International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT228662

A Review on Machine Learning Approaches for HIV Infected Patient Chronic Kidney Disease Stage Classification

Manisha Makwana*[1], Dr. Rocky Upadhyay[2], Dr. Sheshang Degadwala[3], Dhairya Vyas[4]

*[1]Computer Engineering Department, Sigma Institute of Engineering, Vadodara, Gujarat, India
[2]Computer Engineering Department, Sigma Institute of Engineering, Vadodara, Gujarat, India
[3]Computer Engineering Department, Sigma Institute of Engineering, Vadodara, Gujarat, India
[4]Managing Director, Shree Drashti Infotech LLP, Vadodara, Gujarat, India

Article Info

Publication Issue :
Volume 8, Issue 6
November-December-2022

Page Number : 400-408

Article History
Accepted: 20 Nov 2022
Published: 05 Dec 2022

ABSTRACT

Chronic kidney disease (CKD) is sometimes called chronic kidney failure. The kidneys eliminate waste and surplus fluids from the circulation and excrete them as urine. In severe chronic renal disease, fluid, electrolytes, and waste products may build up in the body. HIV patients with additional risk factors for renal disease must have kidney function evaluated annually. HIV may damage kidney filters. The filters won't work properly. CKD has five stages, with more severe symptoms from stage 1 to stage 5. If chronic kidney disease continues to stage 4 or 5, our bodies might accumulate fluid and waste. Machine learning categories HIV-positive CKD patients based on their features. Machine learning relies on feature selection. This research uses feature selection and classification to accurately predict chronic renal illness.

Keywords: Chronic Kidney Disease, Support Vector Machine, K-nearest Neighbor, Random Forest, Decision Tree, Navier Bayes.

I. INTRODUCTION

Changes in kidney structure and function over time characterise chronic kidney disease (CKD). Kidney failure is a slow process that might extend over three months. Roughly 10% of the world's population has chronic kidney disease (CKD), making it a major public health issue. Having this condition means experiencing a gradual decline in renal function that ultimately results in kidney failure. In the first stages of CKD, the patient shows no signs of illness. As a result, the kidney disease could not show up until the organ loses around 25% of its function. Chronic kidney disease (CKD) is a degenerative and terminal disease. So, early detection and identification of CKD is crucial, as it may provide patients the freedom to get more effective therapy closer to home, which in turn may boost the momentum.

The most common methods used to diagnose CKD are a blood test measuring glomerular filtration rate and a urine test looking for albumin. Many people with CKD don't obtain a thorough diagnosis because of the high expense of healthcare and the shortage of nephrologists. As a result, a computer-aided automated diagnostic system is needed to help both patients and doctors.

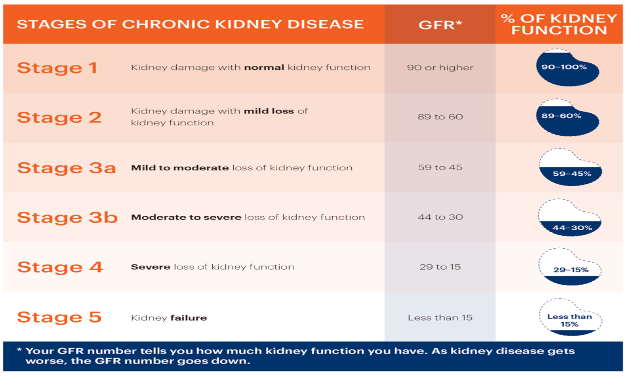

Figure 1. Stages of CKD

Many eminent academics over the last couple of decades have utilised machine learning and data mining to the development of CAD systems, which have shown promising results in accurately and efficiently detecting complicated health problems. The CKD dataset from the UCI Machine Learning Repository has been utilised to categorise CKD patients in this study. The dataset has 24 characteristics total, including several test metrics and a class label. Many previous studies have utilised this dataset to successfully classify CKD.

II. LITERATURE STUDY

No	Title	Publication-Year	Methods use	Summary
1	Contemporary issues and new challenges in chronic kidney disease amongst people living with HIV	AIDS Research and Therapy-2020	-	CKD in PLWHIV is associated with poorer clinical outcomes, including higher morbidity and mortality, and is strongly associated with CVD. Potentially nephrotoxic ART may be avoided if an individual is at high risk of CKD, or if there is established CKD.
2	A Machine Learning Methodology for Diagnosing Chronic Kidney Disease	IEEE ACCESS-2019	KNN imputation	The process of establishing the model, due to the limitations of the conditions, the available data samples are relatively small, including only 400 samples. Therefore, the generalization performance of the model might be limited. In addition, due to there are only two categories (ckd and notckd) of data samples in the dataset, the model cannot diagnose the severity of CKD.

3	Prediction of kidney disease stages using data mining algorithms	Informatics in Medicine Unlocked-2019	PNN, SVM, RBF, MLP	four data mining algorithms used on a clinical/laboratory dataset consisting of 361 chronic kidney disease patients. The results of the addressed algorithms have been compared to define the most accurate algorithm results in classifying the severity stage of CKD. This study recommends that the Probabilistic Neural Networks algorithm is the best algorithm that can be used by physicians to eliminate diagnostic and treatment errors.
4	Classification of Chronic Kidney Disease using Logistic Regression, Feedforward Neural Network and Wide & Deep Learning	International Conference on Innovation in Engineering and Technology (ICIET) 27-29 December, 2018	Logistic Regression, Feedforward Neural Networks and Wide & Deep Learning	experimental results were analyzed for two cases, model performances on actual (imbalanced) data and model performances on oversampled (balanced) data. The feedforward neural networks yielded the best results which are 0.99 F1-Score, 0.97 Precision, 0.99 Recall and 0.99 AUC score for both actual and oversampled data.
5	Role of Attributes Selection in Classification of Chronic Kidney Disease Patients	IEEE-2015	Naïve Bayes, SMO, IBK	Attributes evaluator and classification models have been applied on CKD dataset, attribute evaluator model has performed well by reducing the attributes from 25 to 6, 12 and 7 with NB, SMO and IBK Classifiers, this model result in better classification accuracy on the reduced dataset than the original dataset
6	Predicting the chronic kidney disease using Various Classifiers	ICEECCOT-2014	JRip, SMO, Naive Bayes and IBK	Attributes evaluator used for reduce the attributes of original dataset, and compare the result of original and reduced dataset, as classifier using the original dataset classification results performed

				well, but proposed classifier cannot perform well. As a result JRip produced best performance using 5 parameter.
7	Optimal Feature Selection for Chronic Kidney Disease Classification using Deep Learning Classifier	IEEE-2018	DNN, CNN, NN, BP, KNN	Author Used optimization model and learning procedure to classify CKD, that method select applicable feature of kidney data with the help of Ant Lion Optimization (ALO) technique to choose optimal features for classification. Using DNN get better classification accuracy, precision, F-measure, sensitivity.
8	Chronic Kidney Disease Prediction and Recommendation of Suitable Diet plan by using Machine Learning	ICNTE 2019	---	Diet recommendation for patient will be given according to potassium zone which is calculated using blood potassium level to slow down the progression of CKD.
9	Feature Selection And Dimensionality Reduction Methods For Chronic Disease Prediction	IJSTR-2020	----	This paper provides an analysis of various methods of selection or reduction of dimensionality used to predict and classify these conditions, Author discussed the importance of feature selection methodologies for improving the accuracy and performance of classification systems.
10	Chronic Kidney Disease (CKD) Diagnosis using Multi-Layer Perceptron Classifier	IEEE- 2020	SVM, NB, Multi-	Author use Multi-Layer Perceptron, SVM and NB Classifier to predict whether a patient suffers from the problem of CKD or not, they use 400 patient dataset, model with Multilayer Perceptron perform classification with 92.5% accuracy compare to SVM and NB.

11	Intelligent Diagnostic Prediction and Classification System for Chronic Kidney Disease	SCIENTIFIC REPORT-2019	Density based Feature Selection, Ant Colony Optimization, D-ACO	This research introduces an intelligent prediction and classification system for healthcare, namely Density based Feature Selection (DFS) with Ant Colony based Optimization (D-ACO) algorithm for chronic kidney disease (CKD), this intelligent system eliminates irrelevant or redundant features by DFS in prior to the ACO based classifier construction.
12	Two-Class Classification: Comparative Experiments for chronic kidney disease	IEEE-2019	Decision Forest, Neural Network, K-Nearest Neighbour, SVM, Rule Induction	Itshowed that by conducting experiment on small scope of the features, the Decision Forest algorithm is more reliable than the Neural Network algorithm. In terms of recall, the results showed that the accuracy and precision rate for Decision Forest algorithm is lower than the Neural Network Algorithm, while Neural Network algorithm is the best algorithm in producing a prediction of the chronic kidney disease.
13	KDSAE: Chronic kidney disease classification with multimedia data learning using deep stacked autoencoder network	Springer-2019	Stacked autoencoder (SAE) . SoftMax classifier	The model is constructed using stacked autoencoders which consists of two autoencoders arranged in cascaded fashion with one SoftMax classifier. The stacked autoencoder helps to extract the useful features from the dataset and then a SoftMax classifier is used to predict the final class the model also offers good value of precision and recall as 100% for both.
14	An Analysis on Chronic Kidney Disease Prediction	Springer-2019	NB, SMO, IB 1, Metaclassifier, VFI, Random Forest	classification with higher accuracy, estimation of GFR and finally, deduce the mortality of patients. Our prediction accuracy has been

	System: Cleaning, Preprocessing, and Effective Classification of Data			ranged from 98.30 to 99.55% with reference to the classifier.
15	Performance Based Evaluation of Various Machine Learning Classification Techniques for Chronic Kidney Disease Diagnosis	2016	DT, Linear Discriminant, Quadratic Discriminant	The authors applied various machine learning algorithms to a problem in the domain of medical diagnosis and analyzed their efficiency in predicting the results. The authors evaluated 12 classification techniques, The results indicate that decision-tree performed best with nearly the accuracy of 98.6%, sensitivity of 0.9720, precision of 1 and specificity of 1.

III. METHODOLOGY

A. Datasets

This data was obtained from the machine learning repository at the University of California, Irvine. Of the 25 characteristics in the dataset, 11 are numerical and 14 are categorical. This dataset contains 400 unique cases, all of which were obtained from the Apollo Hospitals in Karaikudi, Tamil Nadu, India.

Blood glucose, blood urea, serum creatinine, sodium, potassium, haemoglobin, packed cell volume, white blood cell count, red blood cell count, age, appetite, pedal oedema, anaemia, hypertension, diabetes mellitus, coronary artery disease, and pus cell, pus cell clumps, bacteria, and blood glucose, random.

The output variable was named as class which contained only two values, “ckd” and “notckd”.

B. Noise Removal Techniques

Abuse of shortenings, data transfer mistakes, duplicate entries, and missing values account for the majority of noise that requires normal noise reduction procedures to rectify. In this context, one important area of study is the de-duplication problem, which involves the detection and removal of duplicate entries from a database. Databases provide a problem for investigation because they may include both precise and hazy versions of a given record. At this point, the numerical dataset has been acquired and any unnecessary information has been removed.

(1) Shortenings [1, 2]: Data is not valuable in a array. Data is only valuable once information, insight or in other words knowledge is extracted from it and is used to make decisions. In data shorting data is divided into equal size raw and column.

(2) Missing esteems [1, 4]: Missing value imputation is crucial in machine learning when data is small, and all available data must be used. It affects model classification performance. As our dataset is tiny and all characteristics except output have missing values, we must impute them. In this study, we employed mean and mode missing value imputation. Mean was used to infer numerical characteristics and mode for categorical.

(3) Copy records [1,3]: Create a single statement to get rid of all the extra copies at once. You need to choose which duplicates to retain before you may delete them.

(4) Data passage botches [2]: In this part data is process in badly manner so the data passage botches algorithm will remove bugs data.

C. Attribute Selection

Attribute selection is a process of reducing the dimension of a dataset by eliminating the attributes of less importance.

(1) Best initial search traversal method [2, 6]: The best initial search traversal method prioritises the most promising nodes to visit first. This is done by letting an evaluation function choose the path taken.

(2) Wrapper subset attribute evaluator [4, 6]: Methods for encapsulating the selection of feature subsets When there are fewer variables, performance improves. Manual processes may begin with a whole collection of attributes, then iteratively eliminate the least useful ones until just the optimal attribute remains. The "wrapper" approach employs a cross-validation loop around a classifier, using the classifier to do an exhaustive search of the attribute space in order to identify the optimal subset of attributes. Forward, backward, and even directed searches are possible from any subset.

D. Machine Learning

(1) SVM [1,4,6]: For this classification task, supervised machine learning techniques such as the support vector machine (SVM) are often utilised. To perform classification and other tasks, such as outlier identification, SVM creates a hyperplane or group of hyperplanes in a high dimensional space. The hyperplane with the greatest distance to the closest training-data point of any class provides the best classification.

(2) NB [9,10,15]: The conditional probabilities are the foundation of a Naive Bayes classification method. Naive Bayes employs the Bayes formula, which determines the likelihood of an event by summing the occurrences of each value and each value combination in the past. The INB model is simple to construct and effective for massive data.

(3) k-NN [3,5]: An example of a supervised machine learning method is the K-nearest neighbour (KNN) algorithm. It's applicable to both regression and classification issues in predictive modelling. The voting system of an object's closest neighbours is used to determine its classification. The parameter "k" in k-NN is left up to the researcher. It's a lazy learning algorithm that doesn't rely on any parameters.

(4) Decision Tree [2,7]: With a decision tree, data is organised in a tree structure, and then branches out to be categorised. All those forks signify different possibilities. The choices and their potential effects and benefits are shown in a tree-like form. Furthermore, it may be integrated with several other algorithms.

(5) RF [11,12,14]: The Random Forest method constructs an ensemble decision tree and is used in supervised machine learning. It may be put to work in both classification and regression analysis. Simple random forest constructs many different decision trees and then merges them into a single, correct conclusion. In RF, the greater the number of trees used, the more precise the final outcome will be.

IV.Comparative Analysis

TABLE I

COMPARATIVE ANALYSIS

Method	Advantage	Disadvantages
SVM [1,4,6]	-SVM is less complex. -Produce very accurate classifiers. -Less over fitting, -Robust to noise.	-SVM is binary classifier, to do multiclass classification -- pair-wise classifications can be used - Computationally expensive, thus runs slow

KNN [3,5]	-Robust to noisy training data -Effective if the training data is large	-Distance based learning is not clear which type of distance to use and which attribute to use to produce the best result. -Computation cost is quite high
Decision Tree [2,7]	-It reduces over-fitting and is therefore more accurate. -Easy to Implement -Works with all types of data. -Multi classification Support	-It may not work if the dependent variables considered in the model are linearly related. Therefore, one has to remove correlated variable by some other technique
RF [11,12,14]	-Efficient on large dataset -Flexibly include missing data from previous node of the tree	-High computational cost -Hard to interpret -overfitting
NB [9,10,15]	-Easy to implement -Requires a small amount of training data to estimate the test data	-All the attributes are mutually independent. -Does not support categorical variable.

V. CONCLUSION

In order to diagnose or categorise chronic kidney disease, it is necessary to understand what makes it distinct. Nevertheless, it is often challenging to identify distinct patterns in qualities, particularly when dealing with a huge volume of data. Data gathered from the natural world is often non-linear, making it impossible to apply conventional techniques for pattern recognition or mathematical modelling. Wrapper subset attribute evaluator approach may provide batter performance, as discussed in the study, which evaluates several attributes selection among best search methods and finds that none of them function. As for learning strategies, machine learning-based techniques are on the rise.

Consequently, in the future, if machine learning is the cutting-edge technology that ultimately succeeds in resolving this issue. Attributes Classification of chronic renal disease is improved by using selection and data balancing approaches.

VI.REFERENCES

[1] J. E. Heron, C. I. Bagnis, and D. M. Gracey, "Contemporary issues and new challenges in chronic kidney disease amongst people living with HIV," AIDS Res. Ther., vol. 17, no. 1, pp. 1–13, 2020, doi: 10.1186/s12981-020-00266-3.

[2] J. Qin, L. Chen, Y. Liu, C. Liu, C. Feng, and B. Chen, "A machine learning methodology for diagnosing chronic kidney disease," IEEE Access, vol. 8, pp. 20991–21002, 2020, doi: 10.1109/ACCESS.2019.2963053.

[3] E. H. A. Rady and A. S. Anwar, "Prediction of kidney disease stages using data mining algorithms," Informatics Med. Unlocked, vol. 15, no. March, p. 100178, 2019, doi: 10.1016/j.imu.2019.100178.

[4] A. Al Imran, M. N. Amin, and F. T. Johora, "Classification of Chronic Kidney Disease using Logistic Regression, Feedforward Neural Network and Wide Deep Learning," 2018 Int. Conf. Innov. Eng. Technol. ICIET 2018, pp. 1–6, 2019, doi: 10.1109/CIET.2018.8660844.

[5] N. Chetty, K. S. Vaisla, and S. D. Sudarsan, “Role of attributes selection in classification of chronic kidney disease patients,” pp. 1–6, 2016, doi: 10.1109/cccs.2015.7374193.

[6] P. Arulanthu and E. Perumal, “Predicting the chronic kidney disease using Various Classifiers,” 4th Int. Conf. Electr. Electron. Commun. Comput. Technol. Optim. Tech. ICEECCOT 2019, pp. 70–75, 2019, doi: 10.1109/ICEECCOT46775.2019.9114653.

[7] K. Shankar, P. Manickam, G. Devika, and M. Ilayaraja, “Optimal Feature Selection for Chronic Kidney Disease Classification using Deep Learning Classifier,” 2018 IEEE Int. Conf. Comput. Intell. Comput. Res. ICCIC 2018, pp. 1–5, 2018, doi: 10.1109/ICCIC.2018.8782340.

[8] A. Maurya, R. Wable, R. Shinde, S. John, R. Jadhav, and R. Dakshayani, “Chronic Kidney Disease Prediction and Recommendation of Suitable Diet Plan by using Machine Learning,” 2019 Int. Conf. Nascent Technol. Eng. ICNTE 2019 - Proc., no. Icnte, pp. 1–4, 2019, doi: 10.1109/ICNTE44896.2019.8946029.

[9] R. Yadav and S. C. Jat, “Feature selection and dimensionality reduction methods for chronic disease prediction,” Int. J. Sci. Technol. Res., vol. 9, no. 4, pp. 2912–2918, 2020.

[10] D. Hu, L. J. Frp, and V. J. Frp, “&kurqlf .lgqh\ ‘lvhdvh &.’ ’ldjqrvlv xvlqj 0xowl /d\hu 3hufhswurq &odvvlilhu,” pp. 346–350, 2020.

[11] M. Elhoseny, K. Shankar, and J. Uthayakumar, “Intelligent Diagnostic Prediction and Classification System for Chronic Kidney Disease,” Sci. Rep., vol. 9, no. 1, pp. 1–14, 2019, doi: 10.1038/s41598-019-46074-2.

[12] A. A. Johari, M. H. Abd Wahab, and A. Mustapha, “Two-Class Classification: Comparative Experiments for Chronic Kidney Disease,” 2019 4th Int. Conf. Inf. Syst. Comput. Networks, ISCON 2019, pp. 789–792, 2019, doi: 10.1109/ISCON47742.2019.9036306.

[13] A. Khamparia, G. Saini, B. Pandey, S. Tiwari, D. Gupta, and A. Khanna, “KDSAE: Chronic kidney disease classification with multimedia data learning using deep stacked autoencoder network,” Multimed. Tools Appl., 2019, doi: 10.1007/s11042-019-07839-z.

[14] S. Singhal, S. Maheshwari, and M. Meena, Recent Findings in Intelligent Computing Techniques, vol. 707. Springer Singapore, 2019.

[15] S. Sharma, V. Sharma, and A. Sharma, “Performance Based Evaluation of Various Machine Learning Classification Techniques for Chronic Kidney Disease Diagnosis,” 2016, [Online]. Available: http://arxiv.org/abs/1606.09581.

Cite this article as :

Manisha Makwana, Dr. Rocky Upadhyay, Dr. Sheshang Degadwala, Dhairya Vyas, "A Review on Machine Learning Approaches for HIV Infected Patient Chronic Kidney Disease Stage Classification", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 6, pp. 400-408, November-December 2022. Available at doi :
https://doi.org/10.32628/CSEIT228662
Journal URL : https://ijsrcseit.com/CSEIT228662