# “PREDICTION OF CHRONIC KIDNEY DISEASE - A MACHINE LEARNING PERSPECTIVE”

A Project Abstract submitted to

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, ANANTAPUR.

In partial fulfillment of the requirements for the award of the degree of

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE AND ENGINEERING**

Submitted by

|  |  |
| --- | --- |
| **Sagun Babu Adhikari** | **(18121A05B9)** |
| **Roshan Kumar Sah** | **(18121A05B8)** |
| **Raju Adhikari** | **(18121A05C0)** |
| **K S M S Sai Saran** | **(18121A0591)** |

Under the Supervision of

**Ms. K. Vidyavathi,** M.Tech., (Ph.D)

**Assistant Professor**

Dept. of Computer Science and Engineering

Department of Computer Science and Engineering

### **SREE VIDYANIKETHAN ENGINEERING COLLEGE**

**(Autonomous)**

(Affiliated to JNTUA, Anantapuramu and Approved by AICTE, Accredited by NBA & NAAC)

Sree Sainath Nagar, Tirupati – 517 102

**2021-2022**

# ABSTRACT

# Chronic kidney disease (CKD) has become a major issue with a steady growth rate. They are now a cause of global morbidity and mortality even in developing countries. According to the report published by researchers and medical professionals from the Department of Nephrology, the All India Institute of Medical Sciences, and the Director-General Health Services, Ministry of Health and Family Welfare, Government of India The approximate prevalenceof CKD is 800 per million population. Early detection and characterization are considered to be critical factors in the management andcontrol of chronic kidney disease. Using efficient data mining techniques can reveal and extract hidden insights from clinical and laboratory patient data, which can be helpful to physicians toidentify disease severity stage with maximum accuracy. In the domain of healthcare, this paper aims in building a model for risk level prediction in CKD considering all of the symptoms and causes contributing to it. The symptoms are the attributes that will define different stages of kidney diseases. Based on the different stages, one can classify a set of patient records to identify towhich class of kidney disease a patient may belong. Classifying patients results in easy recognition of the dominant attributes of CKD. Certain solutions can be provided with respect to the dominant attributes to avoid the progression of CKD. To construct a model on risk prediction of kidney disease, various machine learning techniques canbe applied and then their performance can be compared with respect to the accuracy, specificity and sensitivity of the models. Before the application of any machine learning technique, there is a need of doing feature selection to understand the dominant attributes. A feature selection method called random forest is used to achieve the selection of dominant attributes. This paper is mainly concerned with the use of machine learning techniques namely neuro-fuzzy systems andclustering which is termed unsupervised learning.

**Introduction**

Our Chronic Kidney Disease Prediction model makes use of machine learning to achieve quick, easy and most importantly, accurate predictions of whether or not a person has the disease basedon the symptoms input by the user.

**Problem Statement**

This work proposes a workflow to predict CKD status based on clinical data, incorporating data prepossessing, a missing value handling method with collaborative filtering and attributes selection. The project also considers the practical aspects of data collection and highlights the importance of incorporating domain knowledge when using machine learning for CKD status prediction.

# Objectives

# The objective of this project is to develop a model for risk level prediction in CKD while considering all of the symptoms and causes involved. Because of the building up of waste due to this disease, treatment can help, but this condition can't be cured. Lab tests or imaging is always required, andin later stages, filtering the blood with a machine (dialysis) or a transplant may be required. Due tothis, CKD is a substantial financial burden on patients, healthcare services, and the government.Now, using the advancement of technology, Machine Learning Algorithms are used to detect and predict diseases with more accuracy. The symptoms will be the criteria defining the different stagesof kidney disease. By identifying the different stages, one can identify the class of kidney disease apatient might suffer from by categorizing their records.

# **Drawbacks of the previous model**

# Fluffy Logic has been created till now for the arrangement of patients with CKD, but there are a few classifiers that don't have to fit the informational index in the unique situation. Information mining procedures are utilized to arrive at one specific resolution that relates to the qualities of patients of various types who have Kidney infections. Some of the machine learning approaches that are being considered, do not stand viable for a large volume of data. On analyzing the currently available models, we found that some of the machine learning approaches that are being considered, do not stand viable for a large volume of data. There are also a few classifiers that don't have to fit the informational index when the unique situation arises.

**Advantages of the proposed system**

Detection of CKD earlier helps to save the patient from damaging the kidney and decreases the risk of heart attacks and strokes. CKD detection is less time consuming and helps the doctors to start the treatments early for the patients. Further, it helps to diagnose more patients within a less time period. The five stages (data preprocessing, feature selection, classifier application, SMOTE and analyzing the perfomance of the classifier) are used in this procedure to achieve the accuracy of the algorithm to detect CKD at higher level.

# Hardware and Software Requirements

# Hardware (minimum):

|  |  |
| --- | --- |
| **CPU** | Intel Core i3 8th gen |
| RAM | 4 GB |
| STORAGE | 5 GB HDD |

Software:

* VS Code
* Numpy
* Spyder
* Jupyter

Libraries:

* Pandas
* Matplotlib
* Seaborn
* Plotly Express

# References

1. *Q.-L. Zhang and D. Rothenbacher, ‘‘Prevalence of chronic kidney diseasein population-based studies: Systematic review,’’ BMC Public Health,vol. 8, no. 1, p. 117, Dec. 2008.*
2. *W. M. McClellan, D. G. Warnock, S. Judd, P. Muntner, R. Kewalramani,M. Cushman, L. A. McClure, B. B. Newsome, and G. Howard, ‘‘Albuminuria and racial disparities in the risk for ESRD,’’ J. Amer. Soc. Nephrol., vol. 22, no. 9, pp. 1721–1728, Aug. 2011.*
3. *M. K. Haroun, ‘‘Risk factors for chronic kidney disease: A prospectivestudy of 23,534 men and women in Washington County, Maryland,’’ J. Amer. Soc. Nephrol., vol. 14, no. 11, pp. 2934–2941, Nov. 2003.*
4. *W. D. Souza, L. C. D. Abreu, L. G. D. SilvaI, and I. M. P. Bezerra,‘‘Incidence of chronic kidney disease hospitalisations and mortality in Espírito Santo between 1996 to 2017,’’ Wisit Cheungpasitporn, Univ. Mississippi Medical Center, Rochester, MN, USA, Tech. Rep., 2019, doi:10.1371/journal.pone.0224889.*
5. *W. Mula-Abed, K. A. Rasadi, and D. Al-Riyami, ‘‘Estimated glomerularfiltration rate (eGFR): A serum creatinine-based test for the detection ofchronic kidney disease and its impact on clinical practice,’’ Oman Med. J.,vol. 27, no. 4, pp. 339–340, 2012.*
6. *A. S. Levey, D. Cattran, A. Friedman, W. G. Miller, J. Sedor, K. Tuttle,B. Kasiske, and T. Hostetter, ‘‘Proteinuria as a surrogate outcome in CKD: Report of a scientific workshop sponsored by the national kidney foundation and the US food and drug administration,’’ Amer. J. Kidney Diseases, vol. 54, no. 2, pp. 205–226, Aug. 2009.*
7. *S. Gerogianni, ‘‘Concerns of patients on dialysis: A research study,’’Health Sci. J., vol. 8, no. 4, pp. 423–437, 2014.*
8. *J. R. Chapman, ‘‘What are the key challenges we face in kidney transplantation today?’’ Transplantation Res., vol. 2, no. S1, pp. 1–7, Nov. 2013*