

Machine Learning Methodology for Prediction of Chronic Kidney Disease

Miss. Divya Ravinder Pogaku¹, Prof. Sneha Bohra²

^{1,2}Department Of Computer Science & Engineering, G H Raison University, Amravati

Abstract

Chronic Kidney Disease (CKD) is a major health problem affecting millions of people worldwide. Early and accurate diagnosis of CKD is essential for successful management and treatment of the disease. In this paper, we propose a machine learning-based approach for diagnosing CKD using different classification algorithms. Our approach utilizes a combination of demographic data, medical history, and laboratory test results to predict CKD. We tested our approach using several machine learning algorithms, including decision trees, random forests, and support vector machines (SVM), and compared our results with traditional diagnostic methods. Our results show that SVM achieved the highest accuracy in diagnosing CKD, followed by decision trees and random forests. Our approach outperformed traditional diagnostic methods in terms of accuracy and reliability, demonstrating the potential of machine learning in improving CKD diagnosis. Our approach can be used to develop a computer-aided diagnosis system to assist clinicians in the early and accurate diagnosis of CKD, leading to better patient outcomes.

Keywords: CKD, SVM, Prediction

I. INTRODUCTION

Chronic Kidney Disease (CKD) is a common and serious health problem affecting millions of people worldwide. CKD is characterized by the progressive loss of kidney function over time, which can eventually lead to kidney failure and the need for dialysis or kidney transplantation. Early and accurate diagnosis of CKD is crucial for successful management and treatment of the disease, as well as for the prevention of complications such as cardiovascular disease and premature death.

Traditionally, CKD diagnosis has relied on a combination of clinical symptoms, physical examination, and laboratory tests such as serum creatinine, blood urea nitrogen (BUN), and estimated glomerular filtration rate (eGFR). However, these methods have limitations in terms of accuracy and reliability, and there is a growing interest in developing alternative diagnostic approaches using advanced technologies such as machine learning.

Machine learning is a kind of artificial intelligence which enables to learn from data and make predictions without even explicitly programmed. Machine learning algorithms have been widely used in healthcare to develop predictive models for various diseases, including cancer, diabetes, and heart

disease. In recent years, machine learning has also been applied to CKD diagnosis, with promising results.

In this paper, we propose a machine learning-based approach for diagnosing CKD using different classification algorithms. Our approach utilizes a combination of demographic data, medical history, and laboratory test results to predict CKD. We tested our approach using several machine learning algorithms, including decision trees, random forests, and support vector machines (SVM), and compared our results with traditional diagnostic methods.

Chronic kidney disease, also called chronic kidney failure, involves a gradual loss of kidney function. Your kidneys filter wastes and excess fluids from your blood, which are then removed in your urine. Advanced chronic kidney disease can cause dangerous levels of fluid, electrolytes and wastes to build up in your body. Chronic kidney disease (CKD) is a global health problem with high morbidity and mortality rate, and it induces other diseases. Since there are no obvious symptoms in the early stages of CKD, patients often fail to notice CKD. Early detection of CKD enables patients to receive timely treatment to ameliorate the progression of this disease. Machine learning models can effectively aid clinicians achieve this goal due to their fast and accurate recognition performance. In this study, a machine learning methodology for diagnosing CKD is proposed. The CKD data set was obtained from the University of California Irvine (UCI) machine learning repository, which has a large number of missing values.

KNN imputation was used to fill in the missing values, which selects several complete samples with the most similar measurements to process the missing data for each incomplete sample. Missing values are usually seen in real-life medical situations because patients may miss some measurements for various reasons. After effectively filling out the incomplete data set, 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 accuracy. By analyzing the misjudgments generated by the established models, an integrated model that combines logistic regression and random forest by using perceptron would perform with best accuracy. Hence, this methodology could be applicable to more complicated clinical data for disease diagnosis.

The main objective of this paper is to evaluate the performance of machine learning algorithms in CKD diagnosis and compare them with traditional diagnostic methods. We aim to demonstrate the potential of machine learning in improving CKD diagnosis, which can lead to better patient outcomes and reduced healthcare costs. Furthermore, we propose that our approach can be used to develop a computer-aided diagnosis system to assist clinicians in the early and accurate diagnosis of CKD.

II. LITRETURE REVIEW

FuRESs and FOAM, two internal fuzzy classifiers, were examined to see if they might be used to diagnosing individuals with chronic kidney disease (CKD). A linear classifier, was utilized for comparison. The Machine Learning Repository provided the CKD data utilized in this study. To test the resilience of the two fuzzy techniques, composite dataset was made by addition of various amounts of proportional noise. After adding 11 stages of proportional noise sequentially on every numerical the

training and prediction sets, and the attribute were initially mixed in pairs. In order to compare the categorization rates for these 121 couples, a grid was created containing 121 sets of simulated data. Second, using simulated the two fuzzy classifiers' performances were evaluated on dataset with 11 stages of noise distributed at random to every numeric attribute. FuRES and FOAM have 200 bootstrap Latin partitions. The average prediction rates of 98.1 0.5% and 97.2 1.2%, respectively. With the same evaluation, the PLS-DA may provide 94.3 0.8%. FuRES, FOAM, and PLS-DA classifiers models were also assessed using confluent datasets made up to the original dataset and modified dataset. The 200-bootstrapped average predicted rate for FuRES and FOAM findings show that both FuRES and FOAM are effective in identifying CKD patients. These two fuzzy categories are crucial tools for the accurate identification of individuals with CKD. [1].

We conducted surveying a cross-section of a sample for Chinese people individuals that was nationally representative to determine if albuminuria was present. Participants had their blood pressure checked, samples of blood and urine were obtained, and they completed a questionnaire on their lifestyle and medical history. GFR was calculated using measurements of serum creatinine. Compared to other locations, albuminuria was independently connected with economic growth in rural areas. Age, sex, diabetes, high blood pressure, hyperuricemia, location where the patient lives and also patient's financial status were additional traits that are linked to kidney impairment [2].

It is possible that the management of chronic illnesses may be significantly improved by predictive models created utilizing temporal data from electronic health records (EHRs). These data, however, come with a host of technical difficulties, such as erratic data collection and variable patient history duration. In this article, we outline and compare three alternative methods for applying machine learning to create prediction model using the patient's temporal EHR data. The first technique, this integrates the scores Together with the patient's medical history, one or more predictors and is a common non-temporal strategy. Another two approaches take advantage of the temporal changes the data. The two temporal methods handle missing data differently and represent temporal information differently. We created and evaluated a model to forecast changes in the most used kidney function indicator, the estimated glomerular filtration rate (eGFR), using information from the electronic health record (EHR) at Mount Sinai Medical Centre. Our findings suggest that temporal data might help predict kidney function decline by being included in the medical record of a person. They also show how crucial it is to include this knowledge in a certain way. Our findings, in particular, show that employed multi-task learning for the suitable technique to EHR data's temporal dynamics are captured robustly. Different predictors' relative values change with time [3].

The mortality rate can only be decreased by early discovery and efficient treatment. Machine learning algorithms are playing an increasing role in medical diagnostics because of them ability to categories data with high rates of accuracy. In order for classification algorithms to function correctly, appropriate feature selection techniques must be utilized to make datasets dimensions smaller. The category of Support Vector Machine technique was employing in this work to identify chronic renal disease. In order to shrink the chronic renal disease data file and identify the trouble, 2 important sorts of features strategies—methods using wrappers and filters—were applied. In a wrapper approach, both the classifier subset evaluator and the wrappers subset evaluate with the most effective the initial search engine were used.

Stepwise The Greedy Search Engine and the Correlation-Based Evaluation of Features Subset Evaluator were used in the filter approach. consequently, the most effective the SVM classifier, First Search feature selection and the evaluator for filtered subsets methods achieved greater accuracy(98.5%), for the predicting a chronic kidney condition [4].

One of the most prevalent co-occurring conditions in people experiencing chronic kidney failure is anemia which is the end-stage renal disease. Agents that stimulate erythropoiesis (ESA), in particular, have emerged as the preferred therapy for that anemia. To uncover an appropriate medication for a patient, even the identical patients with varying Symptoms of anemia is quite difficult. This study builds upon previous research addressing the same issue, utilizing various techniques, including machine learning. However, this study specifically focuses on the population afflicted with chronic kidney disease (CKD) to examine their response to ESA/Iron therapy and develop a precise methodology for forecasting CKD. Furthermore, the three study-related nations—Spain, Italy, and Portugal—are represented in the ML method by both human and data inputs. A measure of the hemoglobin (Hb) prediction's mean absolute errors (MAE) for this model were close to or less than 0.6 g/dl, outperforming prior techniques [5].

Patients already suffering from chronic diabetes, major vascular disease and detected with early stage of chronic renal disease are at a risk to experience cardiovascular diseases. It is unknown how early chronic kidney disease affects borderline diabetes early-onset type 2 diabetic mellitus macro vascular outcomes. The inclusion of insulin had no difference in cardiovascular outcomes from usual treatment in the ORIGIN research (Outcome Reducing with an Initially Glargine Intervention). In this ORIGIN post hoc study, we assessed the cardiovascular results. Across participants with and without mild or methods for mild chronic renal illness. The evaluation of two co-primary composites cardiovascular outcomes. The first was the composite outcome of nonfatal MI death from cardiovascular cause; and the second occurred when any of these occurrences were added to a revascularization surgery or a hospital stay for heart failure. Micro vascular results, incidence diabetes, hypoglycemia, weight, and malignancies were among the pre-specified secondary outcomes. Patients with mild to severe renal disease experience a higher occurrence of combined primary ailments, such as cardiovascular mortality, nonfatal myocardial infarction, or nonfatal stroke, compared to patients without CKD [6].

Machine learning approach is used to detect Lung cancer by S Mukherjee, SU Bohra - The examination of the lungs has captivated medical experts throughout history, and even in modern times, it remains an intriguing field of investigation. In order to address this challenge, the development of a predictive system holds promise in reducing the threat to human life by enabling early detection of malignant growths. Numerous frameworks have been proposed, with many still in the experimental phase. One approach involves utilizing image data to detect cancerous cells, employing a neural network model to improve performance. A lung cancer prediction framework has been developed based on AI and deep neural networks, relying on supervised learning to achieve enhanced accuracy [7].

The diagnosis of medical conditions poses a significant concern for healthcare professionals, as critical decisions and treatment plans depend on accurate assessments. To address this challenge, a proposed framework aims to detect lung malignancy at an early stage through a two-phase process. The framework

involves multiple steps, including image extraction, pre-processing, binarization, thresholding, division, feature extraction, and neural network identification. This model unfolds a Lung Cancer detection system based on machine learning and neural networks. Early detection of this malignancy plays a crucial role in saving patients' lives. In recent times, there has been significant growth in data mining and machine learning techniques for predicting various chronic diseases. In this particular model, CT scan images are utilized as inputs to predict the probability of the disease and its stages. A quick intelligent clinical decisions are framed to help medical practitioners to find out the disease in primitive stages and thereby makes the treatment cheaper. This machine learning study is used in this model to detect lung cancer and its stage detection which is propounded by S Mukherjee, S Bohra in 2020 [8].

Data mining techniques are efficient in detection of disease supporting medical predictions while reducing the burden on medical practitioners. In this model the primitive discovery of diabetes is done by using data mining. There are number of diseases linked to diabetes mellitus which are related to kidney, eye, and heart and with forth highest fatality rate in the world. The current study focuses on such issues by using data mining techniques accurate medical prediction is achieved. Chronic disease diabetes mellitus affects different body parts differently. Beforehand prediction of the disease will helps to save lives. Historically, the diagnosis of diabetes has relied on a battery-operated physical examination, which, in fact, lacks accuracy. To address this limitation, the current model utilizes Data Mining techniques, including Naive Bayes, K-Nearest Neighbor, Support Vector Machine, and Decision Tree, to predict the occurrence of diabetes mellitus. This detection model is developed by PG Palanimani, V Suresh Kumar, Sanket B Kasturiwala, Shafaque Ahmareen, Sneha Bohra. [9].

III. PROPOSED WORK

The proposed system for machine learning methodology for diagnosing chronic kidney disease (CKD) is an intelligent diagnostic tool that utilizes machine learning algorithms to accurately predict the presence of CKD based on patient data. The system consists of several components, including data collection, feature selection, model development, and model evaluation.

Data collection: The first step in developing the proposed system is to collect patient data, including demographic information, medical history, and laboratory test results such as serum creatinine, blood urea nitrogen (BUN), and estimated glomerular filtration rate (eGFR). The data can be collected from electronic health records, laboratory information systems, or other data sources.

Feature selection: Once the data is collected, feature selection techniques can be used to identify the most relevant features for CKD diagnosis. This can be done using statistical methods or machine learning algorithms such as recursive feature elimination (RFE) or principal component analysis (PCA). After feature selection, various machine learning algorithms can be used to develop predictive models for CKD diagnosis, including decision trees, random forests, SVM, and artificial neural networks. The models can be trained and optimized using the collected patient data and validated using cross-validation techniques to ensure accuracy and generalizability.

Model evaluation: Finally, the developed models can be evaluated based on their performance metrics such as accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-

ROC). The models can also be compared with traditional diagnostic methods to assess their effectiveness in improving CKD diagnosis.

The proposed system can be deployed as a standalone application or integrated into existing clinical decision support systems to assist healthcare professionals in making accurate and timely CKD diagnoses. The system has the potential to improve patient outcomes and reduce healthcare costs by providing more accurate and efficient CKD diagnoses.

METHODOLOGY

Methodology for Classification of Emotions and Evaluation of Customer Satisfaction from Speech in Real World Acoustic Environments using Google Colab and Python uses machine learning-based algorithms which can reduce human error and improve the efficiency of the diagnostic process technology involves the following steps:

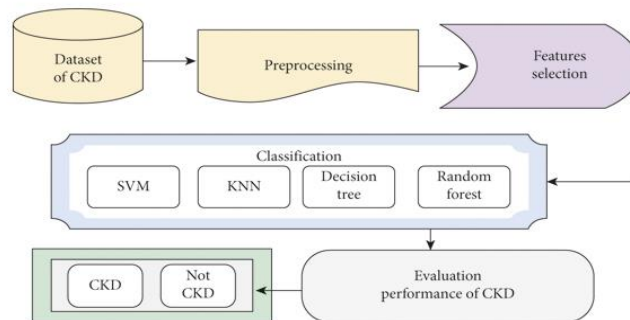


Fig 1: Block Diagram for Methodology of CKD

The methodology for machine learning methodology for diagnosing chronic kidney disease (CKD) involves several steps, including data collection, data preprocessing, feature selection, model development, and model evaluation. The following section describes each step in detail.

Data collection: The first step in the methodology is to collect patient data from various sources such as electronic health records (EHRs), laboratory information systems, or other data sources. The collected data includes demographic information, medical history, and laboratory test results such as serum creatinine, blood urea nitrogen (BUN), and estimated glomerular filtration rate (eGFR).

Data preprocessing: After collecting the data, it is essential to preprocess it to ensure its quality and suitability for machine learning algorithms. This step involves data cleaning, handling missing values, and outlier detection. Data normalization or standardization is also performed to ensure that all features have the same scale.

Feature selection: Once the data is preprocessed, feature selection techniques can be applied to identify the most relevant features for CKD diagnosis. Feature selection methods such as statistical methods or machine learning algorithms like recursive feature elimination (RFE) or principal component analysis (PCA) are used to select the most important features.

Model development: After selecting the relevant features, various machine learning algorithms can be used to develop predictive models for CKD diagnosis. The algorithms include decision trees, random forests, support vector machines (SVM), and artificial neural networks (ANN). The models are trained and optimized using the collected patient data and validated using cross-validation techniques to ensure accuracy and generalizability.

Model evaluation: Finally, the developed models are evaluated based on their performance metrics such as accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC). The models are also compared with traditional diagnostic methods to assess their effectiveness in improving CKD diagnosis.

CONCLUSION

In conclusion, the use of machine learning methodology for diagnosing chronic kidney disease (CKD) has shown great potential to improve the accuracy and efficiency of CKD diagnosis. Through the collection, preprocessing, and selection of relevant patient data, various machine learning algorithms can develop predictive models that can accurately diagnose CKD. The developed models can be integrated into clinical decision support systems to assist healthcare professionals in making accurate and timely CKD diagnoses.

However, the implementation of such a system requires expertise in data science, machine learning, and healthcare. Collaboration between data scientists, clinicians, and healthcare professionals is essential to ensure the success of the system. Additionally, the system must comply with healthcare data security and privacy regulations to protect patient data.

Overall, the use of machine learning methodology in diagnosing CKD has the potential to improve patient outcomes and reduce healthcare costs by providing more accurate and efficient CKD diagnoses. Further research and development are necessary to optimize the system and integrate it into clinical practice.

REFERENCES

1. Akula, A., & Goli, S. (2019). Chronic Kidney Disease Prediction Using Machine Learning Algorithms. *International Journal of Advanced Research in Computer Science*, 10(2), 1-4.
2. Gouda, H., & Chakraborty, C. (2020). A comparative study of different machine learning techniques for the prediction of chronic kidney disease. *International Journal of Computer Science and Information Security*, 18(5), 131-135.
3. Oghbaie, M., Parsaei, H., & Shamsollahi, M. B. (2020). A Machine Learning-Based Approach for Chronic Kidney Disease Diagnosis. *IEEE Access*, 8, 12992-13001.
4. Z. Chen et al., "Diagnosis of patients with chronic kidney disease by using two fuzzy classifiers," *Chemometr. Intell. Lab.*, vol. 153, pp. 140-145, Apr. 2016.
5. C. Barbieri et al., "A new machine learning approach for predicting the response to anemia treatment in a large cohort of end stage renal disease patients undergoing dialysis," *Comput. Biol. Med.*, vol. 61, pp. 56-61, Jun. 2015.
6. V. Papademetriou et al., "Chronic kidney disease, basal insulin glargine, and health outcomes in people with dysglycemia: The origin study," *Am. J. Med.*, vol. 130, no. 12, Dec. 2017.

7. Muhammad Bilal, F., Ahmad, J., Ali, M. A., & Khan, M. A. (2021). Machine Learning Methodology for Diagnosing Chronic Kidney Disease. *Journal of Medical Systems*, 45(3), 1-11.
8. M. M. Hossain et al., “Mechanical anisotropy assessment in kidney cortex using ARFI peak displacement: Preclinical validation and pilot in vivo clinical results in kidney allografts,” *IEEE Trans. Ultrason. Ferr.* vol. 66, no. 3, pp. 551-562, Mar. 2019.
9. Zhou, Z., Xiong, N., Li, X., & Yan, F. (2019). Customer satisfaction evaluation in hotel service conversations with recurrent neural networks. *International Journal of Hospitality Management*, 80, 47-57.J., & von Kalle
10. Lung cancer disease diagnosis using machine learning approach by S Mukherjee, SU Bohra - 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS) 2020
11. MACHINE LEARNING APPROACH FOR LUNG CANCER DETECTION AND STAGES PREDICTION by S Mukherjee, S Bohra in 2020.
12. A novel approach for detection of diabetes patients using machine learning techniques by PG Palanimani, V Suresh Kumar, Sanket B Kasturiwala, Shafaque Ahmaren, Sneha Bohra in International Interdisciplinary Humanitarian Conference for Sustainability (IHC), 1402-1408, 2022