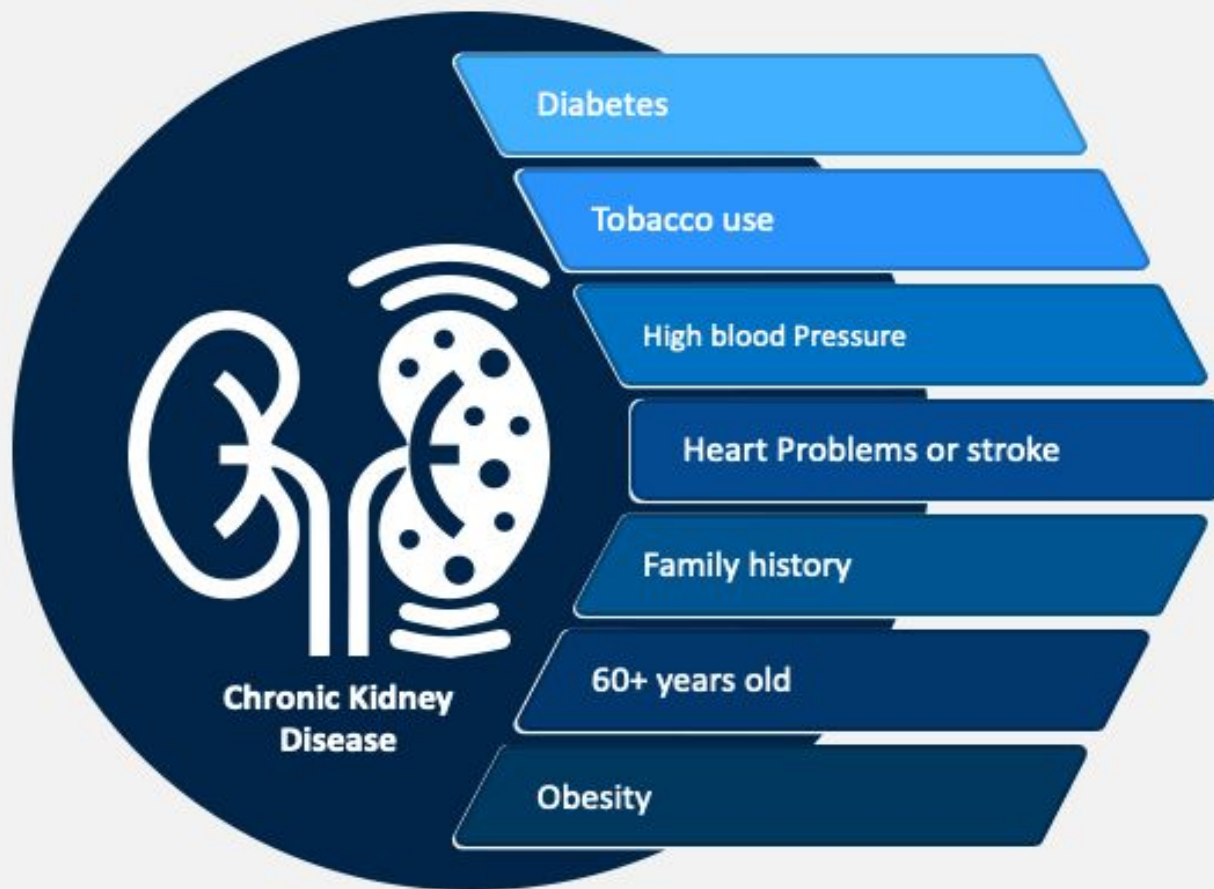


Early Detection of Chronic Kidney Disease using Machine Learning.

Srinidhi R	211719104136
Sivaranjani MG	211719104133
Sowmiya E	211719104134
Swetha JK	211719104148
Swetha S	211719104149

Introduction

Chronic Kidney Disease (CKD) is one of the deadliest diseases that slowly damages human kidney. The disease remains undetected in its early stage and the patients can only realize the severity of the disease when it gets advanced. Hence, detecting such disease at earlier stage is a key challenge now. Machine Learning is one of the emerging field used in the health sectors for the diagnosis of different diseases. In this paper, we compute, analyze and compare between Machine Learning classification approaches to determine which classification approach is the optimal for the prediction of CKD. K Nearest Neighbor Classifier, Decision Tree Classifier, Logical Regression and XGBoost are some renowned machine learning methods which were selected to train the model and based on these results, we can compare and determine which among the following Machine Learning Methods can predict the possibility of CKD at the most accurate level.



Paper - 1

Jager K.J., Kovesdy C., Langham R., et al. A single number for advocacy and communication-worldwide more than 850 million individuals have kidney diseases. *Kidney Int.* 2019;96:1048–1050

Chronic kidney disease (CKD) has emerged as one of the most prominent causes of death and suffering in the 21st century. Due in part to the rise in risk factors, such as obesity and diabetes mellitus, the number of patients affected by CKD has also been increasing, affecting an estimated 843.6 million individuals worldwide in 2017.

On the basis of the results of studies examining the global prevalence of CKD, the current total number of individuals affected by CKD stages 1–5 worldwide was estimated to be 843.6 million.

A Machine Learning Methodology for Diagnosing Chronic Kidney Disease by Jiongming Qin, Lin Chen, Yuhua Liu, Chuanjun Liu, Changhao Feng, Bin Chen

six machine learning algorithms (logistic regression, random forest, support vector machine, k-nearest neighbor, naive Bayes classifier and feed forward neural network) were used to establish models. Among these machine learning models, random forest achieved the best performance with 99.75% diagnosis accuracy. By analyzing the misjudgments generated by the established models, This study proposed an integrated model that combines logistic regression and random forest by using perceptron, which could achieve an average accuracy of 99.83%

Paper -3

Diagnostic decision support system of chronic kidney disease using support vector machine
Mubarik Ahmad, Vitri Tunjungsari, Dini Widiанти, Peny Amalia, Ummi Azizah Rachmawati

This study uses support vector machine(SVM) to predict the classify the dataset. This study resulted in a system that can detect a chronic condition of kidney disease based on several factors with an accuracy of 98.34%.

Paper - 4

Salekin, A., & Stankovic, J. (2016). Detection of Chronic Kidney Disease and Selecting Important Predictive Attributes.

In the research carried by Asif Salekin and John Stankovic, they have introduced a novel approach to detect CKD using machine learning techniques. They have evaluated their research on a dataset with 400 patient records which includes 250 CKD detected patients of early stages with 24 attributes. As the classifiers they have used; k-nearest neighbors, random forest, and neural networks to find an applicable solution. Using a wrapper method they have performed feature reduction analysis to find the attributes which detect CKD with high accuracy and a cost analysis to identify cost effective highly accurate CKD detection classifier by considering 5 attributes which include specific gravity, albumin, diabetes mellitus, hypertension and hemoglobin. The results of this study have introduced new factors to be used by classifiers for detecting CKD more accurately.

Paper -5

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

This research uses a publicly available dataset, which is downloaded from the UCI repository. This dataset includes 400 patient records with 25 attributes. Out of those 25 attributes, reduced dataset with 14 attributes related to CKD patients has been analyzed and predicted for different machine learning classification algorithms: Multiclass Decision Forest, Multiclass Decision Jungle, Multiclass Logistic Regression and Multiclass Neural Network. From this study , it is observed that the Multiclass Decision Forest algorithm provides the highest accuracy of 99.1%.