Machine Learning for Human Intelligence
2024, Vol. 2, Issue 2

Research Article

A Hybrid Decision Support System for Chronic Kidney Disease Diagnosis

Fouzia Ameeen[1,⋆] and Ashfaq Ahmad[2]

[1,2]Department of Mathematics and Statistics, Emerson University, Multan, Pakistan.; Email: fouziaameeen@gmail.com, ashfaqch@gmail.com
⋆Corresponding author: Fouzia Ameeen (fouziaameeen@gmail.com)

Article History

Academic Editor:
Dr. Muhammad Sajid
Submitted: November 06, 2023
Revised: February 19, 2024
Accepted: September 1, 2024

Keywords:
Chronic Kidney Disease (CKD); Feature Extraction; Classifier Integration; E-Health

Abstract

The incidence of chronic renal problems is increasing markedly due to hypertension, diabetes, anemia, and other related conditions. Patients with these diseases may remain unaware of initial symptoms, complicating the diagnostic process. Advanced data mining diagnostic and prediction technologies could facilitate patient self-assessment and assist medical professionals in forming an accurate evaluation of the patient. This research presents a paradigm for a clinical decision support system for Chronic Kidney Disease (CKD) derived from the knowledge and insights of professionals and experts. Diverse classification techniques are utilized and evaluated on the dataset to identify the condition and ascertain the development stage of CKD. The proposed methodology improves accuracy to 91.75% and reduces the cost of forecasting CKD stages by employing LMT algorithms on the dataset.

1 Introduction

Chronic Kidney Disease (CKD) is a persistent illness characterized by the kidneys' inability to function properly. A study indicates that 1.2 million individuals perished globally in 2018 [1], and in Western nations, 5-12% of patients are afflicted with chronic kidney disease (CKD) [2]. The high prevalence of chronic kidney disease (CKD), substantial treatment costs, and irregular access to care present patients and their families with various financial and ethical dilemmas.

Creatinine is an essential parameter for detecting chronic kidney disease (CKD); it is a metabolic waste product generated by muscle metabolism. The normal range is 1.2 mg/dL for women and 1.46 mg/dL for men; however, elevated levels of creatinine are observed in the advanced stages of CKD. Chronic diseases are noncommunicable (NCDs), indicating they are not transmissible between individuals, whereas communicable diseases (CDs) can propagate from one individual to another and reproduce rapidly. Chronic disease arises from behavioral, biological, social, and environmental variables and may result in mortality [3, 9]. This condition is prevalent globally and affects individuals of all age demographics.

The human kidney is essential to bodily function, and associated disorders are typically chronic. The primary function of the kidney is to filter blood via millions of nephrons, eliminating undesirable substances from the human body. The accumulation of these undesirable substances results in chronic renal disease [4].

1 to 5 can be identified using various factors, including hypertension, diabetes, and anemia. Diabetes elevates blood sugar levels, resulting in nerve damage and vascular constriction. Anemia results in hypertension, a deficiency of red blood cells (RBCs), and reduced hemoglobin levels. Erythrocytes

deliver oxygen to bodily tissues. The supply of less oxygen can result in anemia. Anemia diminishes iron levels, reduces red blood cell count, and alters the morphology of red blood cells. Anemia is defined as a hemoglobin concentration below 12 g/dl in women and below 13 g/dl in males.

Another estimate [6] indicates that around one hundred individuals per million are afflicted by kidney illnesses alone. Recent statistics from the National Kidney Foundation (NKF) in 2019 indicate that the death rate of chronic kidney disease (CKD) surpasses that of breast cancer and prostate cancer. It is estimated that in the USA, 37 million individuals, about 90% of those with chronic kidney disease (CKD), are unaware of their condition. It is concerning that over 80 million individuals are at risk for chronic kidney disease (CKD). Recent observations indicate that individuals with kidney illness and transplant patients face an elevated risk of severe consequences from COVID-19.

Feature-based categorization utilizing machine learning techniques might be employed to identify the warning signals of chronic kidney disease (CKD). Classification can be performed utilizing various aspects of accessible data through data mining technologies to diagnose diseases and retrieve pertinent information. Expert systems are beneficial for practitioners since they may generate automated predictions based on the available data of patients, hence enhancing therapeutic outcomes. Numerous instances have demonstrated that expert systems can surpass human specialists in performance owing to their capability for multidimensional data processing [15, 16].

Diverse data mining methodologies are employed to extract insights from current databases and uncover comparable data for decision-making, evaluation, and forecasting [8]. Primarily, data mining employs descriptive and predictive models. Descriptive models classify patterns to examine the characteristics and relationships of data, whereas predictive models forecast outcomes based on many data sources.

This research is significant due to the public health implications of chronic kidney disease (CKD), the potential of machine learning in healthcare, and the critical role of early identification and intervention in enhancing patient outcomes in real-time settings.

This work aims to predict the stages of chronic kidney disease by extracting variables from a specified dataset. Consequently, we created a decision support system to aid both patients and physicians. This method delivers knowledge that may not be easily obtainable by human experts, without any delay. Our objective was to improve algorithm efficiency and precision.

2 Literature Review

Biological data has undergone substantial growth owing to various factors, including advancements in recording technologies, automated data creation techniques, the proliferation of medical facilities, and the rise in human population. Notwithstanding this growth, the death rate attributable to various diseases is concurrently increasing. A primary factor contributing to the elevated death rate is the inability to identify diseases in their early stages.

This section presents a concise overview of recently established categorization models for chronic kidney disease (CKD). Arulanthu and Perumal introduced an optimal cloud- and IoT-based decision support system for chronic kidney disease detection. The presented strategy employs a simulated annealing-based feature selection method with the RMSProp optimizer, referred to as SA-RMSPO-LR, to classify the presence of chronic kidney disease from healthcare information. Noor et al [5] proposed a primary healthcare scheme for the management of Chronic Renal Disease (CKD) patients, utilizing IoT technology. This scheme emphasizes dietary nutrition, particularly salt intake levels, patient logging, activity levels, water consumption, and sleep pattern monitoring to facilitate essential enhancements in their healthcare status. [10] developed a hybrid intelligent solution for chronic kidney disease prediction based on cloud IoT by employing two advanced technologies: neural networks and logistic regression. Neural networks are utilized for predicting chronic kidney disease (CKD). Logistic regression is employed to identify significant characteristics that affected chronic kidney disease.

3 Classification Methods

To effectively and accurately reassess the patient's healthcare status, each critical healthcare parameter and data collected by IoT sensors can be examined using machine learning methods within a predictive model, which has proven to be an efficient solution in prior medical detections. Data mining techniques, particularly classification methods, are frequently utilized as effective tools in anomaly detection and disease prediction within comprehensive research as an efficient strategy. The advancements in existing IoT devices and biomedical sensors have led to the establishment of several health tracking and smart medical care schemes [15]. The current study primarily focuses on the early diagnosis of many chronic diseases, including diabetes mellitus, heart disease, and chronic kidney disease (CKD), utilizing multiple elements that influence these conditions. However, the integration of each critical parameter required for disease prediction, the efficacy of the prediction approach, and the execution time continue to pose a difficulty [16].

3.1 Artificial Neural Network (ANN)

Consequently, the CC technique in healthcare aids stakeholders in storing patient medical records, accessing extensive patient data, applying illness predictions, and facilitating telemedicine, among other functions. Recently, the global community confronts challenges related to public health issues stemming from chronic diseases such as chronic kidney disease (CKD), which in turn escalates healthcare expenditures [18]. Due to threats and rising costs of CKD treatment, especially in developing countries, previous CKD forecasts present a major obstacle to healthcare facilities and physicians in this nation. The input layer receives environmental data as numerical representations, encompassing many data types such as images, text, numerical values, or speech. The input neurons are subsequently linked to hidden neurons, and the subsequent layer of hidden neurons calculates the hyperparameters for each connection. Each subsequent layer neuron calculates the total of products using the following equation:

$$\text{Output} = \sum_{i=1}^{n} w_i \cdot x_i + b \tag{1}$$

where w_i represents the weights, x_i represents the input, and b is the bias term.

3.2 Support Vector Machine (SVM)

Support Vector Machines (SVM) are extensively utilized in machine learning and pattern recognition applications owing to their superior performance relative to alternative machine learning techniques. This is a supervised learning method employed for regression and classification tasks. Linear SVM utilizes a hyperplane to distinguish between the two classes of data [18, 19].

The negative plane signifies a value less than 1 in the SVM methodology:

$$w^T x + b \leq -1 \tag{2}$$

The positive plane signifies values exceeding 1:

$$w^T x + b \geq 1 \tag{3}$$

The hyperplane classifier border is defined by:

$$w^T x + b = 0 \tag{4}$$

3.3 J48 Decision Tree Algorithm

J48 is a decision tree classifier, fundamentally a Java-based implementation of the C4.5 algorithm. It uses information theory to assess the distinct attributes of a certain dataset. The aim of J48 is to attain decision accuracy with adaptability. Decision tree pruning is the procedure that removes some nodes or branches from the tree without compromising the model's accuracy [20, 21].

3.4 Naïve Bayes (NB)

The Naïve Bayes algorithm is a supervised probabilistic classifier. Naive Bayes delineates conditional independence among characteristics. Let v_i represent the probability of a certain data instance belonging to class i; hence, for n classes, the probabilities are expressed as $V = \{v_1, v_2, \ldots, v_n\}$. The likelihood of a certain occurrence belonging to designated classes is expressed as:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \tag{5}$$

3.5 Logistic Model Tree (LMT)

LMT is a synthesis of the decision tree and logistic regression models. The decision tree is the most traditional method for classification, wherein each instance is categorized according to its parameter values. Entropy D is utilized to quantify data impurity. For a dataset of instances, entropy is defined as:

$$\text{Entropy}(D) = -\sum_{i=1}^{k} p_i \log_2(p_i) \tag{6}$$

The information gain for a specific characteristic is defined by:

$$\text{Gain}(D, A) = \text{Entropy}(D) - \sum_{v \in Values(A)} \frac{|D_v|}{|D|} \text{Entropy}(D_v) \tag{7}$$

3.6 Functional Tree

The functional tree constructs the decision tree and implements logistic regression at nodes and/or leaves [14]. The functional tree algorithm constructs the decision tree in two parts: conducting a depth-first traversal of the decision tree for the purpose of pruning, and assessing two metrics at non-leaf nodes for tree pruning [22, 23].

4 Proposed Methodology

The study advocates a multiphase methodology for constructing a decision support system (DSS) for the diagnosis of Chronic Kidney Disease (CKD), delineating the process into distinct phases to guarantee comprehensiveness and orderly advancement [24, 25].

4.1 Data Acquisition

This multivariate dataset is tailored for classification problems, consisting of 800 cases and 25 attributes, which include 11 numeric and 14 nominal features with multi-labels. The dataset is categorized into five stages of chronic kidney disease (CKD) and one normal condition. The five phases are as follows:

- CKD Stage 1: eGFR 90 or greater
- CKD Stage 2: eGFR between 60 and 89
- CKD Stage 3: eGFR between 30 and 59
- CKD Stage 4: eGFR between 15 and 29
- CKD Stage 5: eGFR less than 15

A distinct dataset (designated as TEST) is compiled from pertinent data sourced from local hospitals and clinics to evaluate the efficacy of classification algorithms. This test dataset comprises 66 occurrences and an equivalent number of characteristics.

4.2 Selection of Data Mining Tools

A data mining tool, Weka 3.6, has been selected to implement the data mining algorithms and associated data processing methods. The selection is made based on its characteristics including several integrated data mining techniques along with data preparation capabilities.

4.3 Data Preprocessing

This step involves the elimination of noise and the imputation of missing values from the dataset. Missing values are substituted with mean values. Various features of the dataset possess values within disparate ranges; hence, all such attributes are normalized to the range of 0 to 1.

4.4 Feature Prioritization

The subsequent phase in diagnosis involves feature ranking, wherein high-value features are prioritized in the mining process. Information gain and entropy theory are utilized for this aim. The most significant qualities encompass Serum Creatinine, blood urea, hemoglobin, specific gravity, hypertension, red blood cell count, and diabetes mellitus, which regulate the risk factors associated with kidney function.

4.5 Classification

Classification involves a two-step process: the mapping of features to labels and the implementation of a classification method. The classifier is trained using the training dataset, and its performance is assessed on the test/validation dataset. This study employs various categorization algorithms, with their performance evaluated based on precision, recall, and F-measure.

5 Results and Discussion

Cross-validation is a technique employed to yield outcomes with elevated confidence. We employed 10-fold cross-validation for our studies and present the average score. To enhance the reliability of our results, we utilized an additional dataset for testing purposes, which was obtained from local hospitals.

5.1 Experiment One: Binary Classification

Initially, we conducted binary classification. The outcomes for binary categorization are presented in Table 1.

Table 1: Results of the Binary Classification Dataset

Algorithms	**Precision (%)**	**Test Duration (seconds)**
LMT	98.0	0.99
FT	97.5	0.09
J48	99.0	0.01
Artificial Neural Network	99.75	5.05
Support Vector Machine	97.75	0.02
Naïve Bayes	95.0	0.01

The ANN approach attains 99.75% accuracy in binary classification, albeit it requires more time than alternative algorithms. While the ANN approach requires marginally more time than alternative classification algorithms, it provides superior accuracy.

5.2 Experiment Two: Multi-class Classification without Gender

In our second experiment, we performed multi-class classification on the chosen dataset. The results for multi-classification are presented in Table 2.

Table 2: Results of Multi-class Dataset-2

Algorithms	**Precision (%)**	**Test Duration (seconds)**
LMT	84.5	4.91
FT	84.5	0.2
J48	84.75	0.01
Artificial Neural Network	66.75	6.22
Support Vector Machine	67.0	0.2
Naïve Bayes	77.75	0.1

The J48 algorithm attains an accuracy of 84.75% in 0.01 seconds. Dataset-2 exclusively concentrates on the identification of CKD phases, omitting the gender characteristic. We noted that the gender feature is significantly important for CKD stages.

5.3 Experiment Three: Multi-class Classification with Gender

Dataset-3 comprises 27 attributes, one of which is gender. An equivalent number of instances is added for each gender to achieve class balance. The results are shown in Table 3.

Table 3: Results for Multi-class Dataset 3

Algorithms	**Precision (%)**	**Test Duration (seconds)**
LMT	91.75	10.65
FT	87.5	0.49
J48	85.875	0.03
Artificial Neural Network	76.625	14.0
Support Vector Machine	75.5	0.26
Naïve Bayes	81.5	0.01

This work proposes a hybrid knowledge modeling technique that integrates domain knowledge and patients' clinical situations for complicated decision-making regarding the commencement, adjustment, monitoring, or withdrawal of medicine. Additionally, we suggest a PII that offers a comprehensive overview of the patient data for a specified duration. The PII helps identify previous similar cases that resulted in positive patient-centric outcomes, such as improvements in laboratory results after following the prescribed medication regimen, thereby enabling the recommendation of the same regimens to similar patients. Medication dosage estimation is conducted on reference cases (derived from analogous patient cases) utilizing the IQR to aid clinicians in determining suitable dosages. The prevalence of chronic renal issues is significantly rising due to hypertension, diabetes, anemia, and other associated disorders. Individuals afflicted with these conditions may remain unaware of the initial symptoms, which complicates the diagnostic process. Advanced data mining diagnostic and predictive technologies could enable patient self-assessment and aid medical experts in making an accurate appraisal of the patient. This research introduces a framework for a clinical decision support system for Chronic Kidney Disease (CKD) based on the knowledge and insights of professionals and experts. Various categorization methods are employed and assessed on the dataset to determine the status and establish the progression stage of CKD. The proposed methodology enhances accuracy to 91.75% and decreases the cost of forecasting CKD stages by utilizing LMT algorithms on the dataset.

References

[1] Luyckx, V. A., Tonelli, M., & Stanifer, J. W. (2018). The global burden of kidney disease and sustainable development goals. *Bulletin of the World Health Organization*, 96(6), 369-440. doi: 10.2471/BLT.17.206441

[2] Versino, E., & Piccoli, G. B. (2019). Chronic kidney disease: The intricate history of long-term care organization and bioethics: Why action is needed now more than ever. *International Journal of Environmental Research and Public Health*, 16(5), 785. doi:10.3390/ijerph16050785

[3] Pandya, D., Nagrajappa, A. K., & Ravi, K. S. (2016). Evaluation and connection of urea and creatinine concentrations in saliva and serum of patients with chronic renal disease, diabetes, and hypertension: A research study. *Journal of Clinical and Diagnostic Research*, 10(10), ZC58.

[4] Charleonnan, A., Fufaung, T., Niyomwong, T., Chokchueypattanakit, W., Suwannawach, S., & Ninchawee, N. (2016). Predictive analytics for chronic kidney disease utilizing machine learning techniques. In *2016 Management and Innovation Technology International Conference (MITicon)* (pp. 80-83). IEEE.

[5] Polat, H., Mehr, H. D., & Cetin, A. (2017). Diagnosis of chronic kidney disease utilizing support vector machine via feature selection techniques. *Journal of Medical Systems*, 41(4), 55.

[6] Hill, N. R., Fatoba, S. T., Oke, J. L., Hirst, J. A., O'Callaghan, C. A., Lasserson, D. S., & Hobbs, F. D. R. (2016). Global prevalence of chronic kidney disease: A systematic review and meta-analysis. *PLoS One*, 11(7), e0158765.

[7] Yadollahpour, A., Nourozi, J., Mirbagheri, S. A., Simancas-Acevedo, E., & Trejo-Macotela, F. R. (2018). Developing and executing an ANFIS-based medical decision support system to forecast the course of chronic kidney disease. *Frontiers in Physiology*, 9, 1753.

[8] Mohapatra, S., Patra, J. K., Laishram, M., Thatoi, H. N., & Mohanty, S. N. (2018). Intelligent healthcare system utilizing data mining. In *2018 International Conference on Information Technology (ICIT)* (pp. 44-49). IEEE.

[9] Webster, A. C., Nagler, E. V., Morton, R. L., & Masson, P. (2017). Chronic kidney disease. *The Lancet*, 389(10075), 1238-1252.

[10] Hamedan, F., Orooji, A., Sanadgol, H., & Sheikhtaheri, A. (2020). Clinical decision support system for predicting chronic renal disease: A fuzzy expert system methodology. *International Journal of Medical Informatics*, 138, 104134.

[11] Ahmad, M., Tundjungsari, V., Widianti, D., Amalia, H., & Rachmawati, I. (2017). Diagnostic decision support system for chronic renal disease utilizing support vector machines. In *2017 Second International Conference on Informatics and Computing (ICIC)* (pp. 1-4). IEEE.

[12] Norouzi, J., Yadollahpour, A., Mirbagheri, S. A., Mazdeh, M. M., & Hosseini, S. A. (2016). Forecasting the advancement of renal failure in chronic kidney disease with an integrated intelligent fuzzy expert system. *Computational and Mathematical Methods in Medicine*, 2016, 6080814.

[13] Vijayarani, S., & Dhayanand, M. S. (2015). Classification algorithms in data mining for predicting kidney disease. *International Journal on Cybernetics and Informatics*, 4(4), 217-227.

[14] Gama, J. (2004). Functional trees. *Machine Learning*, 55(3), 219-250.

[15] Elhoseny, M., Shankar, K., & Uthayakumar, J. (2019). Advanced diagnostic prediction and classification system for chronic kidney disease. *Scientific Reports*, 9(1), 16940.

[16] Sobrinho, A., Silva, L. D., Pinheiro, M. E., Cunha, A. M., & Albuquerque, A. B. (2020). Computer-aided diagnosis of chronic kidney disease in developing nations: A comparative analysis of machine learning methodologies. *IEEE Access*, 8, 25407-25419.

[17] Subasi, A., Alickovic, E., & Kevric, J. (2017). Diagnosis of chronic kidney disease by using random forest. In *CMBEBIH 2017* (pp. 589-594). Springer.

[18] Kazakeviciute, A., & Olivo, M. (2017). Point separation in logistic regression on Hilbert space-valued variables. Statistics & Probability Letters, 128, 84-88.

[19] Patel, J., Upadhyay, T., & Patel, S. (2015). Heart disease prediction using machine learning and data mining technique. Heart Disease, 7(1), 129-137.

[20] Xing, C., Geng, X., & Xue, H. (2016). Logistic boosting regression for label distribution learning. In Proceedings of the IEEE conference on computer vision and pattern recognition.

[21] Siraj, F., & Abdoulha, M. (2017). Mining Enrollment Data Using Descriptive and Predictive Approaches. In Knowledge oriented approaches in data mining (pp. 54-68).

[22] Thiyagaraj, M., & Suseendran, G. (2019). Research of Chronic Kidney Disease based on Data Mining Techniques. International Journal of Recent Technology and Engineering (IJRTE), 8(2S11).

[23] Elhoseny, M., Shankar, K., & Uthayakumar, J. (2019). Intelligent Diagnostic Prediction and Classification System for Chronic Kidney Disease. Scientific Reports, 9(1), 1-14.

[24] Sobrinho, A., et al. (2020). Computer-Aided Diagnosis of Chronic Kidney Disease in Developing Countries: A Comparative Analysis of Machine Learning Techniques. IEEE Access, 8, 25407-25419.

[25] Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. Future Healthcare Journal, 6(2), 94.

[26] Bhargava, N., et al. (2013). Decision tree analysis on j48 algorithm for data mining. Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering, 3(6).