

Prediction of Abnormality in Kidney Function Using Classification Techniques and Fuzzy Systems



Mynapati Lakshmi Prasudha , Sukhavasi Vidyullatha ,
and Yeluri Divya 

Abstract Kidney diseases are life threatening. Its development is prevented by early detection and vigorous management. It is important to discover such disorders at an early stage in order to extend a patient's lifespan and to classify the abnormalities in kidney function based on pathological data. The primary goal is to identify the stages of the kidney disease and check the performance for various classifiers of the model. In this paper, classification algorithms are used to find out the accuracy of the supervised data. Not all machine learning classifiers predict the accurate results because of imprecision. So, fuzzy expert system (FES) is used to deal with imprecise data. To predict the disease at an early stage and also to identify the stages of the disease, FES is used. FES has shown promising results in identifying the stages of the patients. The accuracy of the pathological data is found by using machine learning algorithms. In addition, the probability of the occurrence of the disease is found by combining various parameters and identified the stages of the patient's disease.

Keywords Fuzzy expert system (FES) · Logistic regression (LR) · Support vector machine (SVM) · Decision tree (DT) · Random forest (RF) · Chronic kidney disease (CKD) · Acute kidney injury (AKI)

1 Introduction

Subsequent paragraphs, however, are indented. Kidneys are bean shaped organs in the human body located at the backside. Healthy kidneys are about 5 inches in size. The change in kidney size indicates an unhealthy kidney condition. Kidneys purify about 200 L of blood per day. The major function of the kidneys is to filter excess water, salts, and waste from the blood. The appropriate operation of this entire process is

M. Lakshmi Prasudha · S. Vidyullatha (✉) · Y. Divya
BVRIT HYDERABAD College of Engineering for Women, Bachupally, Hyderabad, Telangana,
India
e-mail: vidyullatha.1988@gmail.com

required to maintain electrolytes in a healthy level. Figures 1 and 2 show the healthy and unhealthy kidneys.

Ailments related to kidney are becoming more prevalent. Kidney damage happens slowly among many people over many years, generally as a result of diabetes mellitus or blood pressure, and it can be termed as CKD, whereas AKI happens when a person’s renal function changes suddenly due to illness, accident, or by the use of

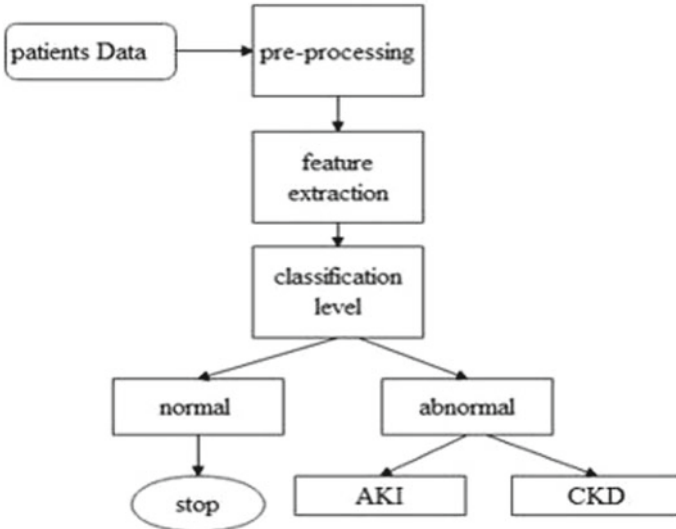


Fig. 1 Proposed model for kidney disease prediction [29]

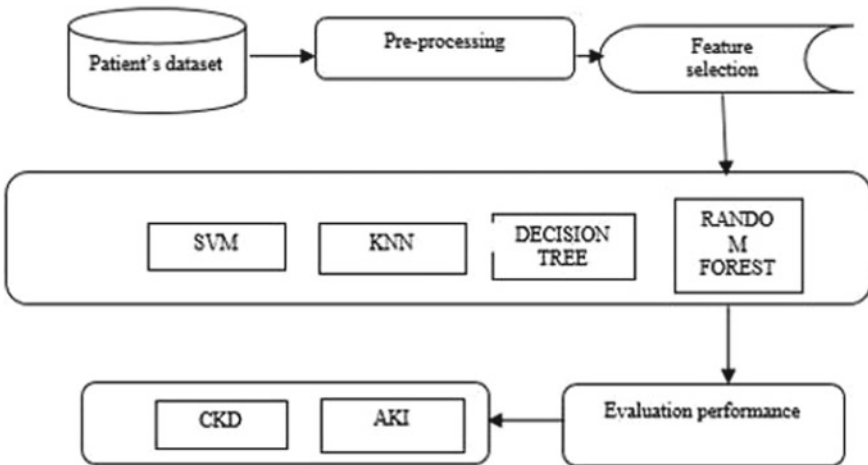


Fig. 2 Proposed framework for CKD and AKI prediction

certain drugs. This can affect the healthy people whose healthy kidneys or problems have related to kidneys. Chronic kidney disease (CKD) is usually a dangerous condition if not identified at an early stage. Its progression is prevented by early detection and effective management [1–6]. It is vital to discover such disorders at an early stage in order to extend a patient's lifespan. Kidney disease is a quiet and serious disease that affects people all over the world. It is harmful since the symptoms do not appear until the kidney's functions have deteriorated by 85–90%. According to Global Burden of Diseases (GBDs), over 1.2 million individuals died from kidney disease in some form. Since 2005, the proportion has raised by 32%, implying that the death rate of renal patients has increased by 32% over a ten-year period. According to the findings of the study, around 5–10 million individuals die each year as a result of kidney failure.

2 Literature Survey

To predict renal disorders, SVM and ANN were used [7, 8]. The study examined the accuracy and execution time of the two methods mentioned above. To develop a set of features that can predict kidney damage, effectively feature selection algorithms are employed. The reduced feature set reduces costs, improves efficiency, and eliminates ambiguity [9]. To predict at an early stage, the combination of machine learning algorithms and predictive modeling is proposed [10]. ANN models were assessed for predicting patient's lifespan, especially while suffering with CKD [11, 12]. K-means algorithm was used to extract information about how CKD markers interact with patient's mortality and analyzed clustering methods to predict dialysis patient's lifespan. By using Hadoop environment, different machine learning algorithms are used, and KNN and SVM with an AUROC 0.83 is achieved [13]. Gradient boosting algorithms and clinical information from EHR to present a one-year prediction model for CKD [14] among diabetic patients [15]. Convolutional auto-encoder is used to encode the temporal features, which exceeded baseline models by using EHR data containing sequences of lab test results to predict the risk of progressing from the first to the second stage of diabetic nephropathy [16]. The prediction model in kidney disease patients is proposed, especially for hypertension individuals using textual and numeric data from EHR. A neural network, based on bidirectional long- and short-term memories and auto-encoders, were used to encode both textual, numerical data. Under-sampling is used to balance the data and is able to get an accuracy of 89.7% using tenfold cross-validation [17]. Dataset containing missing values are dealt since it results in reduction of the model's accuracy and prediction outputs. They discovered a solution to this by performing a recompilation process on CKD stages, which resulted in unknown values. They recalculated missing data to fill up the gaps [18]. Using several machine learning classifications techniques, the authors worked on reducing diagnosis time and increased accuracy for the same. The classification of different stages of CKD based on severity is proposed. Using algorithms such as

the RBF, RF, and BPNN, the results shown that RBF algorithm performs better than other classifiers, with an accuracy of 85.3% [19, 20].

3 Proposed Work

3.1 *Kidney Disease Classification Using Machine Learning Based on Pathological Data*

Analyzing the medical data is a very sensible matter and that must be done correctly for disease prediction, detection, and analysis. This results in developing accurate tools and usage of such effective machine learning algorithms [21] which accurately detect or for diagnosing the disease. Appropriate and effective analysis of medical data has ushered in a revolution in machine learning field, especially for the widespread usage computationally demanding algorithms in recent years. However, existing number of clinical issues, namely accuracy, dependability, and rapid decision models, must be solved in order to guide physicians while diagnosing disease effectively [22, 23]. The classifiers' performance for disease prediction is determined by the obtained quality medical data and the classifier models used for the classification process. As a result, it is critical to employ various classifiers to correctly and accurately assess sensitive medical data in order to anticipate and detect diseases. In machine learning, classification [19, 20, 24–28] is a crucial challenge to extract knowledge from various real-time issues. Hence, a well-developed model shown in Fig. 1 is required to accurately predict the target class using collected data at multiple categorization levels.

3.2 *Early-Stage CKD Prediction Using Fuzzy Systems*

The major goal of developing this fuzzy expert system [30] is to assist doctors in detecting CKD in patients. This medical expert system can detect a disease and assist specialists in providing proper and appropriate treatment. Here, a patient's data is taken as input and classifying the stage of the patients is expected output. Inference and defuzzification are used to process the output as shown in Fig. 2 [7, 31–33]. The fuzzy system contains various methods like fuzzification, inference, and defuzzification for processing the output. A fuzzy expert system [34] is a conceptual framework used to diagnose and also to manage chronic kidney disease. The rule-based model receives its membership value through the defuzzification technique. This method converts output (linguistic values) into crisp values [35, 36].

4 Methodology

4.1 Dataset

We used a publicly available chronic kidney disease dataset from UCI repository. It contains 400 instances and 25 attributes.

4.2 ML Classifiers

Algorithm for proposed model

Input. Patient's dataset

Output. Correct classification of patient's dataset under various classification algorithms.

Step 1. Load dataset.

Step 2. Pre-processing the data.

- Row-elimination technique to deal with missing values.
- Convert the categorical values into numerical values

Step 3. Construct the classifier model (LR, DT, RF, and SVM) for preprocessed dataset.

Step 4. Performance analysis of constructed classifier models in step 3.

4.3 Classification Accuracy

Equation (1) is used to calculate the accuracy of given models:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (1)$$

where TP, TN, FP, and FN are observations and prediction values which are given in terms of true, false, positive, and negative.

4.4 Fuzzification

Medical diagnosis frequently requires a thorough examination of a patient in order to determine whether the patient is suffering from a suspected condition. If we consider sugar level, it may be high for one patient and low for another patient or no sugar for others. So here are combined features and its strengths to obtain an accurate diagnostic

conclusion. Here, physicians' experience is used in the current study to create a database of various fuzzy rules. Based on fuzzy decisions [37], a computer software can be developed to automatically evaluate if a patient with specific symptoms is suffering from one or other kind of a diseases.

The profile table can be determined as $[r_{pij}, rij, v]$.

$$\sigma_i = \frac{\sum_{j=1}^{j=ki} (W_{ij}\delta_{ij})}{\sum_{j=1}^{j=ki} W_{ij}} \quad (2)$$

Equation (2) is used to take a diagnosis decision by adding the impact of K_i relevant features by adding weighing factor w_{ij} . In this case, all the features will have equal weighted factor.

$$\sigma_i = \frac{1}{k_i} \sum_{j=1}^{j=ki} \delta_{ij} \quad (3)$$

Equation (3) is to obtain precise crisp numbers which indicates the probability of each disease in the set S .

For the given data, the first step is to perform normality test. The main risk factors for CKD are SCR, blood sugar, blood pressure, age, GFR, and smoke. Here, normality tests are performed for GFR and SCR because these are main factors for CKD prediction.

Normality. Normality check is very important while considering any pathological or numerical data because the obtained data contains lots of imprecision. To deal with imprecise data, normality check is must.

Confidence Indicator. CIs can be used to determine the ranges that will function as the fuzzy [38] sets in the outputs and input variables of a given model.

After normality test is done, to measure uncertainty in variables, a confidence indicator is used. Equation (4) for confidence interval is:

$$CI = \bar{X} \pm Z \frac{s}{\sqrt{n}} \quad (4)$$

CI = confidence interval.

Z = confidence level value.

s = sample standard deviation.

n = sample size.

IF-THEN-RULES (knowledge base). The fuzzy variables for output categorization are linked with set of rules in this step. Mandani fuzzy rule-based model is used to store fuzzy rules [39]. Different membership functions are selected and analyzed for certain results, such as parameter as normal, moderate, or critical, using the MATLAB-FIS editor. Finally, the condition of a patient is established using the

prepared rule bases, taking into account the status of individual parameters [40]. GFR = low (0–15), moderate (15–60), high (60 and above).

From the obtained data, to classify the stages of abnormality in kidney disease, six variables are considered. From these six variables, we get 324 rules.

Confidence indicator is used to estimate the performance using Eq. (5).

$$CI = \frac{\text{Success number}}{\text{total tests}} * 100. \tag{5}$$

5 Result Analysis

While calculating CI, we got 92% accuracy when fuzzy expert system is used. So, we can say that fuzzy system [41] can help many physicians for diagnosing CKD. Various metrics such as accuracy, precision, specificity, and sensitivity can be used for performance evaluation. In order to evaluate the performance metrics, Table 1 and Table 2 show the performance of classifiers and the confusion matrix must be reduced to 2 × 2 matrix and is shown in Table 3.

Table 1 Performance of various classifiers

Algorithms	Total predictions	Right predictions	Wrong predictions	Accuracy	precision	Recall	F1-score	Support
LR	210	193	17	94.56	0.98	0.91	0.95	162
SVM	210	168	42	88.52	0.79	1.00	0.89	162
DT	210	205	5	98.45	0.99	0.98	0.98	162
RF	210	206	4	98.75	0.99	0.98	0.99	162

Table 2 Confusion matrix for various classifiers

Classifier	TP	FP	FN	TN
LR	148	14	3	45
SVM	162	0	42	6
DT	159	3	2	46
RF	159	3	1	47

Table 3 2 × 2 matrix

Yes	No	Class
136	15	Yes
11	162	No

6 Conclusion

To predict the kidney disease at an early stage, the given input data was first classified into two levels, i.e., AKI and CKD using binary classification. A model was built using LR, SVM, DT, and RF. The random forest performed better while comparing to other classifiers, with an accuracy of 98.7%. In order to identify various stages of the disease, the attributes are combined to predict the outcome. By using fuzzy system, the inference rules were built and trained a model on these rules and achieved 96% accuracy when 200 rules were used for training using fuzzy expert system. So, it can be concluded that this can be helpful for many physicians in taking decisions to predict the severity of the disease at an early stage.

References

1. Lakshmi Prasudha M, Kasumolla R, Sukheja D (2021) Research reviews: towards identification and classification kidney disease using computational technology. In: 2021 5th international conference on computing methodologies and communication (ICCMC), pp 1387–1391. <https://doi.org/10.1109/ICCMC51019.2021.9418454>
2. Prasudha ML et al (2021) Comprehensive analysis of state-of-the-art CAD tools and techniques for chronic kidney disease (CKD). *IJBDAH* 6(2):1–12. <https://doi.org/10.4018/IJBDAH.287605>
3. Sivasankar E, Pradeep R, Sinandham S (2019) Identification of important biomarkers for detection of chronic kidney disease using feature selection and classification algorithms. *Int J Med Eng Inform* 11(4)
4. Aditya K, Babita P (2020) “A novel integrated principal component analysis and support vector machines-based diagnostic system for detection of chronic kidney disease. *Int J Data Anal Tech Strateg (IJDATS)* 12(2)
5. Pramila A, Eswaran P (2021) An efficient oppositional crow search optimization-based deep neural network classifier for chronic kidney disease identification. *Int J Innov Comput Appl (IJICA)* 12(4)
6. Khaleed MA (2021) Prediction of chronic kidney disease using different classification algorithms. *Inform Med Unlock* 24:100631, ISSN2352-9148. <https://doi.org/10.1016/j.imu.2021.100631>
7. Fazel Zarandi MH, Abdolkarimzadeh M (2022) Fuzzy rule based expert system to diagnose chronic kidney disease. In: Springer NAFIPS 2017 annual conference, vol 648, pp 323–328
8. Panwong P, Iam-On N (2021) Predicting transitional interval of kidney disease stages 3 to 5 using data mining method. In: 2016 second Asian conference on defence technology (ACDT), Chiang Mai, pp 145–150
9. Vijayarani S, Dhayanand S (2021) Kidney disease prediction using SVM and ANN algorithms. *Int J Comput Bus Res (IJCBR)* 6(2)
10. Aljaaf J et al (2022) Early prediction of chronic kidney disease using machine learning supported by predictive analytics. In: 2018 IEEE congress on evolutionary computation (CEC), Rio de Janeiro, pp 1–9
11. Zhang H, Hung C, Chu WC, Chiu P, Tang CY (2021) Chronic kidney disease survival prediction with artificial neural networks. In: 2018 IEEE international conference on bioinformatics and biomedicine (BIBM), Madrid, Spain, pp 1351–1356
12. Tazin N, Sabab SA, Chowdhury MT (2022) Diagnosis of Chronic Kidney disease using effective classification and feature selection technique. In: 2016 international conference on medical engineering, health informatics and technology (MediTec), Dhaka, pp 1–6

13. Kaur G, Sharma A (2022) Predict chronic kidney disease using data mining algorithms in Hadoop. In: 2017 international conference on inventive computing and informatics (ICICI), Coimbatore, pp 973–979
14. Al-Hayari AYA, Al-Tae MA (2021) Clinical decision support system for diagnosis and management of chronic renal failure. In: IEEE Jordan conference on applied electrical engineering and computing technologies, pp 1–6
15. Johansson M, Buijs JOD, Song X, Waitman LR, Yu AS, Robbins DC, Hu Y, Liu M (2020) Longitudinal risk prediction of chronic kidney disease in diabetic patients using a temporal-enhanced gradient boosting machine: retrospective cohort study. *JMIR Med Inform* 8:e15510 [CrossRef]
16. Katsuki T, Ono M, Koseki A, Kudo M, Haida K, Kuroda J, Makino M, Yanagiya R, Suzuki A (2018) Risk prediction of diabetic nephropathy via interpretable feature extraction from EHR using convolutional autoencoder. *Stud Health Technol Inform* 247:106–110
17. Ren Y, Fei H, Liang X, Ji D, Cheng M (2021) A hybrid neural network model for predicting kidney disease in hypertension patients based on electronic health records. *BMC Med Inform Decis Mak* 19:131–138. [CrossRef] [PubMed]
18. Dilli Arasu S, Thirumalaiselvi R (2021) Review of chronic kidney disease based on data mining techniques. *Int J Appl Eng Res ISSN 0973–4562* 12(23):13498–13505
19. Ramya S, Radha N (2022) Diagnosis of chronic kidney disease using machine learning algorithms. *Proc Int J Innov Res Comput Commun Eng* 4(1)
20. Polat H, Mehr HD, Cetin A (2021) Diagnosis of chronic kidney disease based on support vector machine by feature selection method. *Springer* 41(4):1–11
21. Michie D, Spiegelhalter DJ, Taylor CC (1994) Machine learning. *Neural Statistic Class* 12(12)
22. Sebasky M, Kukla A, Leister E, Guo H, Akkina SK, El-Shahawy Y, Matas AJ, Ibrahim HN (2009) Appraisal of GFR-estimating equations following kidney donation. *Am J Kidney Dis* 53(6):1050–1058
23. Jahantigh FF (2015) Kidney diseases diagnosis by using fuzzy logic
24. Sahani R, Rout C, Badajena JC, Jena AK, Das H (2021) Classification of intrusion detection using data mining techniques. In: Progress in computing, analytics and networking, Springer, Singapore, pp 753–764 CrossRefView Record in Scopus Google Scholar[4]
25. Das H, Naik H, Behera HS (2020) Classification of diabetes mellitus disease (DMD): a data mining (DM) approach Progress in computing, analytics and networking, Springer, Singapore, pp 539–549 CrossRefScopus Google Scholar
26. Dey N, Ashour A (2016) Classification and clustering in biomedical signal processing, IGI global Hershey, Google Scholar
27. Kamparia A, Saini G, Pandey B, Tiwari S, Gupta D, Kahna A (2021) KDSAE: Chronic kidney classification with multimedia data learning using deep stacked autoencoder network. Springer, pp 1–6
28. Hua C, Wu R, Kei C, An Wang S (2019) A cloud based fuzzy expert system for the risk assessment of chronic kidney disease. *Indrescience* 9(4)
29. Fig 1 and fig 2 (google images)
30. Ahmed S, Tanzir Kabir M, Mehmood NT, Rehman RM (2022) Diagnosis of kidney disease using fuzzy expert system. In: IEEE The 8th international conference on software, Dhaka, pp 1–8, April, 2022.
31. Ramesh R (2022) Chronic kidney disease prediction using machine learning models. 9:6364. <https://doi.org/10.35940/ijeat.A2213.109119>
32. Al-Hayari AYA, Al-Tae AA, Al-Tae MA (2018) Clinical decision support system for diagnosis and management of chronic renal failure. In: IEEE Jordan conference on applied electrical engineering and computing technologies, pp 1–6
33. Ahmed S, Tanzir Kabir M, Mehmood NT, Rehman RM (2021) Diagnosis of kidney disease using fuzzy expert system. In: IEEE the 8th international conference on software, Dhaka, pp 1–8
34. Fazel Zarandi MH, Abdolkarimzadeh M (2021) Fuzzy rule based expert system to diagnose chronic kidney diseases. In: Springer NAFIPS 2017 annual conference, vol 648, pp 323–328, September, 2021.

35. Zadeh LA (1965) Fuzzy sets. *Inf Control* 8:338–353 Article Download PDF Scopus Google
36. Himansu D, Bighnaraj NHS, Behera C (2020) Medical disease analysis using neuro-fuzzy with feature extraction model for classification. *Inf Med Unlocked* 18(1–12):100288. *Inform Med Unlock* 18:100299
37. Hua C, Kei Chiu R, An Wang S (2015) A cloud based fuzzy expert system for the risk assessment of chronic kidney disease. *Indrescience* 9(4)
38. Clinical decision support system to predict chronic kidney disease: a fuzzy expert system approach <https://doi.org/10.1016/j.ijmedinf.2020.104134>
39. Norouzi J, Yadollahpour A, Mirbagheri SA, Mazdeh MM, Hosseini SA (2022) Predicting renal failure progression in chronic kidney disease integrated fuzzy expert system. *Hindawi Comput Mathematic Methods Med* 2016:1–9
40. Shubhajit RC, Dipankar C, Hiranmay S (2008) Development of an FPGA based smart diagnostic system for spirometric data processing applications. *Int J Smart Sens Intell Syst* 1(4)
41. Zarandi MF, Abdolkarimzadeh M (2017) Fuzzy rule based expert system to diagnose chronic kidney disease North American fuzzy information processing society annual conference, Springer, pp 323–328