IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.047 ⁜ Vol. 10, Issue 5, May 2022

DOI: 10.17148/IJIREEICE.2022.10552

# Artificial Intelligence to Predict Chronic Kidney Disease (CKD)

**Sonia S B[1], Shwetha T[2], Smrithi L Shetty[3], Sujatha M[4]**

Assistant Professor, Department of Information Science & Engineering, Atria Institute of Technology, Bangalore, India[1]

Student, Department of Information Science & Engineering, Atria Institute of Technology, Bangalore, India[2,3,4]

**Abstract**: Chronic Kidney Disease (CKD) has emerged as a global health concern. Many people get diseases unexpectedly as a result of risk factors such as food, environment, and living standard. A normal kidney has one million filtering units. Glomerulas are the name given to each filtering unit. High blood pressure and diabetes both contribute significantly to kidney damage. The following kidney functions can be harmed by this disease. And they are: damaging their filtering units, collecting tubules, and causing scarring. A normal or healthy kidney can remove waste products from the blood and maintain an equal chemical level in the human body.

CKD is also defined as either kidney structural damage or a decrease in GFR of less than 60ml/min/1.73m2 for three months or more. It means that slow progressive loss of kidney functions over time caused by progressive destruction of renal mass. Because of this gradual loss of kidney function, CKD frequently goes undetected and undiagnosed until it worsens over time. Gradual kidney function loss can result in End Stage Renal Disease (ESRD), accelerated Cardio Vascular Disease (CVD), and death. This paper proposes a framework to predict the Chronic Kidney disease so that it reduces the time and effort required to detect and predict chronic kidney disease also to create a reliable analysis system. Using such models, most patients at risk of having proteinuria less than 1g/24 hrs can be classified as low risk and potentially treated slowly by their primary patient follow-up.

**Keywords**: Chronic Kidney Disease, Machine Learning, SVM, KNN, Random Forest, CKD

## I. INTRODUCTION

Chronic Kidney Disease cannot be predicted in the early stage since it has no sign or any symptoms. Chronic Kidney Disease impairs kidney health and causes waste to accumulate in the bloodstream. It causes a variety of symptoms, including high blood pressure, low blood count, bone deterioration, poor nutrition, nerve damage, and heart and blood vessel illness. As a result, it's crucial to detect kidney illness early on. Clinical data can be used to predict the kidney. The most critical stage of CKD is stage 5, when the kidneys are unable to perform the majority of their duties.

Diagnosis of CKD is a difficult task because there are no major symptoms that serve as a classification feature in detecting this disease. Accurate disease prediction not only informs patients about their health but also assists doctors in making medication recommendations well in advance. In today's world, advanced health knowledge and proper care can add a number of living days to a patient's life. ECG (Electrocardiogram) signals are now being used to detect the presence of CKD. The CT (computed tomography) scan images can then be used to identify the affected kidney.

Chronic failure can be treated with hemodialysis, peritoneal dialysis, or a kidney transplant. The ECG is capable of absorbing the dynamic changes seen by CKD patients. When a biopsy is performed, for example, the patient may be exposed to infection, surgical anxiety, and misdiagnosis. For many years, imaging investigations (mammograms, sonograms, and renal MRIs) have been utilised to detect the condition.

However, there are some drawbacks to employing them. The effects of radiation exposure are more specifically discussed. Imaging data is insufficient to diagnose CKD, in addition to being dangerous. Chronic kidney disease diagnosis is typically intrusive, costly, time-consuming, and hazardous. As a result, many patients, particularly in countries with low resources, reach late stages of the disease without receiving treatment. As a result, the disease's early detection method is still vital, especially in underdeveloped nations, where diseases are typically in late stages.

Finding answers to the aforementioned issues and overcoming the disadvantages has been a powerful motivator for conducting this research. Machine learning is an artificial intelligence application that allows computers to learn and develop without being explicitly designed. It's a branch of data science that focuses on developing algorithms that can learn from male data forecasts. Its algorithms are widely employed in the prediction and diagnosis of a wide range of essential diseases.

Machine learning has demonstrated success in delivering answers for early stage diagnosis in a range of medical areas by utilising computational methodologies. Data analysis is the process of examining, cleansing, manipulating, and modelling data with the objective of uncovering usable information, drawing conclusions, and assisting decision-making.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.047 ⁜ Vol. 10, Issue 5, May 2022

DOI: 10.17148/IJIREEICE.2022.10552

Data analysis has many facets and approaches, encompassing various techniques under various names and used in various business, science, and social science domains.

Prediction begins with identifying symptoms in patients and then distinguishing sick patients from a large number of sick and healthy ones. Predictive analytics for healthcare using machine learning is a difficult task that can assist doctors in determining the best treatment to save lives. In this paper, we collected a few samples from a public hospital and analysed selected fields to create a CKD prediction model.

To improve the statistical analysis of given data, data analysis and visualisation are used. The data is subjected to logistic regression because it has a large number of columns with categorical values. The model's accuracy, precision, and f1 score were all evaluated. This independent dataset can be used to infer a variety of conclusions, and it can be saved as historical data for future examination. This paper shows how to predict CKD using machine learning algorithms using best prediction model.

## II. LITERATURE SURVEY

[1] "Chronic Kidney Disease Analysis Using Data Mining Classification Techniques"

The authors for this survey paper is Veenita Kunwar, Khushboo Chandel, a Sai Sabitha and Abhay Bansal

In this case-study Data-Mining functionality comes into process, as it is been a major trend for obtaining diagnostic results. The main objective of data-mining is to understand, analyse and conceptualize the hidden information from a large-set of data, classifying them into various categories and sub-categories and understanding different patterns and learning those patterns to understand the diagnostic results more efficiently and effectively, such that effective decision-making would be attained in the process, such data are called as unmined data in the language of Data-mining. i.e.,classification, regression, analysis based on association, classification, and clustering.

[2] "Detection of Chronic Kidney Disease and Selecting Important Predictive Attributes"

In this case study we consider 24 predictive attributes or parameters and create a classification model which is used to categorize the data into classes, using Machine Learning algorithms to determine Chronic Kidney Disease(CKD). This feature selection determines the most significant attribute which would play a major role in analysing and detecting CKD in first place and adding that attribute functionality in the priority list. They also discover new attributes or sub-attributes which is not used by any previously available functionality like Glomerular Filtration Rate(GFR) estimator etc. The main objective of this research paper-work mainly focuses on discovering attributes related to CKD and introduce a ML approach which is cost-effective and detect the kidney disease at a very early stage of diagnosis instead of GFR estimator equation.

[3] "Predictive Analytics for Chronic Kidney Diseade Using ML(Machine Learning) Techniques"

This technique include four major Machine Learning methods or algorithms that are used, and are as follows: (SVM)Support Vector Machine, (KNN)K-Nearest-Neighbors, (LR) Logistics Regression, Decision Tree Classifier. And eventually this dataset which is collected is been categorized into training and testing. And accuracy of the historical data and the currently obtained data is been compared recursively and aggressively and difference obtained from this comparsion, is been noted in-accordance for further more comparsion. Then it is compared among the four predictive models to obtain the best classifier and exceeds the accuracy rate from the other three techniques, then they take that one technique which has exceeded the accuracy level and is selected for further detection of CKD(Chronic Kidney Disease).

[4] "Extraction of Action Rules for Chronic Kidney Disease using Naïve Bayes Classifier"

The author of this paper-work are: Dr Uma N Dulhari and Mohammad Ayesha

Chronic Kidney Disease is mainly due to the decreased funcionality rate of kidney, or findings of damage in kidney or in the decrease in the Glomerular Filtration rate would eventually be affected with CKD, which is also known as (CRD) Chronic Renal Disease. In this proposed system, the extraction of Action Rules becomes significant as it is like a production rules which is been exhibited by equipments or the model performance through which necessary steps and treatments to be followed in a more detailed and specific manner such that to avoid the symptoms showcasing the chances of CKD into next stage of diagnosis.

[5] "Dietary Prediction for patients with CKD by considering Blood Potassium Level using Machine Learning Algorithm"

The authors of this paper-work includes: M P N M Wickramasinghe, D M Perera, KA D C P Kahandawaarachchi.

The main objective of this research paper includes finding an appropriate diet plan for patients affected with CKD by applying specific classification algorithm on the medical test records of the patients. The diet-plan for patients can be obtained by analysing and predicting potassium zone-level in CKD patients in-accordance with the Blood Potassium level using classification algorithm. This experiment is been performed using different algorithms such as: Multiclass Logistic Forest, Multiclass Neural Network, Multiclass Decision Jungle, Multiclass Logistic Regression. Through this

DOI: 10.17148/IJIREEICE.2022.10552

comparsion between these algorithms carried out, Multiclass Decision Forest algorithm provides better accuracy rate and result.

[6] "Chronic Kidney Disease Prediction on Imbalanced data by MultiLayer Perceptron"
This paper-work mainly focus on the essence of imbalancement in training data, if found it could be erected through certain techniques of action rules, hence they developed a Neural Network classifier for detection of CKD using decision-making aspect as an attribute. Neural Network is used in wide number of applications including Decision systems and Data-Mining etc. Backpropagation technique in neural network is one of the important and significant aspect of which is used to train the dataset to identify and analyse different unusal patterns. Sampling algorithms eventually improves the performance rate of the classification algorithm, and learning rate algorithm is one the most abstract parameter, which would affect the Multi-Layer Perceptron.

[7] "A Novel Imputation Method for Effective Prediction of Coronary Kidney Disease"
The authors for this paper-work includes: S Dilli Arasu, Dr R Thirmumalaiselvi.
Prediction of CKD is sometimes a highly complicated task, as handling multiple dataset of numerous patients medical records and identifying and analysing would be time-consuming and sometimes irrelevant at-times, it would be more difficult if there is any missing field found in the patient's medical records which would be majorily important for the accuracy rate of the disease predictability, Hence through this technique of value imputation, the missing values from the patients records can be eventually predicted in no time with the help of Regression tree, classification and RF etc.
There is a proposed model WAELI algorithm which usually predicts the missing value.

[8] "Preemptive Diagnosis of Diabetes Mellitus Using Machine Learning"
This case-study basically revolves around the study of disease variants in various regions in-accordance with their ethnic groups, they have a different pattern of development of CKD depending upon the region. This diagnosis using DM(Diabetes Mellitus) is used to study about the development of disease from a region which has never been explored before. In this case-study in co-relation to this technique in King Fahd University Hospital, Saudi Arabia from this hospital the data has been retrieved and it undergoes certain pre-processing technicalities for the classification or the identification process.Through this experimental approach, the result obtained was ANN outperformed SVM, K-Nearest & Naïve bayes with an accuracy of testing of 77.5%. This experimental setup was carried in Weka software tool.

[9] "Analysis of Chronic Kidney Disease Dataset by Applying Machine Learning Methods"
The authors for this case-study includes: Yedilkhan Amiragaliyev, Shahrira Shamiluulu, Azamat Serek.
Chronic Kidney Disease is spread through several risk factors such as environmental imbalance, live-style and food consumption habits, these are the factors through which CKD could be spread without understanding the condition. In this case-study the patients who are affected by CKD are been classified using clinical trials on the basis of Support Vector Machine (SVM) Algorithm. Using this algorithm, the experimental result obtained was 93% of the success rate in classifying patients with CKD on the basis of 3 important performance metrics which was the f1 measure i.e., Specificity, Sensitivity and Accuracy.

[10] "Chronic Kidney Disease(CKD) Diagnosis using Multi-Layer Perceptron Classifier"
The author for this case-study includes: Shubham Vashisth, Ishika Dhall, Shipra Sarawat.
In this research-paper this classifier uses a fully-connected Deep Neural Network to determine or predict whether the patient suffers from CKD or not. This model is trained dataset of around 400 patients, and by observing and understanding various symptoms exhibited by CKD affected patients is been depicted or observed in a graph, out of which 250 patients are affected with CKD while the remaining 150 have not, hence attributes such as age, BP level, albumin level, RBC and WBC count level etc, by using these indicators the resultants get displayed in the graph and appropriate accuracy percentage levels can be obtained for the same and SVM and Naïve Bayes Classifier has surpassed and exceeded the accuracy levels in detecting the CKD.

## III. MACHINE LEARNING ALGORITHMS TO PREDICT CHRONIC KIDNEY DISEASE (CKD)

Machine Learning algorithms are capable of learning from the info we tend to provide. As new data is provided, the model's accuracy and potency to create choices improve with the resulting training.
Here, are few machine learning algorithms we used to predict chronic kidney disease.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.047 ⁜ Vol. 10, Issue 5, May 2022

DOI: 10.17148/IJIREEICE.2022.10552

### A. SUPPORT VECTOR MACHINE

A linear model for classification and regression is Support Vector Machine (SVM) which will be used to solve each linear and nonlinear problem. The algorithm classifies knowledge victimization as a hyperplane. during this algorithm, each data item is premeditated in an n-dimensional area (where n is the variety of features). The worth of every feature is the value of an explicit coordinate. Classification will be performed by finding the proper hyper-plane which might differentiate the 2 categories efficiently.

### B. K NEAREST NEIGHBOUR

K-Nearest Neighbour is one in all the best Machine Learning algorithmic rules supported the supervised Learning technique. It assumes the similarity between the new case/data and obtainable cases and place the new case into the class that's most kind of like the available categories. It can be used for Regression further as for Classification however largely it's used for Classification problems. it's a non-parametric algorithmic rule, which suggests it doesn't create any assumptions on underlying information. it's conjointly known as a lazy learner algorithm as a result of it does not learn from the coaching set straight off instead it stores the dataset and at the time of classification, it performs an action on the dataset. KNN algorithm at the training part simply stores the dataset and once it gets new data, then it classifies that data into a class that's abundant kind of like the new data.

### C. DECISION TREE

Decision Tree one of the algorithm, is used to solve regression and type problems. The fashionable objective of the use of the Decision Tree is to create a version that predicts instructions or values of goal variables via way of means of producing choice guidelines derived from schooling statistics sets. The decision tree algorithm follows a tree shape with roots, branches, and leaves. The attributes of choice-making are the inner nodes and class labels are represented as leaf nodes. The Decision Tree algorithm is simple to apprehend in comparison with different types of algorithms.

### D. RANDOM FOREST

The random forest rule constructs multiple call trees as an associate ensemble of categorification and regression processes. style of decision trees employs random subsets of the employment data sets, an outsized assortment of decision trees offers higher accuracy of results. The runtime of the algorithmic rule is comparatively fast and collectively accommodates missing information. Random forest randomizes the algorithm and not the coaching data set. the selection class is the mode of classes generated by decision trees..

## IV. CONCLUSION

In this paper, chronic kidney disease prediction models are proposed using some machine learning algorithms. The data set was taken from a trusted source and pre-processed it accordingly. The data being cleaned was divided into train and test data wherein 70% of the data was taken for training and remaining considered for test. We have calculated the accuracy from each models and were happy to conclude the accuracy is 97.50%. This model gives the high reliability, accuracy and precision for any samples.

## ACKNOWLEDGMENT

We would like to take this opportunity to thank our guide **Mrs. Sonia S B**, for giving us all the help and guidance needed. We are grateful to him for his kind support and valuable suggestions.

## REFERENCES

[1]. Sahil Sharma, Vinod Sharma, Atul Sharma, “Performance Based Evaluation of Various Machine Learning Classification Techniques for Chronic Kidney Disease Diagnosis”, July 18, 2016.
[2]. Venita Kunwar, Khushboo Chandel, A. Sai Sabitha, Abhay Bansal. “Chronic kidney disease analysis using data mining classification techniques” 2016 IEEE.
[3]. Yedilkhan Amirgaliyev, Shahriar Shamiluulu, Azamat Serek. “Analysis of Chronic Kidney Disease Dataset by Applying Machine Learning Methods”, 2018 IEEE.

DOI: 10.17148/IJIREEICE.2022.10552

[4]. Reem A. Alassaf, Khawla A. Alsulaim, Noura Y. Alroomi, Nouf S. Alsharif, Mishael F. Aljubeir, O. Olatunji, Alaa Y. Alahmadi, Mohammed Imran, Rahma A. Alzahrani, Nora S. Alturayeif "Preemptive Diagnosis of Diabetes Mellitus Using Machine Learning".

[5]. Shubham Vashisth, Ishika Dhall, Shipra Sarawat "Chronic Kidney Disease (CKD) Diagnosis using Multi-Layer Perceptron Classifier" 2019.

[6]. Anusorn Charleonnan, Thipwan Fufaung, Tippawan Niyomwong, Wandee Chokchueypattanakit, Sathit Suwannawach, Nitat Ninchawee, "Predictive Analytics for Chronic Kidney Disease Using Machine Learning Techniques".

[7]. Gunarathne W.H.S.D, Perera K.D.M, Kahandawaarachchi K.A.D.C.P, "Performance Evaluation on Machine Learning Classification Techniques for Disease Classification and Forecasting through Data Analytics for Chronic Kidney Disease (CKD)"

[8]. Pinar Yildirim, "Chronic Kidney Disease Prediction on Imbalanced Data by Multilayer Perceptron: Chronic Kidney Disease Prediction".

[9]. Asif Salekin, John Stankovic, "Detection of Chronic Kidney Disease and Selecting Important Predictive Attributes".

[10]. S.Dilli Arasu and Dr. R.Thirumalaiselvi, "Review of Chronic Kidney Disease based on Data Mining Techniques", International Journal of Applied Engineering Research**".**