

CONCEPTUAL REVIEW ON CHRONIC KIDNEY DISEASE PREDICTION USING FEATURE SELECTION METHODS

Rakesh Kumar Pandey*
Dr. Anoop Kumar Tiwari**

ABSTRACT

As the human race is progressing, the diseases which are impacting the humans are also progressing and its deadliest example is COVID-19. With the advancement of technology, Healthcare sector is also deeply influenced by the growth of the Information technology. The disease predictions using data mining as the best example of this case. In medical diagnosis, it is very important to identify most significant risk factors related to disease. Relevant feature identification helps in the removal of unnecessary, redundant attributes from the disease dataset which, in turn, gives quick and better results. Chronic Disease Prediction plays a pivotal role in healthcare informatics. It is crucial to diagnose the disease at an early stage. This paper reviews the various of feature selection-based approaches which can be used for Chronic Kidney Disease Prediction.

Keywords: *Chronic Kidney Disease, Feature Selection, Data Mining, Healthcare, Disease Prediction.*

Introduction

Chronic kidney disease, additionally called chronic kidney disappointment, includes a steady loss of kidney work. Your kidneys channel squanders and overabundance liquids from your blood, which are then eliminated in your pee. Progressed chronic kidney disease can prompt risky degrees of liquid, electrolytes and squanders to develop in your body. [1]

In the beginning phases of chronic kidney disease, you may have not many signs or manifestations. You probably won't understand that you have kidney disease until the condition is progressed. [1]

Therapy for chronic kidney disease centers around easing back the movement of kidney harm, generally by controlling the reason. Yet, in any event, controlling the reason probably won't keep kidney harm from advancing. Chronic kidney disease can advance to end-stage kidney disappointment, which is lethal without counterfeit sifting (dialysis) or a kidney relocate. [2]

Signs and side effects of chronic kidney disease create after some time on the off chance that kidney harm advances gradually. Loss of kidney capacity can cause a development of liquid or body waste or electrolyte issues. Contingent upon how serious it is, loss of kidney capacity can cause: [3]

- Sickness
- Heaving
- Loss of hunger
- Weariness and shortcoming
- Rest issues
- Peeing pretty much

* Research Scholar, Department of Computer Science and Engineering, Dr. K.N. Modi University, Rajasthan, India.

** Assistant Professor, Department of Computer Science and Engineering, Dr. K.N. Modi University, Rajasthan, India.

- Diminished mental sharpness
- Muscle cramps
- Expanding of feet and lower legs
- Dry, bothersome skin
- (Hypertension) that is hard to control
- Windedness, if liquid develops in the lungs
- Chest torment, if liquid develops around the covering of the heart

Signs and side effects of kidney disease are regularly vague. This implies they can likewise be brought about by different sicknesses. Since your kidneys can compensate for lost capacity, you probably won't foster signs and side effects until irreversible harm has happened. [3]

Causes of Chronic kidney disease

Chronic kidney disease happens when a disease or condition hinders kidney work, causing kidney harm to deteriorate more than a while or a long time. [3]

Diseases and conditions that cause chronic kidney disease include:

- Type 1 or type 2 diabetes
- Hypertension
- Glomerulonephritis (gloe-mer-u-low-nuh-FRY-tis), an irritation of the kidney's sifting units (glomeruli)
- Interstitial nephritis (in-tur-STISH-ul nuh-FRY-tis), an irritation of the kidney's tubules and encompassing constructions
- Polycystic kidney disease or other acquired kidney diseases
- Delayed block of the urinary parcel, from conditions, for example, expanded prostate, kidney stones and a few malignant growths
- Vesicoureteral (ves-ih-koe-yoo-REE-tur-ul) reflux, a condition that makes pee back up into your kidneys
- Repetitive kidney contamination, additionally called pyelonephritis (pie-uh-low-nuh-FRY-tis) [3]

Literature Review

P. A. Moreno-Sanchez, 2020 [4] Chronic Kidney Disease (CKD) is an overall chronic disease that, assuming it is perceived late, drives a greater part of patients to experience untimely mortality and personal satisfaction decrease because of an ever-evolving loss of kidney work. Information mining classifiers would add to an early diagnosis and consequently forestalling kidney extreme harm since unpretentious examples in CKD pointers can be found. By utilizing Cross Industry Standard Process of Data Mining (CRISP-DM) approach alongside features significance procedures, this work fosters a classifier model that would support medical care experts in early diagnosis of CKD patients. Through an information workflow pipeline, a computerized information change, demonstrating and assessment is applied to the CKD dataset removed from the University of California Irvine-Machine Learning (UCI-ML) storehouse. The pipeline created is utilized to do a thorough inquiry of the best information mining classifier and the various boundaries of the information arrangement's sub-stages like information missing and feature selection. Therefore, Ada Boost is chosen as the best classifier with a 100 percent as far as exactness, accuracy, sensitivity, explicitness, and f1-score; outflanking the characterization results got by the connected works even with inconspicuous information from a test set. As to's interpretability, the use of feature selection lessens from 24 to 12 the gathering of features to be utilized in the classifier model created, accomplishing more logical model's results. Besides, an investigation of the significance of features chose is done to investigate the pertinence of each chosen feature.

O. A. Jongbo, et.al 2020 [5] Chronic Kidney Disease (CKD) is a general medical problem which is viewed as a huge danger to human existence because of unusual working of kidney over a time of months or a long time which, whenever left untreated, may harm imperative organs in the body prompting an expanded rate in cardiovascular mortality which might bring about abrupt passing if not early distinguished. Information mining methods are utilized in a few clinical findings for making savvy diagnostics decisions that can be applied in disease expectation. The exhibitions of these methods are extremely encouraging in the administration of various afflictions to diminish the large quantities of individuals that bite the dust yearly because of wrong diagnosis of various disease conditions. This

review assesses the exhibition of a sacking troupe procedure on CKD dataset with a compelling feature selection strategy to yield a solid and exact prescient model able to do accurately grouping diseased from non-diseased patients. The review was examined on a genuine patient dataset got from the UCI machine learning storehouse comprising of 400 occurrences with 24 restrictive characteristics and a decisional class. Radom timberland algorithm was utilized as an action to choose the best subset of features for the prescient models. Credulous Bayes, k-Nearest Neighbor, and Decision Tree algorithms fill in as the base classifiers whose presentation were amassed utilizing the sacking outfit way to deal with further develop base students' exhibitions. Results got from the review showed the impact of feature selection and group method in working on the exactness of information mining arrangement algorithms. The model's ideal outcome is accomplished utilizing 7 best-chosen features on the outfit classifier with 100 percent exactness of CKD diagnosis contrasted with 98.3% precision without feature selection. Henceforth, making the model reasonable for proficient diagnosis of CKD.

M. S. Wibawa et.al 2017 [6] Chronic kidney disease (CKD) is a disease brought about by degeneration capacity of the kidneys. CKD is top ten driving reasons for death on the planet. There are two driving reasons for CKD, for example diabetes and hypertension. At the point when the indications become more serious, the disease must be treated with dialysis and kidney transplantation. This disease can be dealt with if the analyze is led fittingly and quickly. Notwithstanding, the signs and side effects are regularly not explicit. Hence, diagnosis from clinical work force is may emotional and fluctuate. This review created machine learning strategy utilizing gathering learning and feature selection to work on the nature of CKD diagnosis. The CKD dataset was taken from UCI machine learning archive, it contain 400 cases. CKD dataset have 24 ascribes including signs, manifestations and risk factors that may show up because of CKD. In this review, features were chosen utilizing a Correlation-based Feature Selection (CFS) and AdaBoost was utilized for group learning to work on the location of CKD. K-Nearest Neighbor algorithm (kNN), Naive Bayes and Support Vector Machine (SVM) was utilized as base classifier. In general, the best outcome was accomplished by blend of kNN classifier with CFS and AdaBoost, with 0.981 precision rate, 0.980 review rate and 0.980 f-measure rate. Most elevated accuracy rate was accomplished by the blend of Naive Bayes classifier with CFS and AdaBoost, with 0.981 accuracy rate.

Peng, X. Zhu et.al 2019 [7] Traditional Chinese Medicine (TCM) has been broadly applied in Chronic Kidney Disease (CKD), syndrome order is a vital stage in the diagnosis of TCM. In view of clinical diagnosis encounters and the similarity hypothesis of TCM, creators propose a strategy that is grounded on relationship hypothesis of TCM for syndrome diagnosis (ATTSCM). The ATTSCM technique uses the feature selection algorithm to choose the critical side effects, and the grouping algorithm join multi-mark learning algorithm to recognize the syndromes. The proposed strategy is assessed by the chronic kidney disease dataset which is gathered by University of Chinese Medicine. Three assessment proportions of multi-name grouping are taken to assess the presentation of syndrome characterization, including order precision, miniature F1 measure and hamming misfortune. Trial results show that the proposed strategy performs well in the application for TCM syndrome separation of CKD.

N. Tazin, et.al 2016 [8] The enormous measure of information gathered by medical services area can be powerful for examination, diagnosis and decision making assuming it is mined appropriately. Secret data removed from the voluminous information can give assistance and solution for handle basic medical services circumstances. Chronic kidney disease is a deadly sickness of kidney which can be forestalled with early right expectations and appropriate precautionary measures. Information mining of the data gathered from recently analyzed patients opened up another period of clinical progression. In any case, explicit methods should be executed to achieve better outcome. In this original copy the ability of the grouping of Support Vector Machine, Decision tree, Naïve Bayes and K-Nearest Neighbor algorithm, in dissecting the Chronic Kidney Disease dataset gathered from UCI store, was examined to foresee the presence of kidney disease. Informational collection has been dissected as far as exactness, Root Mean Squared Error, Mean Absolute Error and Receiver Operating Characteristic bend. In the current review, Decision tree shows promising outcomes when executed through WEKA information mining instrument. Ranking algorithm furnishes fundamental enhancements in arrangements with appropriate number credits. 15 ends up being the enchanted number for choosing ascribes for the given dataset coming about most noteworthy percent of progress in precision.

J. R. Lambert et.al 2020 [9] Diagnosis of chronic kidney disease (CKD) has for the most part huge in the clinical field of information mining. The goal of this paper is to foresee chronic kidney disease utilizing just ostensible characteristics which results contrasted with mathematical and both ostensible and mathematical qualities. The characterization and forecast of this paper presents connection based feature selection (CFS) strategy applied to separate significant characteristics and grouping them into

CKD and not CKD. The CFS approach is applied on the information sorts of ostensible, mathematical and both ostensible and mathematical characteristics in feature selection. The aftereffect of the CFS approach contrasted with three ranker methodologies like Information Gain, Gain Ratio, and ReliefF approach for feature selection. The exactness of relationship based feature selection and consecutive least streamlining (CFS-SMO) approach accomplished 98.5% for ostensible, 95.25% for mathematical and 98.5% for ostensible and mathematical. These test results pronounce that the relationship based feature selection (CFS) effectively separated features from the benchmark endorsement of own and unique chronic kidney disease (CKD) dataset and SMO ordered the situation with kidney disease. In this way CFS-SMO considered as a hopeful apparatus to diagnosis kidney disease precisely which supports clinical specialists to make decision accurately.

A. Salekin and J. Stankovic, 2016 [10] Chronic kidney disease (CKD) is a significant general wellbeing worry with rising predominance. In this review creators think about 24 prescient boundaries and make a machine learning classifier to distinguish CKD. Creators assess their methodology on a dataset of 400 people, where 250 of them have CKD. Utilizing their methodology creators accomplish a discovery precision of 0.993 as per the F1-measure with 0.1084 root mean square mistake. This is a 56% decrease of mean square blunder contrasted with the cutting edge (i.e., the CKD-EPI condition: a glomerular filtration rate assessor). Creators additionally perform feature selection to decide the most important properties for recognizing CKD and rank them as per their consistency. Creators distinguish new prescient qualities which have not been utilized by any past GFR assessor conditions. At long last, creators play out an expense exactness tradeoff investigation to recognize another CKD discovery approach with high precision and minimal expense.

M. Manonmani and S. Balakrishnan, 2020 [11] Feature Selection is a significant antecedent to expectation and arrangement of clinical information. Clinical information mining is advancing at a quicker rate and the current machine learning algorithms should be guzzled with smart expectation and arrangement frameworks to handle the gigantic clinical information. Since the gigantic volume of information put away in the clinical data set might be inclined to 'revile of dimensionality', it becomes important to embrace strategies that handle the issues of high layered information and work on the soundness of the chose strategy. One such arrangement as visualized in ongoing writing is the troupe feature selection procedure that joins the aftereffects of channel and covering strategies to choose the most biased features for diagnosis of chronic diseases. In this paper, a troupe feature selection method is applied to Chronic Kidney Disease (CKD) dataset. Thickness based Feature Selection (DFS) strategy is utilized as a channel way to deal with rank the features of CKD. The aftereffects of DFS strategy is given to a covering based streamlining procedure named Improved Teacher Learner Based Optimization (ITLBO) algorithm to observe the ideal feature set that contains the main features for expectation of CKD with high precision. The consequences of group feature selection technique are assessed utilizing SVM, Gradient Boosting, and CNN characterization algorithms. Test results uncover that the proposed strategy can accomplish high grouping exactness of 93% for SVM, 97% for Gradient Boosting and 97.75% for CNN individually for the determined ideal feature subset.

Conclusion

In developed countries, Chronic Kidney Disease (CKD) is developing as a significant general medical issue. Other than causing untimely mortality and bleakness and bringing down the existence quality, this is a disease which forces a tremendous monetary weight not just on parental figures, and additionally on the federal medical insurance framework and country. The better system for tackling CKD is early and precise diagnosis of the disease and move ought to be made to slow its encouraging towards end stage renal disappointment (ESRD). The utilization of feature selection approaches for CKD diagnosis in information mining picks the factors that much of the time affect the thought about event. Likewise, the computation time and precision of the chose characterization approach additionally lessens.

References

1. M. Lenart N. Mascarenhas R. Xiong and A. Flower "Identifying risk of progression for patients with Chronic Kidney Disease using clustering models" Systems and Information Engineering Design Symposium (SIEDS) 2016 IEEE pp. 221-226 2016.
2. S. Dhamodharan "Liver Disease Prediction Using Bayesian Classification" 4 th National Conference of Advance Computing applications and Technologies pp. 1-3 2014.

3. Oguntimilehin A. O. Adetunmbi and K. A. Olatunji "A Machine Learning Based Clinical Decision Support System for Diagnosis and Treatment of Typhoid Fever" *International Journal of Advanced Research in Computer Science and Software Engineering* vol. 4 no. 6 pp. 1-9 2014.
4. P. A. Moreno-Sanchez, "Features Importance to Improve Interpretability of Chronic Kidney Disease Early Diagnosis," *2020 IEEE International Conference on Big Data (Big Data)*, 2020, pp. 3786-3792.
5. O. A. Jongbo A. O. Adetunmbi R. B. Ogunrinde and B. Badeji-Ajisafe "Development of an Ensemble Approach to Chronic Kidney Disease Diagnosis" *Scientific African* 2020 .
6. M. S. Wibawa, I. M. D. Maysanjaya and I. M. A. W. Putra, "Boosted classifier and features selection for enhancing chronic kidney disease diagnose," *2017 5th International Conference on Cyber and IT Service Management (CITSM)*, 2017, pp. 1-6.
7. Peng, X. Zhu, H. Liao and P. Zhang, "Application of Multi-label Learning Model for Chronic Kidney Disease Syndrome Classification," *2019 IEEE 5th International Conference on Computer and Communications (ICCC)*, 2019, pp. 1729-1733.
8. N. Tazin, S. A. Sabab and M. T. Chowdhury, "Diagnosis of Chronic Kidney Disease using effective classification and feature selection technique," *2016 International Conference on Medical Engineering, Health Informatics and Technology (MediTec)*, 2016, pp. 1-6.
9. J. R. Lambert, P. Arulanthu and E. Perumal, "Identification of Nominal Attributes for Intelligent Classification of Chronic Kidney Disease using Optimization Algorithm," *2020 International Conference on Communication and Signal Processing (ICCSP)*, 2020, pp. 0119-0125.
10. Salekin and J. Stankovic, "Detection of Chronic Kidney Disease and Selecting Important Predictive Attributes," *2016 IEEE International Conference on Healthcare Informatics (ICHI)*, 2016, pp. 262-270.
11. M. Manonmani and S. Balakrishnan, "An Ensemble Feature Selection Method for Prediction of CKD," *2020 International Conference on Computer Communication and Informatics (ICCCI)*, 2020, pp. 1-6.

