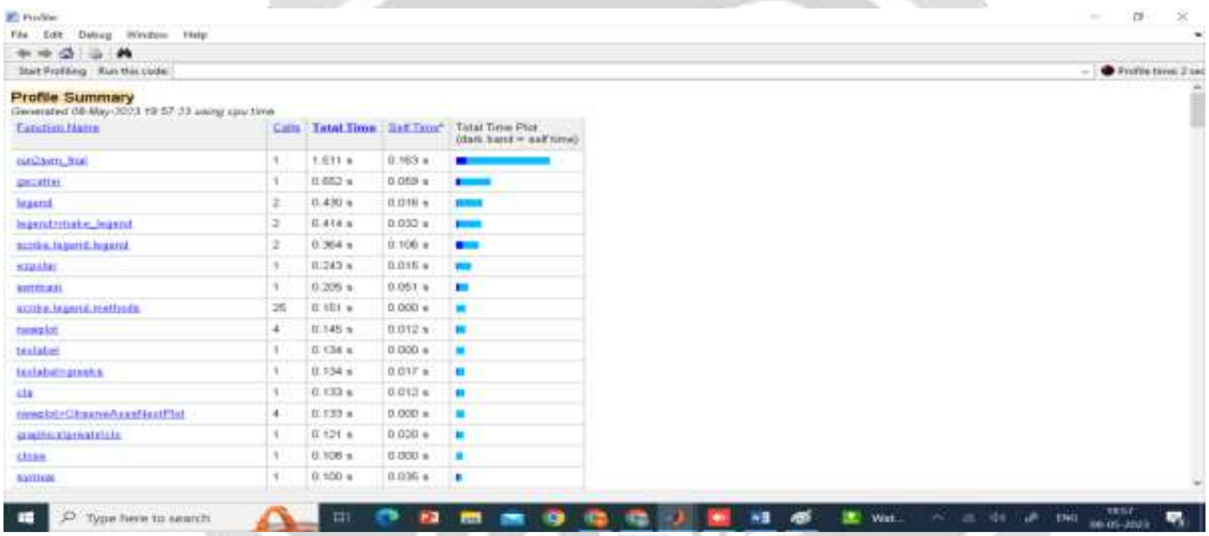




```
accuracy =

    0.9877
```





Result Analysis

Dataset	Algorithm	Accuracy
Kiggle Dataset	SVM	98.75
Kiggle Dataset	KNN	96.56
Kiggle Dataset	ANN	97.14

Timing Analysis

Dataset	Algorithm	Time
Kiggle Dataset	SVM	1.15 Sec
Kiggle Dataset	KNN	1.43 Sec
Kiggle Dataset	ANN	1.41 Sec

6. CONCLUSION

The survey of various algorithms and classifications is examined in this project, followed by the detection of kidney stones. As a result of this implementation, the limitations of the current system are deduced, and a new design is suggested to overcome them. For instance, level set techniques necessitate a lot of thought in order to construct velocities in order to produce an ideal advanced level set function. This implies that a lot of data should be available to determine the accuracy rate, which may not always be the case.

7. REFERENCES

[1] Zahid Ullah, and Mona Jamjoom, "Early Detection and Diagnosis of Chronic Kidney Disease Based on Selected Predominant Features", Hindawi, 2022.

[2] Pankaj Chittora, Sandeep Chaurasia, Prasun Chakrabarti, Gaurav Kumawat, Tulika Chakrabarti, Zbigniew Leonowicz, Michał Jasiński, Łukasz Jasiński, Radomir Gono, Elżbieta Jasińska, And Vadim Bolshev, "Prediction of Chronic Kidney Disease - A Machine Learning Perspective", IEEE Access, 2021.

[3] Nikitha Saurabh, Tanzila Nargis, "Chronic Disease Prediction using Effective Feature Selection", IJRTE, 2019.

[4] Pawan Agarawal, Sanjay Kumar, "Chronic Kidney Disease Prediction using Classification techniques", Journal of Critical Reviews, 2020.

[5] Reshma S, Salma Shaji, S R Ajina, Vishnu Priya S R, Janisha A, "Chronic Kidney Disease Prediction using Machine Learning", IJERT, 2020.

[6] Elias Dritsas and Maria Trigka, "Machine Learning Techniques for Chronic Kidney Disease Risk Prediction", MDPI, 2022.

[7] Dibaba Adeba Debal and Tilahun Melak Sitote, "Chronic kidney disease prediction using machine learning techniques", Debal and Sitote Journal of Big Data, 2022.

[8] Namrata Ghuse, Shravani Nirbhavane, Ashish Tayade, Shreya Gavhane, Jagruti Patil, "Prediction of Chronic Kidney Disease (CKD) Using Machine Learning", IJRMETS, 2022

[9] Deepak K.N, Adhwaidh P.S, Akshay P.D, Athira K.S, Jisna Jayan, "Survey on Chronic Kidney Disease Prediction System", IJARCCE, 2021.

[10] Pooja Mane, Naresh Thoutam, Neha Tiwari, Gauri Mandlik and Nutan Pandey, "Chronic Kidney Disease Prediction Using Machine learning", IJRPR, 2021.

[11] Dr. S. Vijayarani, Mr.S.Dhayanand, "Data Mining Classification Algorithms For Kidney Disease Prediction", IJCI, 2015.

[12] Jayalakshmi V, Lipsa Nayak 2 and 3 K.Dharmarajan, "A Survey on Chronic Kidney Disease Detection Using Novel Methods", International Journal of Pure and Applied Mathematics, 2018.

[13] S.Priya, S.Nirmal Kumar, G.Sibi Saravanan, E.Pandiyan, Ashuthosh Kumar Pandey, "Chronic Kidney Disease Prediction Using Neural Networks", IRJET, 2020.

[14] Hasnain Iftikhar, Murad Khan, Zardad Khan, Faridoon Khan, Huda M Alshanbari , and Zubair Ahmad, "A Comparative Analysis of Machine Learning Models: A Case Study in Predicting Chronic Kidney Disease", MDPI, 2023.