International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Ensemble Prediction of Chronic Renal Disease by Using Fuzzy Clustering Technique

Dr. P. Nithya[1], Dr. G. Sumathi[2], R. Vijayalakshmi[3]*

Submitted: 05/02/2024 **Revised**: 13/03/2024 **Accepted**: 19/03/2024

Abstract: The term "data mining" refers to the process of discovering previously unknown patterns in massive databases. It is possible to extract valuable medical information from the medical field's heterogeneous data, which includes text, graphics, and photographs. The severity of a patient's survival, illness, etc. following a medical condition can be predicted using medical data that reveals a pattern of the disease. An automated calculation was utilized to generate the patient data set utilized for the analysis of patients with renal illness. Predictions are employed in patients with renal illness based on past predictions. Since this pertains to the patient's life and an accurate result is required, conventional theory is preferable to the probability theory utilized to get the outcome. As the population ages, chronic kidney illness will only become worse. In order to give patients the best care possible, it is crucial to be able to detect and anticipate renal illness. The traditional methods employed to identify patients suffering from renal disease, as well as the outcomes of the traditional methods applied in the if-then rule and in conjunction with the generated agency. This novel approach takes the output data set as input and generates results using a combination of two fuzzy systems—neu-ral blur systems—and neural networks. Instead of using probabilistic neural networks, this new method combines fuzzy logic with other types of systems that provide mathematical conclusions. Mathematical computations typically yield more precise outcomes.

Keywords: *Data Mining, Fuzzy System, Renal Disease, Neural Network, Pre- diction.*

1. Introduction

The use of data mining techniques is widespread and has several applications in healthcare. A great volume of patients, illnesses, medical facilities, complex data, claims processing expenses, medical equipment, etc. have all been generated by the medical business, necessitating the handling and analysis of information extraction. In order to help healthcare providers make better judgments and enhance patient management, data mining offers a collection of tools and approaches that may be used to processed data. Based on the patient's medical history, current symptoms, diagnosis, and past treatments, a successful treatment regimen can be suggested for patients with comparable health issues.

A number of ailments, including chronic renal disease, are caused by people's present-day work patterns, diets, and ways of living. Diagnosing and treating chronic kidney disease (CKD) quickly is essential because it is a worldwide health concern that is becoming more common. Important for the proper functioning of many bodily systems, the kidneys filter out waste products and extra blood. A gradual decline in renal function, as seen in chronic kidney disease (CKD), makes it more difficult for harmful waste products to escape the body. In order to address the question of CKD in their most recent study, researchers employed data mining techniques.

To forecast the membership function of data instances, data mining techniques like classification and clustering are employed. Similar to clustering, classification likewise uses classes to break down information retrieval into its component elements. Applying target or predicted qualities, the algorithm processes the training data set that contains a collection of characteristics and specified outcomes in order to arrive at the forecast.

The field should pay attention to CKD because it has become a major concern on a global scale. When both kidneys are impaired, the body is unable to filter out harmful waste. In this study, we apply clustering techniques like FCM and ANN to the problem of detecting potentially fatal diseases like chronic kidney disease.

2. Literature Review

In order to adjust the health knowledge foundations of the care business without offering a formal knowledge test, Allen et al. [1] demonstrate a symbolic FCM clustering algorithm that incorporates fuzzy knowledge into its structure. The approach also includes better attributes and records, which provide higher accuracy throughout an extra range. The system can detect and forecast when a

1. Department of Computer Science, SRM arts &science college, Kattankulathur, Chennai – 603203, Tamilnadu, India. Email: nithyaraju.r@gmail.com
2. Department of Mathematics, CEG Campus, Anna University, Chennai - 600 025, Tamilnadu, India. Email: sumisundhar@auist.net
3. Department of Computer Science, Seethalakshmi Ramaswami college (Autonomous), Tiruchirapalli – 620002, Tamilnadu, India.*
Corresponding author: R.Vijayalakshmi
Email: giripriya710@gmail.com

patient will have renal failure. Instead of relying on medical tests administered by doctors to determine the severity of the grouped diseases, the suggested system can assist in predicting the aforementioned risk factors for renal failure. In order to gather patient records for renal disease, the FCM-related clustering method is used to identify individuals who have received inadequate therapy. Initial preprocessing information is derived totally from all boilerplate records with missing data added; the changing technique effectively connects the best cohorts with suitable tweaks to the various FCMs, discovering anomalies and legacy cases. The categorization step involves dividing the FCM classification into data regarding the severity of renal disease risk.

In order to understand the relationships between patient survival and several recorded factors, Chowdhury et al., [2] employed data transformation, data extraction, and data preparation techniques. Decision rules are the basis for two distinct data mining techniques. Algorithms for making decisions employ these guidelines to forecast how long new patients will live. Their significance in medicine explains why data mining is being used to uncover key factors. Using data gathered from four different dialysis centers, they have been developing and testing a novel concept detailed in a recent study article. A less time-consuming and expensive method is suggested in their work for choosing research subjects for clinical studies. Using the anticipated outcome and the most critical parameters identified, patients can be chosen.

In order to forecast the long-term results of a kidney transplant, Senan et al. [3] distinguished between logistic regression and artificial neural networks. Using ten datasets for training and validation, this study uses logistic regression and neural networks to forecast the specificity and sensitivity of artificial renal exclusion in renal transplant recipients. Clinical decision-making and the kidney transplant procedure could be enhanced by combining the two methods, which, according to the experimental results, are complementing algorithms.

According to Dovgan et al. [4], in order to ensure that patients receive appropriate and accurate therapy, it is crucial to be able to diagnose and forecast renal illness. The results of traditional renal illness detection systems are generated utilizing and/or mechanisms and if-then rules applied to patient datasets. This innovative method takes an input dataset and employs a neuro-fuzzy system, a hybrid of a fuzzy system and a neural network, to produce output. The new approach, which combines fuzzy logic with neural networks, generates outcomes not from probability theory but from mathematical computations. The accuracy of results is typically better when they are based on mathematical computations. An optimized version of the ANFIS system, which is thought to better collect valuable data, is showcased.

According to Sobrinho et al. [5], CKD is predicted in that study by combining ANN with naïve Bayesian categorization. Compared to artificial neural networks, Na-ive Bayes outperforms them in terms of accuracy. The identification and investigation of renal disease make extensive use of classification systems.

Data mining and its medical applications were shown on the screen by Makino et al. [6]. From that point on, data is gathered. When dealing with kidney disease, this has been carried out. These findings demonstrate the usefulness of data mining in healthcare and its potential to enhance numerous medical applications. Determining the number of groups in large datasets is an important aspect of the k-means (KM) algorithm. The ADT Naive Bayes J48 renal illness dataset, which is tree-based, was examined. Various machine learning algorithms (e.g., AD tree, J48, K-star, Bayesian naive, random forest, etc.) analyze data statistically and utilize algorithms to forecast the onset of renal illness.

3. Problem Formulation

The patient's life is at stake, so using probability theory to arrive at a conclusion would be a mistake. Predicting probability theory using outcome-based algorithms, such as statistical approaches, Bayesian classification, or association mechanism prediction, will yield false findings. New methods for predicting the cause of renal disease are a direct outcome of the reliability of the findings. Predicting diseases should be essential for improving hygiene and saving patients' lives. So

So far, it has been utilized to get findings in accordance with technical standards; nonetheless, the precision of these outcomes is far from satisfactory. Consequently, medical technology becomes more efficient because the demand for new technologies can be sensed more accurately.

4. Proposed Work

Renal disease is a growing problem in an aging population. Monitoring and prediction of renal disease is very important so that patients can receive good and appropriate treatment. Conventional systems are used to detect the use of datasets in renal disease patients, and whether to use the results of the presence or presence of rules. The disad- vantage of using this tradition is the high probability of getting more and more accurate results. Mis-predictions of the disease can lead to patients losing their lives. A new technique is proposed to predict and detect renal disease from patient datasets. This new technique uses two fuzzy systems and a neural network called a neuro-fuzzy sys- tem, which is obtained based on results

from an input dataset. This new system is a combination of fuzzy systems that produce the results of mathematical calculations, not a probabilistic neural network. Mathematical calculations tend to produce results with greater precision, thereby increasing the efficiency of the system. The proposed system is to accurately predict the development of renal disease.

5. Methodology

5.1. Fuzzy Model

Data is provided via fuzzy grouping in the form of an important graph belonging to each group pattern. Unlike other classification techniques, fuzzy modeling allows human reasoning models to develop and manage data despite technological ambiguity. Fuzzy logic's simplicity and flexibility are its key features. In conjunction with the conventional statistical model, a fuzzy system may represent any complicated nonlinear function, and fuzzy logic can handle cases with missing or erroneous data. Because of its ability to offer a more transparent model and its rule-based linguistic explanation, fuzzy modeling is often preferred. Actually, developing rules that could serve as informational clinical guidelines is one of the many advantages of fuzzy logic. When it comes to medical classification, various non-linear modeling strategies rely on fuzzy models to display comparison results..

5.2. Fuzzy C Means

Fuzzy c in many algorithms fuzzy clustering analysis data can be "fuzzified", but here only consider the fuzzy mean of K mean, called fuzzy c mean. In clustering, k means that the cluster is not used to update the cluster center of gravity, sometimes referred to as the c-mean, and is suitable for the fuzzy version of the community by blurring to the K-mean community. The fuzzy C-means algorithm is also called FCM.

5.2.1. FCM Algorithm

1 One must first choose an initial fuzzy pseudo-partition, which entails giving each wij a value.

2. Do it again.

3. Use the fuzzy pseudo-partition to get the centroid of every cluster.

4. Find the Wij, which is the fuzzy pseudo-partition, again.

5. as long as there is no change to the centroids

5.3. Artificial Neural Network (ANN)

For each tuple, the prediction network is compared with each known tuple via backpropagation, which iteratively processes the tuple training data set. For classification issues, the target values can be continuous values (predictions), or they could be well-known tag tuple classes. When training a tuple, the weights are adjusted so that the network's prediction falls as close as possible to the target value while minimizing the mean square error. The reason behind the name "back propagation" is that these adjustments are done in the "backwards" direction, meaning from the output layer all the way down to the first concealed layer. In most cases, the learning process will end when the weights converge, however this is by no means guaranteed. Inputs, outputs, and mistakes are the building blocks of the process. Still, if you get the hang of it, there's really not much of a learning curve.

5.3.1. Pseudo code for Backpropagation

Input: Dinput, a dataset including training tuples and their corresponding goal values; output, a multilayer feed-forward network; and learning rate, l.

Output: A trained neural network. Methods:

(1) Initialize all weights and biases in networks;

(2) While terminating condition is not satisfied {

(3) for each training tuple X in D {

(4) // Propagate the inputs forwards;

(5) for each input layer unit j {

(6) $O_j = I_j$; // output of an input unit is its actual input value.

(7) for each hidden or output layer unit j {

(8) $I_j = \sum_i W_{iji}\, O_i + \theta_j$ere/ compute the net input of unit j with respect to the previous layer, i

(9) $j = \frac{l}{i+e^{-1}{}_j}$;}//computer the output of each unit j

(10) // Backpropagate the errors;

(11) for each unit j in the output layer

(12) $Err_j = O_i(1 - O_j)\,(T_j - O_j)$; computer the error

(13) For each unit j in the hidden layers, for the last to the first hidden layer.

(14) $Err_j = O_i(1 - O_j) \sum_k Err_k\, W_{jk}$;//computer the error with respect to the next higher layer, k

(15) for each weigh $wi\ j$ in network {

(16) $\Delta W_{ij} = (1)\ Err_j\ 0_j$ //weight increment

(17) $W_{ij} = W_{ij} = +\Delta\, W_{ij}$ //weight update

(18) for each bias θ_j in network {

(19) $\Delta\, O_{j=}$ (1) Err// Bias increment

(20) $O_j = O_j + \Delta\, O_j$;//bias update

(21) }}

6. Experimental Results

The MathWorks Matlab software is used for this purpose. utilize MATLAB to work with data, functions, and matrices; develop algorithms, design user interfaces, and communicate with other languages' programs. A combination of precise performance measures and confused matrix grouping determines the C-mixing experimental blur.

6.1. Dataset

The data set is retrieved from the UCI library machine, which is considered the reference point. They are part of the machine learning community that uses the ML algorithm to empirically analyze database theory and data production in the UCI machine learning library. In the 1987 FTP file, David Aha and other UC Irvine graduate students created the document. Students, teachers, and researchers from all across the globe have been using it since then, and it's a big reason why data collection machines.

6.2. Fuzzy C Means

With its introduction by Ruspini and its extensions by Dunn and Bezdek, the fuzzy c-means has found extensive application in fuzzy clustering-clustering (FCM) applications such cluster analysis, pattern recognition, and image processing. It all starts with the fundamental principle of K-Means K indicates whether a data point is a member of a particular cluster or not, while FCM assigns each data point a degree of cluster membership. Therefore, FCM employs fuzzy partitioning to ensure that members range from 0 to 1 when they are capable of belonging to more than one group. Nevertheless, FCM continues to employ the cost function in an effort to partition the data set when minimizing, with the degree of membership provided by the degree of data points.

Permits a number between zero and one to appear in the member matrix U.

When the algorithm detects that there is no further improvement, it continues to iteratively process the previous two conditions. Here are the procedures that FCM follows when operating in batch mode to find the cluster centers i and c as well as the membership matrix U:

Step 1: To ensure that the requirement in Equation (1) is met, initialize the membership matrix U with random values ranging from 0 to 1.

Step 2: Calculate c fuzzy cluster centers, Ci ,i=(1……c), using Equation (3).

Step 3: In accordance with Equation (2), compute the cost function. Stop if it falls below a predetermined threshold for improvement over the previous iteration or if it falls below a specific tolerance value.

Step 4: Calculate a new U using Equation (4). Go to step 2.

Because the initial value of the membership matrix determines how well the FCM performs, it is advisable to execute the method multiple times, with varying degrees of data points belonging to the value each time.

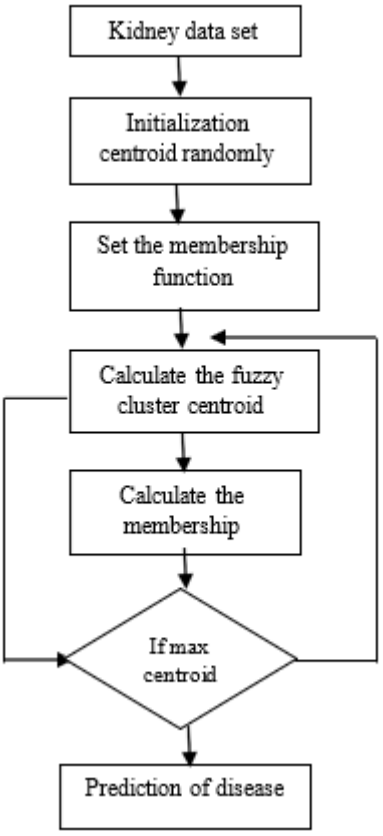

Figure 1 performance of proposed work

6.3. Fuzzification Score

For every value entered as a score in the associated table of the query's contents, the algorithm determines the fuzzy C meaning as the diffuse score. As the degree of similarity to the string increases, so does the anticipated score. If the fuzzy score is 1.0 or 0.9, then clustering is extremely dangerous. The associated symptoms are less affected or not at risk if the risk level is 0.0%. Each query is scored by the individual, and FCM is separated into two categories with the lowest and highest levels discovered again with the results within the supplied range of values. The user can enter the minimum and maximum risk factors that are set to call the doctor and the base. Determine the lowest and highest possible scores for their limitations. As a result, FCM can offer three low-risk scores: fuzzy average sub-risk, cluster-based outcomes, and fuzzy scores for discovering high-risk results.

6.4. Results

Clustering	No of	Sensitivity	Specificity	Accuracy
K-Means	450	92.33	93.47	92.33
K-Medoids	450	95.66	95.63	95.66
ANN	450	97.13	98.13	98.33
Fuzzy C Means	450	97.66	98.75	98.66

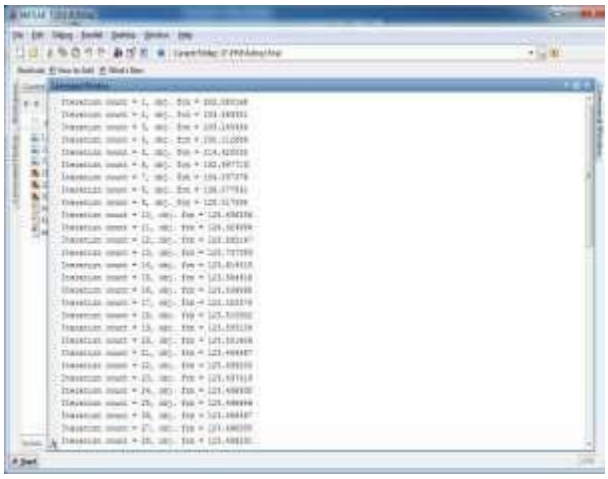

Figure 2 Objective function of Proposed Fuzzy C Means Method

For every value entered as a score in the associated table of the query's contents, the algorithm determines the fuzzy C meaning as the diffuse score. As the degree of similarity to the string increases, so does the anticipated score. If the fuzzy score is 1.0 or 0.9, then clustering is extremely dangerous. The associated symptoms are less affected or not at risk if the risk level is 0.0%. Each query is scored by the individual, and FCM is separated into two categories with the lowest and highest levels discovered again with the results within the supplied range of values. The user can enter the minimum and maximum risk factors that are set to call the doctor and the base. Determine the lowest and highest possible scores for their limitations. As a result, FCM can offer three low-risk scores: fuzzy average sub-risk, cluster-based outcomes, and fuzzy scores for discovering high-risk results.

7.Conclusion

In the proposed investigation, a projected association FCM clustering algorithm was used to localize the likelihood of maltreatment of patients with renal disease into the collected patient data. With the correct adjustment of the FCM classification, the strat- egy will effectively develop the best associated varieties, find outliers and traditional cases, and complete the initial preprocessing information by removing all duplicate rec- ords and adding missing data. In the classification phase, the FCM classification is classified into information about the risk level of renal disease. Clustering results on a pre- formed information set obtained from 450 patients showed that the FCM clustering algorithm accomplished higher accuracy than most existing procedures. It turns out that the expected performance of FCM is well known in terms of accuracy.

References

[1] Allen, Z. Iqbal, A. Green-Saxena et al., “Prediction of diabetic kidney dis- ease with machine learning algorithms, upon the initial diagnosis of type 2 diabetes mellitus,” BMJ Open Diabetes Research & Amp; Care, vol. 10, no.

[2] 1, p. e002560, 2022.

[3] N. H. Chowdhury, M. B. Reaz, F. Haque et al., “Performance analysis of Con- ventional machine learning algorithms for identification of chronic kidney dis- ease in type 1 diabetes mellitus patients,” Diagnostics, vol. 11, no. 12, 2021.

[4] E. M. Senan, M. H. Al-Adhaileh, F. W. Alsaade et al., “Diagnosis of chronic kidney disease using Effective classification algorithms and recursive feature Elimination techniques,” Journal of Healthcare Engineering, vol. 2021, p.1004767, 2021.

[5] E. Dovgan, A. Gradišek, M. Luštrek et al., “Using machine learning models to predict the initiation of renal replacement therapy among chronic kidney disease patients,” PLoS One, vol. 15, no. 6, p. e0233976, 2020.

[6] Sobrinho, A. C. M. D. S. Queiroz, L. D. Da Silva, E. D. B. Costa, M. E. Pinheiro, and A. Perkusich, “Computer-aided diagnosis of chronic kidney dis- ease in developing Countries: a comparative analysis of machine learning techniques,” IEEE Access, vol. 8, pp. 25407–25419, 2020.

[7] M. Makino, R. Yoshimoto, M. Ono et al., “Artificial intelligence predicts the progression of diabetic kidney disease using big data machine learning,” Sci- entific Reports, vol. 9, no. 1, pp. 1–9, 2019.

[8] N. A. Almansour, H. F. Syed, N. R. Khayat et al., “Neural network and support vector machine for the prediction of chronic kidney disease: a comparative study,” Computers in Biology and Medicine, vol. 109, pp. 101–111, 2019.

[9] Y. Hayashi, “Detection of lower albuminuria levels and early development of diabetic kidney disease using an artificial intelligence-based rule extraction Approach,” Diagnostics, vol. 9, no. 4, 2019.

[10] S. Ravizza, T. Huschto, A. Adamov et al.,

"Predicting the early risk of chronic kidney disease in patients with diabetes using real-world data," Nature Medi- cine, vol. 25, no. 1, pp. 57–59, 2019.

[11] T. R. Gadekallu, N. Khare, S. Bhattacharya et al., "Early detection of diabetic retinopathy using pca-firefly based deep learning model," Electronics, vol. 9, no. 2, pp. 1–16, 2020.

[12] K. Al-Rubeaan, K. Siddiqui, M. Alghonaim, A. M. Youssef, and D. AlNaqeb, "The Saudi Diabetic Kidney Disease study (Saudi-DKD): clinical characteris- tics and biochemical parameters," Annals of Saudi Medicine, vol. 38, no. 1, pp. 46–56, 2018.

[13] H. Polat, H. Danaei Mehr, and A. Cetin, "Diagnosis of chronic kidney disease based on support vector machine by feature selection methods," Journal of Medical Systems, vol. 41, no. 4, p. 55, 2017.

[14] O. Corporation, Machine Learning-Based Adaptive Intelligence: The Future of Cybersecurity Executive Summary. January, 2018.

[15] J. J. Khanam and S. Y. Foo, "A comparison of machine learning algorithms for diabetes prediction," ICT Express, vol. 7, no. 4, pp. 432–439, 2021.

[16] E.-H. A. Rady and A. S. Anwar, "Prediction of kidney disease stages using data mining algorithms," Informatics in Medicine Unlocked, vol. 15, p.100178, 2019.

[17] M. Sohail, H. M. Ahmed, M. Shabbir, and K. Noor, "Predicting chronic kid- ney disease by using classification algorithms in," WE!, vol. 11, no. 6, pp.1047–1050, 2020.