International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys.16 , No.1 (Jul-24)

http://dx.doi.org/10.12785/ijcds/160132

Ferritin Level Prediction in Patients with Chronic Kidney Disease using Cluster Centers on Fuzzy Subtractive Clustering

Linda Rosita[1,2], Sri Kusumadewi[3*], Tri Ratnaningsih[4], Nyoman Kertia[5], Barkah Djaka Purwanto[6] and Elyza Gustri Wahyuni[7]

[1]*Department of Clinical Pathology, Faculty of Medicine, Universitas Islam Indonesia, Yogyakarta, Indonesia*
[2]*Doctoral Program, Faculty of Medicine, Public Health, and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia*
[3,7]*Department of Informatics, Faculty of Industrial Technology, Universitas Islam Indonesia, Yogyakarta, Indonesia*
[4]*Department of Clinical Pathology and Laboratory Medicine, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia*
[5]*Department of Internal Medicine, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada,Yogyakarta, Indonesia*
[6]*Department of Internal Medicine, Faculty of Medicine, Universitas Ahmad Dahlan, Yogyakarta, Indonesia*

Received 18 Jan 2024, Revised 11 Mar. 2024, Accepted 6 Apr. 2024, Published 1 Jul. 2024

Abstract: It is important to know about iron reserves in patients on hemodialysis who have chronic kidney disease (CKD). Early detection of iron insufficiency or raised serum iron levels is crucial. Ferritin levels are one thing that can be used to cool this. Regretfully, ferritin testing is still seen as a highly costly procedure. Therefore, utilizing straightforward and affordable variables including height, weight, blood pressure, the duration of hemodialysis, history of comorbidities, and Hb levels before and after hemodialysis, this study predicts ferritin levels. The cluster center is used to help in ferritin level prediction. Due to the wide diversity of the sample data, the clustering technique is applied for clustering. Fuzzy subtractive clustering (FSC) was used to adaptively categorize 50 patient states using the dense concept. After clustering, we wound up with eight final clusters that had an accept_ratio of 0.75, a reject_ratio of 0.25, and an influence range of 0.5. The blood pressure variable has the strongest link with ferritin levels, according to the correlation coefficient. The mean degree of agreement between ferritin levels in the real and predicted samples was 62.53%. After evaluating nine sets of test data, the average similarity value was 83.74%. When data is clustered using the K-Means approach, the application of cluster centers yields a result of 50.91%; this result is significantly higher. This study's limitation is that it is unable to identify the ideal cluster in the presence of numerous outliers. Consequently, it is necessary to conduct additional study while taking the ideal number of clusters into consideration.

Keywords: fuzzy clustering, prediction, density, chronic kidney disease, ferritin

1. INTRODUCTION

One consequence of the metabolic syndrome is chronic kidney disease (CKD) [1][2]. According to reports, CKD affects 9.1% to 13.4% of people worldwide [3]. A significant residual risk of cardiovascular (CV) events and the advancement of CKD is a characteristic of CKD patients [4], [5]. In low- to middle-income nations, CKD screening is typically infrequently carried out in individuals with diabetes [6]. Anemia is a common side effect of severe chronic kidney disease (CKD) in most people [7], [8]. Iron deficiency is a major contributing factor to anemia associated with CKD. This can happen if there is an absolute or relative iron deficit that prevents the body from using its stored iron reserves [9]. Treatment for anemia in hemodialysis patients depletes iron reserves, which raises hemoglobin but causes iron deficiency [10].

Ferritin is a protein in the body that binds iron. Most of the iron stored in the body is bound to these proteins. Higher serum ferritin levels are associated with increased mortality in hemodialysis (HD) patients, which may be influenced by iron use and inflammation [11], [12]. For CKD patients receiving hemodialysis, ferritin level information is crucial. Decision-makers can utilize this information to guide their next steps. Regretfully, ferritin tests are still somewhat costly in underdeveloped nations. Predicting ferritin levels in CKD patients receiving hemodialysis becomes crucial as a result. These forecasts must to be supported by affordable, simple, and pertinent elements.

Several studies have been conducted on the prediction of chronic kidney disease (CKD) using intelligent systems. These include the use of machine learning to track the progression of CKD [13], [14], and intelligent prediction and classification systems to use the Ant Colony-based Optimization (D-ACO) method for CKD [15]. In the ma-

E-mail address: linda.rosita@uii.ac.id, sri.kusumadewi@uii.ac.id, triratnaningsih@ugm.ac.id, nyoman.kertia@ugm.ac.id, barkah.purwanto@uad.ac.id, elyza@uii.ac.id

jority of these three research, expensive or difficult-to-get variables or components were used. Prediction of iron reserves in the blood or Serum Iron (SI) has been carried out [16]. Likewise, total Iron Binding capacity (TIBC) prediction has also been carried out [17]. However, this prediction has not been made for Ferritin. Even though this prediction is also very important. In actuality, this can be accomplished by gathering information from patient serum ferritin test results in medical records.

The existence of outliers is one of the challenges in analyzing medical data. There are data that are very separated from other data. To assure the quality of data analysis, this outlier data needs to be identified and handled differently [18], [19], [20]. One method for processing and analyzing data that is frequently utilized is clustering. In order to support the decision-making process, the features of each cluster that was created as a result of the clustering findings are examined. In order to examine the heterogeneity of specific diseases, cluster analysis is being employed more and more in the medical profession. A hypothesis- or data-based approach is used in this investigation. According to several studies [21], [22], [23], it provides a workable method for categorizing entities with significant clinical variability. After clusters develop, inference is frequently performed [24]. Predictions or classifications are the goals of this inference.

Using fuzzy logic, Fuzzy Subtractive Clustering (FSC) is a popular clustering technique. This subtractive clustering method is unsupervised learning as opposed to semi-supervised learning like K-Means or Fuzzy C-Means (FCM) [25], [26], [26]. This method's ability to create clusters based on the maximum permitted distance between cluster members and the cluster center is one of its benefits. The cluster center is a component of the clustered data in this method. The aim of this study is to utilize the cluster center of FSC information to predict ferritin levels in hemodialysis patients with CKD. Easy and affordable variables are used in research. The three major contributions of this study were as follows: 1) collected a group of individuals with CKD and examined the characteristics of every group; 2) utilize the cluster center to determine ferritin levels in each group; and 3) propose an easy-to-obtain model for ferritin levels prediction. Group formation is carried out through a learning process, with the number of groups formed adjusted to the density of each dataset.

This paper is organized into four sections. Section 1 is an introduction that describes the background of the problem, objectives, and contribution of the study. Section 2 systematically describes the research methodology. Section 3 describes the results of the research and discussion. Finally, Section 4 is a conclusion and recommendation for further research.

2. METHODS

This study was completed in stages, as shown in Figure 1.

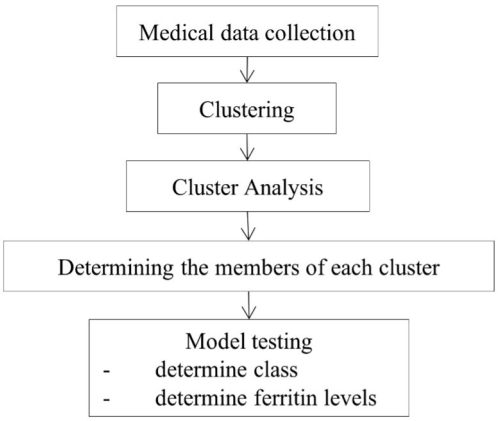

Figure 1. Research stage

Stage 1: medical data collection. Medical data collection was carried out on CKD patients undergoing hemodialysis therapy at two hospitals in Yogyakarta, Indonesia. The inclusion criteria included: 1) undergoing hemodialysis twice a week and having undergone it for at least three months from medical record data; 2) being at least 18 years old; 3) having Hb ¡ 13 g/dl for men and Hb ¡ 12 g/dl for women (WHO criteria); 4) being willing to follow the research from the results of the interview. While the exclusion criteria included: 1) laboratory examination results from peripheral blood smear showing thalassemia, hemolytic anemia, and megaloblastic anemia; 2) pregnant women; 3) history of blood transfusion in the last three months; 4) iron therapy in the last three weeks; 5) Have a history of infectious diseases, hematological diseases (hemolytic anemia, thalassemia, and megaloblastic anemia.

Stage 2: clustering. The clustering process is carried out in order to group data into several clusters that have certain characteristics. The clustering process is done using unsupervised learning. The FSC method was chosen because this method is proven to be able to adapt well to uncertain environments. Cluster centers are the end result of this clustering process. The clustering flowchart using the FSC algorithm is shown in Figure 2.

Since the collected data is not yet normal, it must be normalized in order to place all of the data within the range of 0 to 1. This is the normalization formula that is applied:

$$x_{ij} = \frac{x_{ij} - x_{min}}{x_{max} - x_{min}} \quad (1)$$

where xij is the $i-th$ data sample on the $j-th$ variable, x_{min} is the minimum data on the $j-th$ variable, and x_{max} is the maximum data on the $j-th$ variable. The process of clustering is initiated by determining the density value of each data point after the sample data has been standardized. For i = 1 to n, the density of X_i is called D_i, and it may be

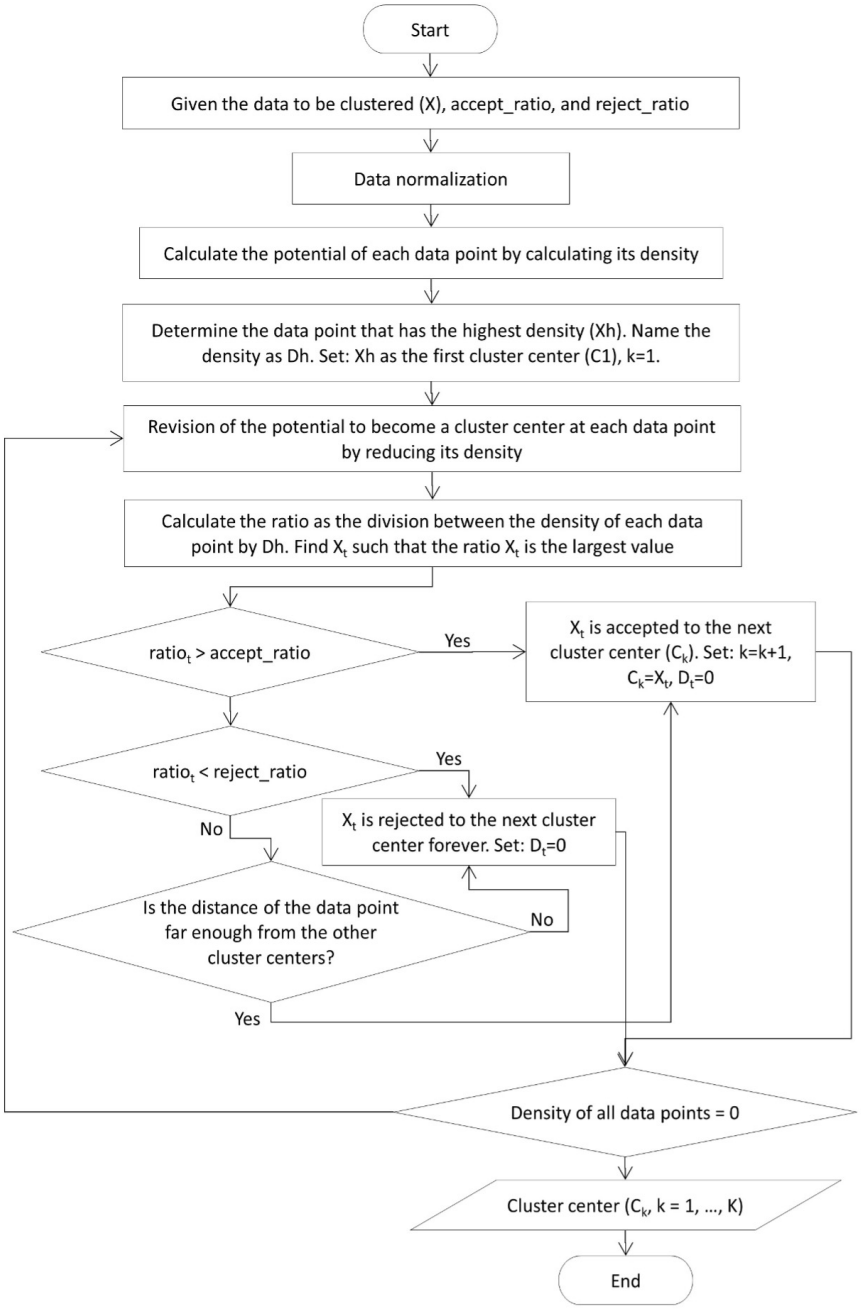

Figure 2. Flowchart of fuzzy subtractive clustering

computed using this formula [27]:

$$D_i = \sum_{t=0}^{n} exp(-\frac{4||X_t - X_i||^2}{r^2}) \quad (2)$$

Set the influence range (r) to 0.5. Since fewer clusters will form the closer to one, 0.5 was determined to be the ideal number. On the other hand, the larger the clusters will be, the fewer that form. A data point is said to have a large density if it has a large number of close neighbors. The cluster center will be chosen from among the data points with the highest density. The density value is called $D_t i$, and the data point with the biggest density value is called $X_t i$. Moreover, the first cluster center (C1) will be $X_t i$. This implies that the density of the surrounding data will be reduced by applying the following formula:

$$D_i = D_i - D_{ti}(exp(-\frac{4||X - X_{ti}||^2}{r_b^2})) \quad (3)$$

A positive number called r_b influences how much a cluster's data density decreases. Because $r_b > r$ in general, it can be calculated as

$$(qr_b)$$

, where q is a squash factor between 1,2 and 1,5. This implies that the density of the data will drastically drop toward the cluster's center. The data is therefore unlikely to succeed as the future cluster center. Next, the second cluster center (C2) will be searched after the updated data density. The density of every data point will be updated once more after C2 is determined, and so forth. In case D_k represents the maximum density during an iteration, the ratio for X_k can be computed as follows:

$$ratio = \frac{D_k}{D_h} \quad (4)$$

where D_h is the highest density at the beginning of the iteration. The parameters accept_ratio and reject_ratio are required in order to identify the subsequent cluster center. In each iteration, one of three circumstances could apply when X_k wishes to be designated as the cluster center: 1) X_k is accepted as the new cluster center if the ratio is greater than the accept_ratio; 2) If the ratio is greater than both the accept_ratio and the reject_ratio, X_k will be accepted as a new cluster center, but only provided that it is sufficiently far away from other cluster centers. If not, X_k will be rejected as the cluster center and will not be taken into consideration going forward; 3) If the ratio is equal to or less than the reject_ratio, then X_k is not taken into consideration as a candidate for the cluster center. The iteration ends up coming a close if there is no more data suitable for being a cluster center.

Stage 3: cluster analysis. Cluster analysis is done to assist decision makers in interpreting the characteristics of each cluster. The analysis was conducted using a data-driven approach. We use Matlab software as a tool for computing.

Stage 4: determining the members of each cluster. Members of each cluster will be assigned after they have been formed. A single data point is restricted to belonging to a single cluster. Equation (5), which represents the Euclidean distance formula, is utilized in this context:

$$d_{ik} = \sqrt{\sum_{j=1}^{m}(x_{ij} - v_{kj})^2} \quad (5)$$

where d_{ik} is the distance of the i-th sample data from the k-cluster center, x_{ij} is the i-th sample data in the j-variable, and v_{kj} is the k-th cluster center in the j-variable. The i-th sample data will occupy the k-cluster if the distance between the i-th data and the k-cluster center is the shortest distance.

Stage 5: model testing. Model testing is done in two stages. First, classify the testing data to determine the appropriate cluster. The shortest distance is used in this process. Second, after the cluster information is obtained, the appropriate ferritin level will be determined. The ferritin level in the output model (Y) is compared with the real ferritin level obtained based on the measurement results (F). If $T_i - \theta \leq Y_i \leq T_i + \theta$, it is said that the model output is in accordance with real conditions.

3. RESULT AND DISCUSSION

The data collection yielded 59 complete data samples. A total of 50 data points will be clustered, with nine serving as testing data. Data taken from each patient included height, weight, systolic blood pressure, diastolic blood pressure, duration of hemodialysis, hemoglobin level before hemodialysis, hemoglobin level after hemodialysis, history of comorbidities (diabetes, hypertension, and urinary tract stones), and ferritin level. BMI was obtained from height and weight; meanwhile, the increase in Hb was obtained from hemoglobin levels before and after hemodialysis. The data obtained from the measurement results are shown in Table 1.

Based on the data in Table 1., the H and W variables are used to calculate Body Mass Index (BMI) with the following formula:

$$BMI = \frac{W}{(H)^2} \quad (6)$$

While the HbPre and HbPost variables are used to calculate the increase in the patient's Hb level before and after hemodialysis with the following formula:

TABLE I. Data set

Variable	Unit
Input	
Height (H)	m
Weight (W)	kg
Systolic Blood Pressure (SBP)	mmHg
Diastolic Blood Pressure (DBP)	mmHg
HD duration (HDD)	months
Hypertension comorbidities (HT)	1: yes, 0: no
Diabetes comorbidities (DM)	1: yes, 0: no
Urinary tract stone comorbidities (UTS)	1: yes, 0: no
Hb prehemodialysis (HbPre)	gram/dl
Hb posthemodialysis (HbPost)	gram/dl
Output	
Feritin (F)	ng/ml

$$dHb = HbPost - HbPre \quad (7)$$

The data sample profile based on gender and ferritin levels is shown in Figure 3 and 4. The data samples were balanced between male and female patients. As output data, the highest ferritin level is between 15 – 200 ng/ml, which still shows normal ferritin levels.

Figure 5. shows the detailed data distribution for each input and output variable. Several data points are seen as outliers, such as three data points on the BMI variable, two data points on the DBP variable, four data points on the HDD variable, two data points on the dHb variable, and two data points on variable ferritin levels.

In completing out the clustering process, it begins with determining the parameters accept_ratio and reject_ratio. These two parameters really determine the number of clusters formed. We tried a combination of several values for these two factors with the help of factor alpha () with a value of 0 – 0.3 (Figure 5). The smaller the accept_ratio value or the smaller the reject_ratio value, the more clusters will be formed. The combinations we did were: 1) set an accept_ratio of 1- and a reject_ratio of (Figure 5a); 2) set accept_ratio of 1- and reject_ratio = 0.25 (Figure 5b); and 3) set accept_ratio = 0.75 and reject_ratio of (Figure 5c). From Figures 5a and 5c, it can be seen that in the range of 0.2 0.25 there is a very significant decrease in the number of clusters. The same thing is shown in Figure 5b, by setting = 0.25 a stable condition is obtained, which produces a total of eight clusters.

In this study, the parameters accept_ratio = 0.75 and reject_ratio = 0.25 were used. Eight clusters were obtained with cluster centers as shown in Table II. Eight data points had the highest density and were selected as cluster centers, namely data 30, data 31, data 20, data 41, data 1, data 47, data 23, and data 18. Each cluster has a center from C1 to C8. Figure 6a shows the location of the cluster center on the SBP variable, and Figure 6b shows the location of the cluster center on the ferritin level output variable.

The first cluster (Cluster 1) is a group of patients who are overweight, have high blood pressure, have a short duration of hemodialysis, have diabetes comorbidities, and have a significant increase in hemoglobin levels during the hemodialysis process. The ferritin level for Cluster 1 was 90 ng/mL. The second cluster (Cluster 2) is a group of patients who are obese grade 1, have high blood pressure, have a long duration of hemodialysis, have diabetes comorbidities,

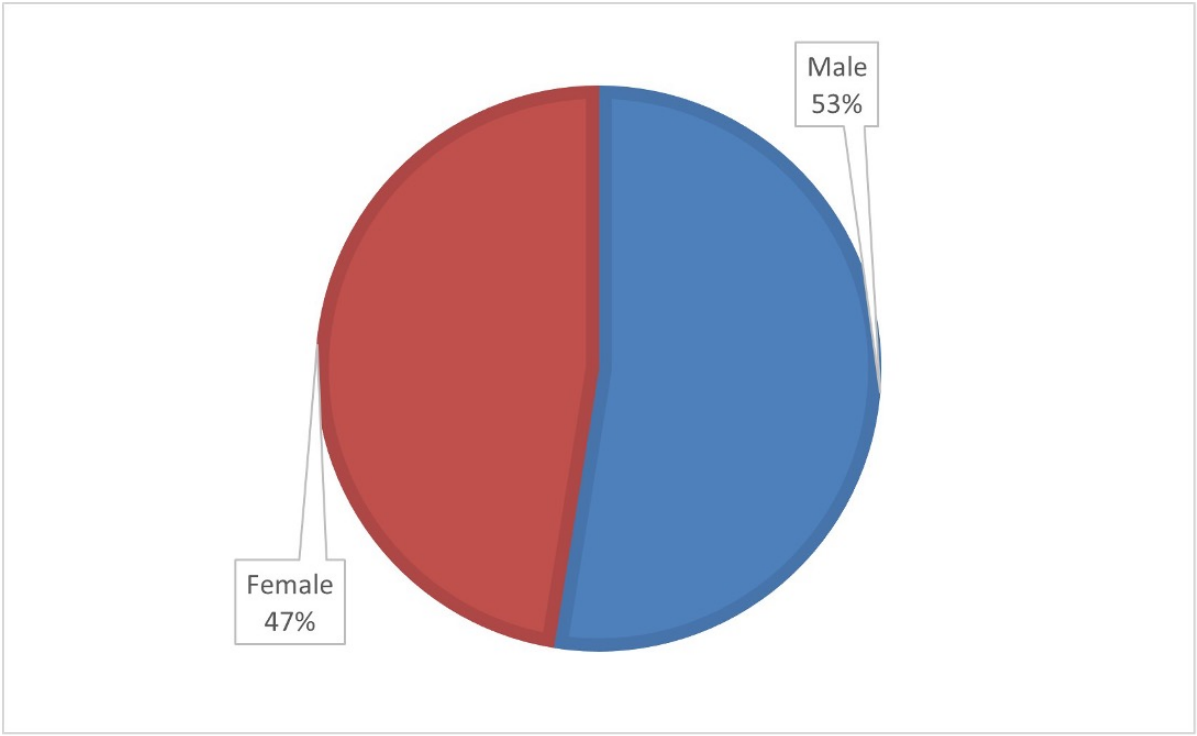

Figure 3. Data sample profile based on: (a) gender

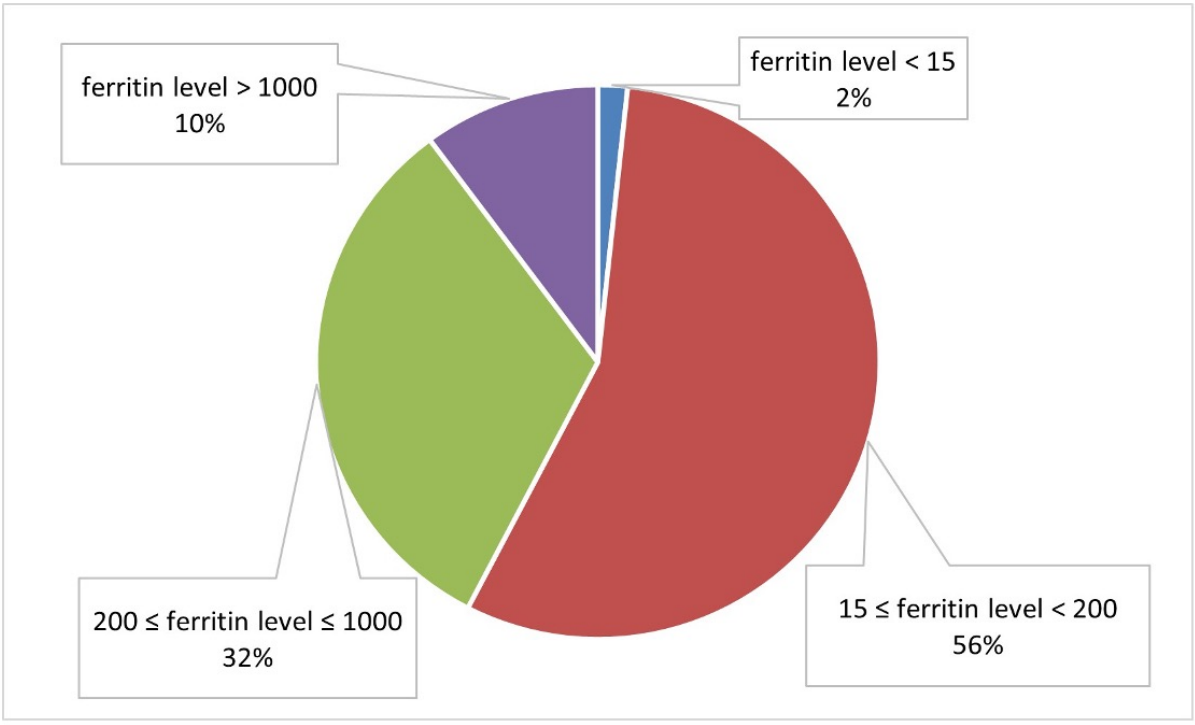

Figure 4. Data sample profile based on: (b) ferritin levels

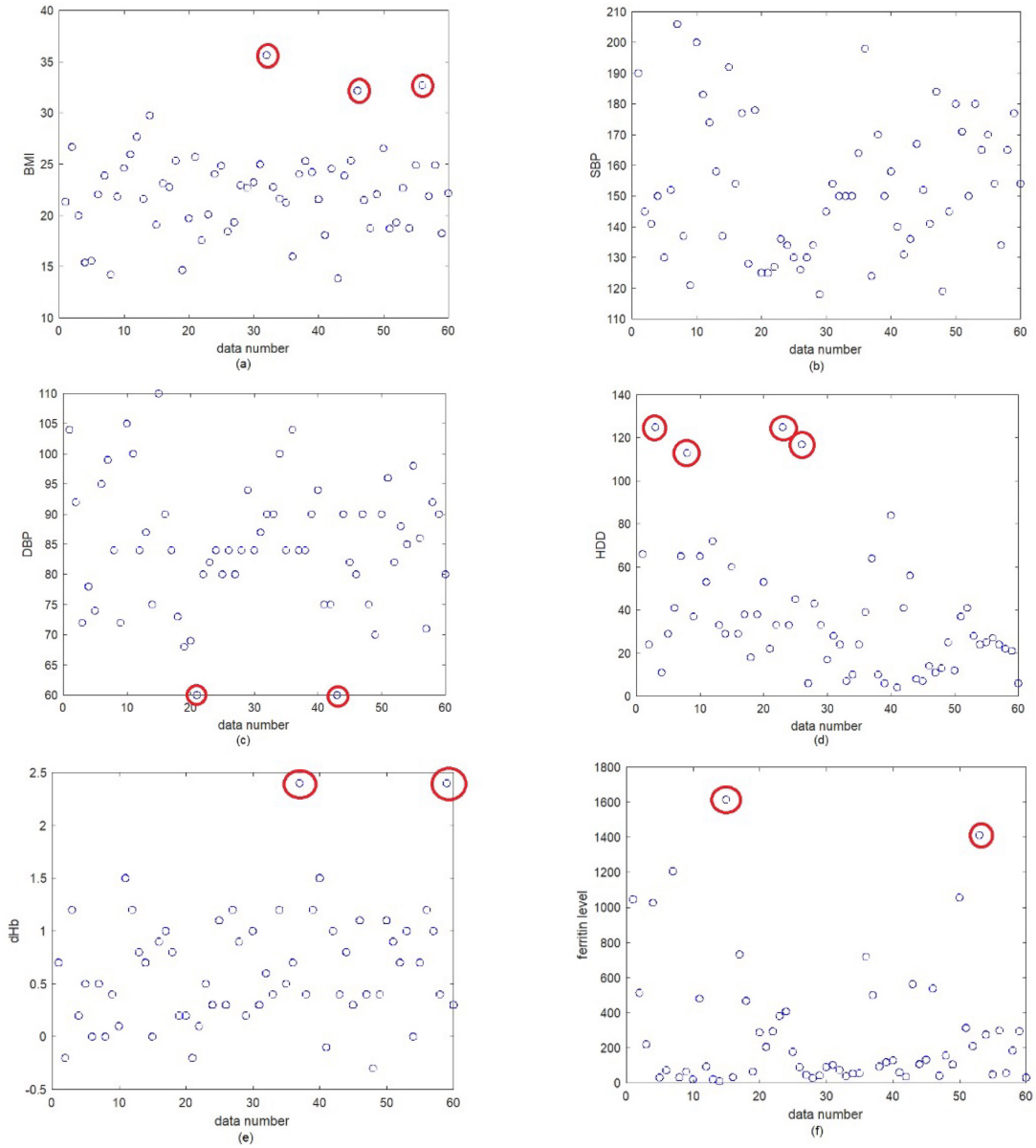

Figure 5. Distribution of sample data in each variable: (a) BMI; (b) SBP; (c) DBPs; (d) HDDs; (e) dHb; and (f) Ferritin level

TABLE II. Cluster Center and Standard Deviation

Cluster	i-th data as cluster center	IMT x_1	SBP x_2	DBP x_3	HDD x_4	HT x_5	DM x_6	UTS x_7	dHb x_8	Ferritin x_9
C1	30	23.20	145	84	17	0	1	0	1.00	90.00
C2	31	25.00	154	87	28	0	1	0	0.30	103.50
C3	20	19.70	125	69	53	0	0	0	0.20	288.30
C4	41	18.10	140	75	4	1	1	0	-0.10	60.20
C5	1	21.30	190	104	66	1	1	0	0.70	1047.00
C6	47	21.50	184	90	11	0	1	0	0.40	41.40
C7	23	20.10	136	82	125	0	1	0	0.50	381.80
C8	18	25.30	128	73	18	0	0	1	0.80	468.00

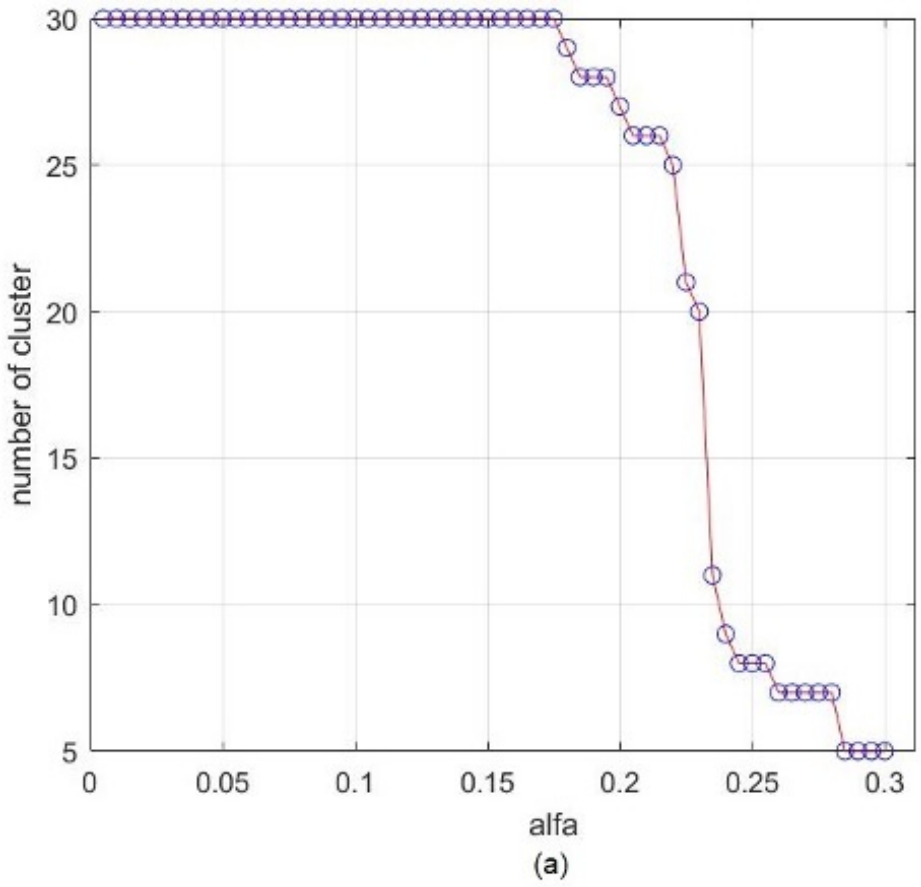

Figure 6. Number of clusters based on decrease/increase accept_ratio and reject_ratio: (a) accept_ratio = (1-) and a reject_ratio = ;

and have an increase in hemoglobin during the hemodialysis process. The ferritin level for Cluster 2 was 103.5 ng/ml. The third cluster (Cluster 3) is a group of patients with normal nutritional status, normal blood pressure, a long duration of hemodialysis, no comorbidities, and an increase in hemoglobin during the hemodialysis process.

The ferritin level for Cluster 3 was 288.3 ng/ml. The fourth cluster (Cluster 4) consists of thin patients with high blood pressure, very short hemodialysis duration, diabetes comorbidities and hypertension, and decreased hemoglobin during the hemodialysis process. The ferritin level for Cluster 4 was 60.2 ng/ml. The fifth cluster (Cluster 5) consists of patients who have a normal nutritional status, extremely high blood pressure, a long duration of hemodialysis, co-morbid diabetes and hypertension, and a significant increase in hemoglobin during the hemodialysis process. The ferritin level for Cluster 5 was 1047 ng/ml. This ferritin level is very high, so patients who fall into this group need special attention. Ferritin levels of more than or equal to 1000 ng/ml indicate an accumulation of iron in the body [28]. The sixth cluster (Cluster 6) is a group of patients with normal nutritional status, very high blood pressure, short duration of hemodialysis, co-morbid diabetes, and an increase in hemoglobin during the hemodialysis process. The ferritin level for this Cluster 6 was 41.4 ng/ml. The seventh cluster (Cluster 7) is a group of patients with normal nutritional status, high blood pressure, a very long duration of hemodialysis, co-morbid diabetes and hypertension, and a significant increase in hemoglobin during the hemodialysis process. The ferritin level for Cluster 7 was 381.8 ng/ml. The eighth cluster (Cluster 8) is a group of patients who are overweight, have normal blood pressure, have a long duration of hemodialysis, have UTS comorbidities, and have a significant increase in hemoglobin during the hemodialysis process. The ferritin level for Cluster 8 was 468 ng/ml. This ferritin level is quite high, so patients who fall into this group need special attention. The members of each cluster are shown in Figure 7. Cluster 2 has the most members, with 11 data, while Cluster 7 has the fewest, with just one.

To determine the relationship of each variable to ferritin levels, information from the correlation coefficient can be used. The correlation coefficient between each variable and ferritin before and after clustering is shown in Figure 8. Before clustering, the variables with the highest correlation coefficient values were SBP, DBP, DM, HDD, UTS, dHb,

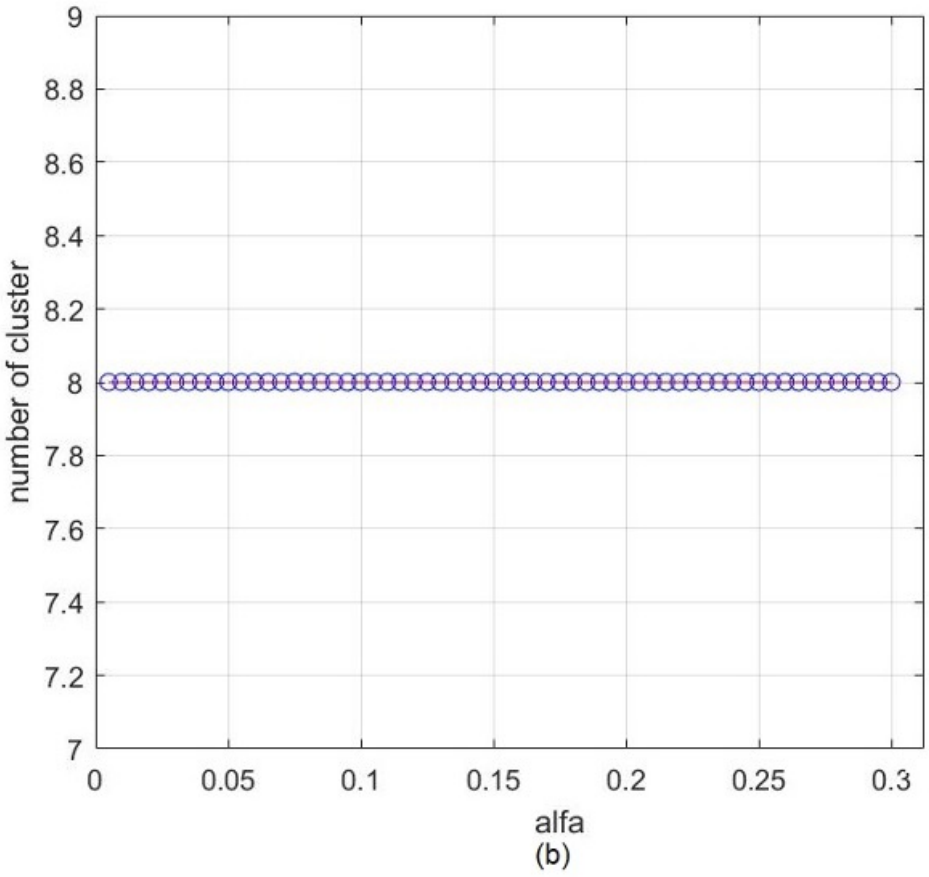

Figure 7. Number of clusters based on decrease/increase accept_ratio and reject_ratio: (b) accept_ratio = (1-) and reject_ratio = 0.25;

HT, and BMI. After clustering, the largest correlation coefficient values are in DBP, HDD, DM, dHb, SBP, UTS, BMI, and negative values are in HT. The decrease in the correlation coefficient only occurs in the SBP and HT variables.

The system output is derived from the center of the cluster on the ferritin level variable, in contrast to previous inference techniques that start with clustering and use a particular function to get the system output. This is done to anticipate the presence of outlier data. The weighted average idea is incorrect in this situation since the clustering results show the uniqueness and uniqueness of each cluster.

Training and testing data are used to conduct model testing. We calculate the similarity of the results. The ferritin levels from the model output are compared to the actual ferritin levels obtained from the measurement results to conduct the test. The similarity value is equal to the model output divided by the real data if the model output is less than or equal to the real data. On the other hand, the similarity value is equal to the real data divided by the model output if the real data is less than the model output.

The test results from the training data are shown in Figure 9.a. In the training data, the average similarity level is 62.53%. The main reason for this low level of similarity is that three data points have ferritin levels that are quite far from the ferritin levels in the cluster center. The three data points are: 1) The 8th data point, with a ferritin level of 32 ng/ml, is in a cluster with a central cluster ferritin level of 90 ng/ml; 2) The 14th data point, with a ferritin level of 9.80 ng/ml, is in a cluster with a ferritin level of 60.20 ng/ml; and 3) The 28th data point, with a ferritin level of 27.8 ng/ml. If these three variables are removed from the training data, the similarity value becomes 77.42%. The test results for the testing data are shown in Figure 9.b. and Table III. In the testing data, the average similarity level is 83.74%.

A comparison of test results using the K-Means approach is presented as well in Table III. We perform clustering with K-Means for the value K = 8. It is adjusted for the number of clusters acquired using FSC. An average similarity of 50.91% can be seen in the test results using K-Means.

Using these results, we can predict ferritin levels in CKD patients on hemodialysis by using information on

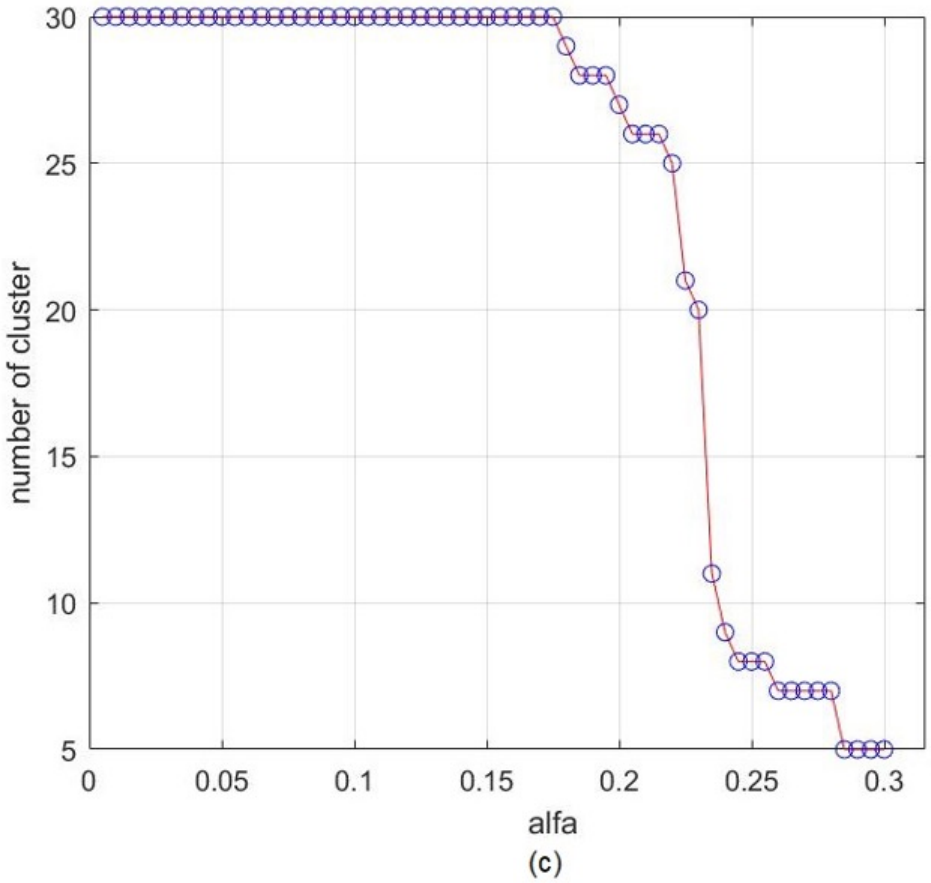

Figure 8. Number of clusters based on decrease/increase accept_ratio and reject_ratio: (c) accept ratio = 0.75 and reject ratio of α

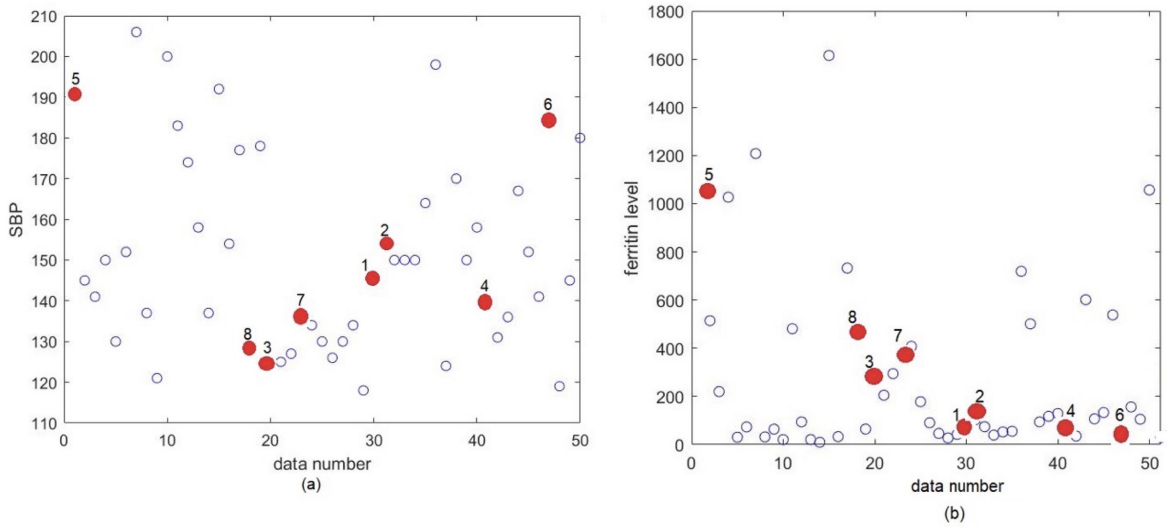

Figure 9. Eight cluster centers for accept ratio = 0.75 and reject ratio = 0.25: (a) SBP input variable; (b) ferritin level output variable

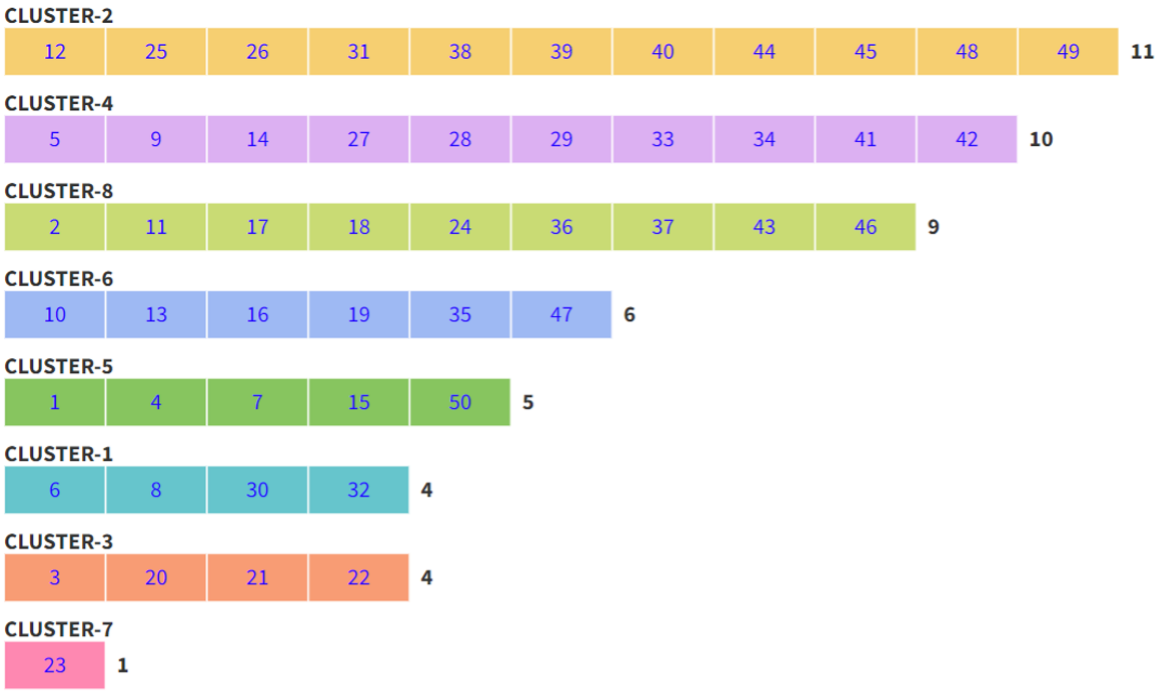

Figure 10. Cluster members

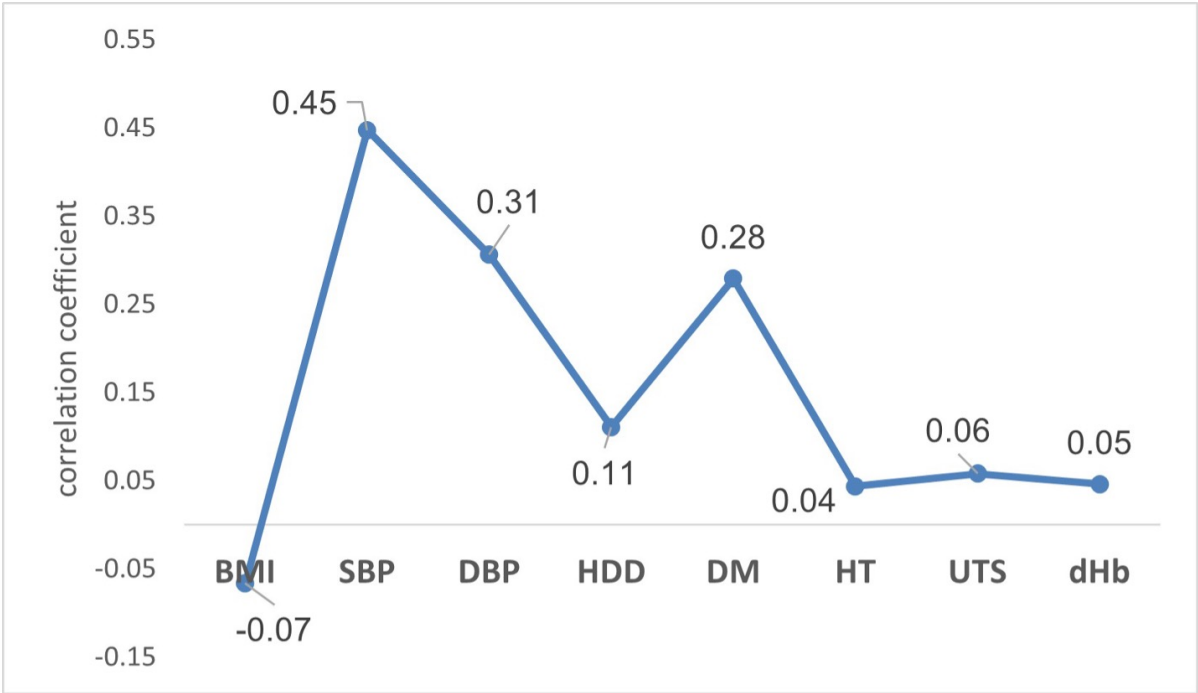

Figure 11. Correlation coefficient with ferritin: (a) all data sample before clustering;

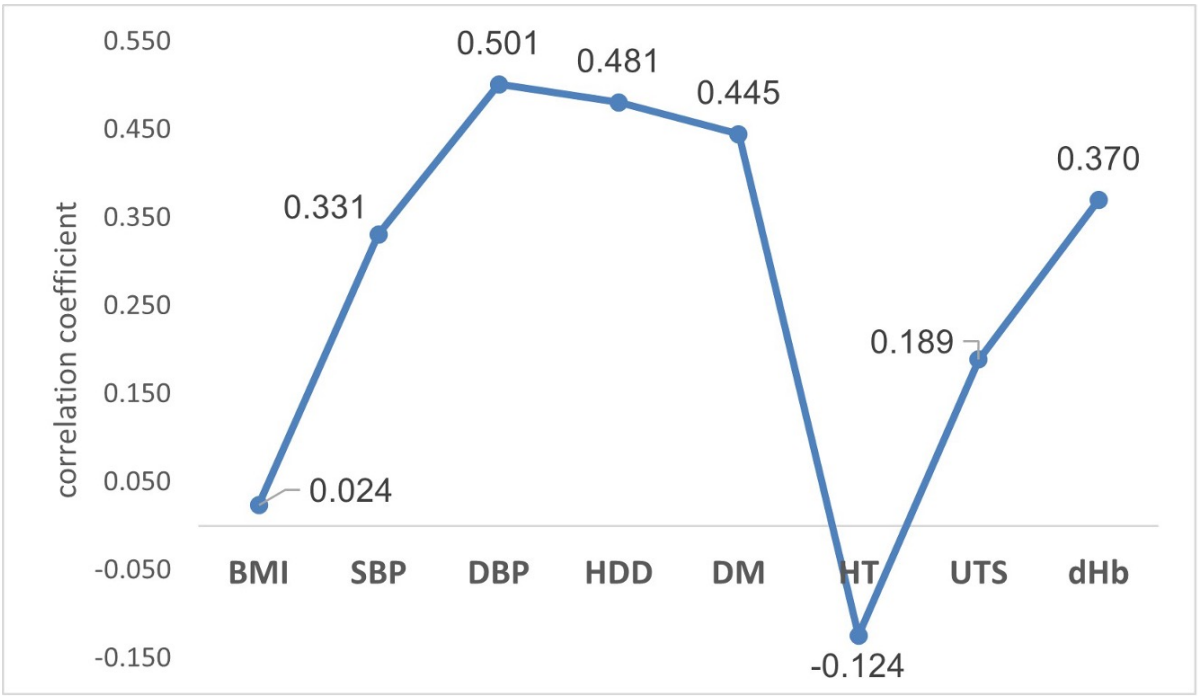

Figure 12. Correlation coefficient with ferritin:(b) cluster center after clustering.

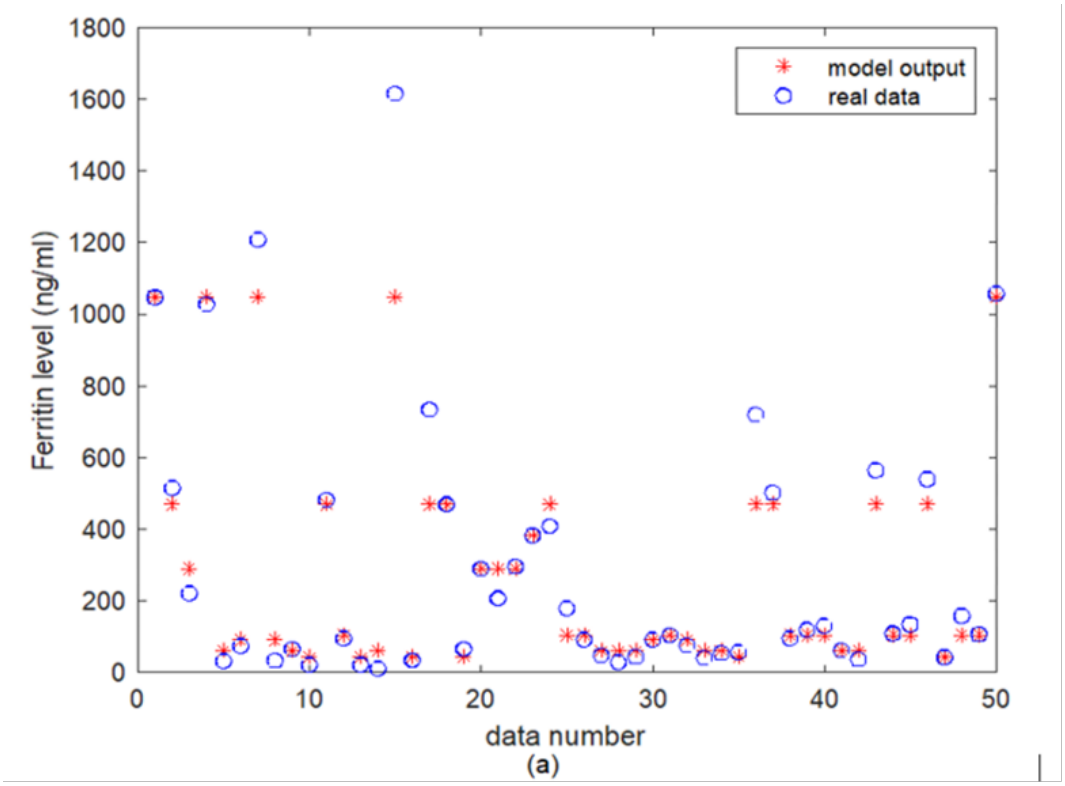

Figure 13. Model output vs real data: (a) data training;

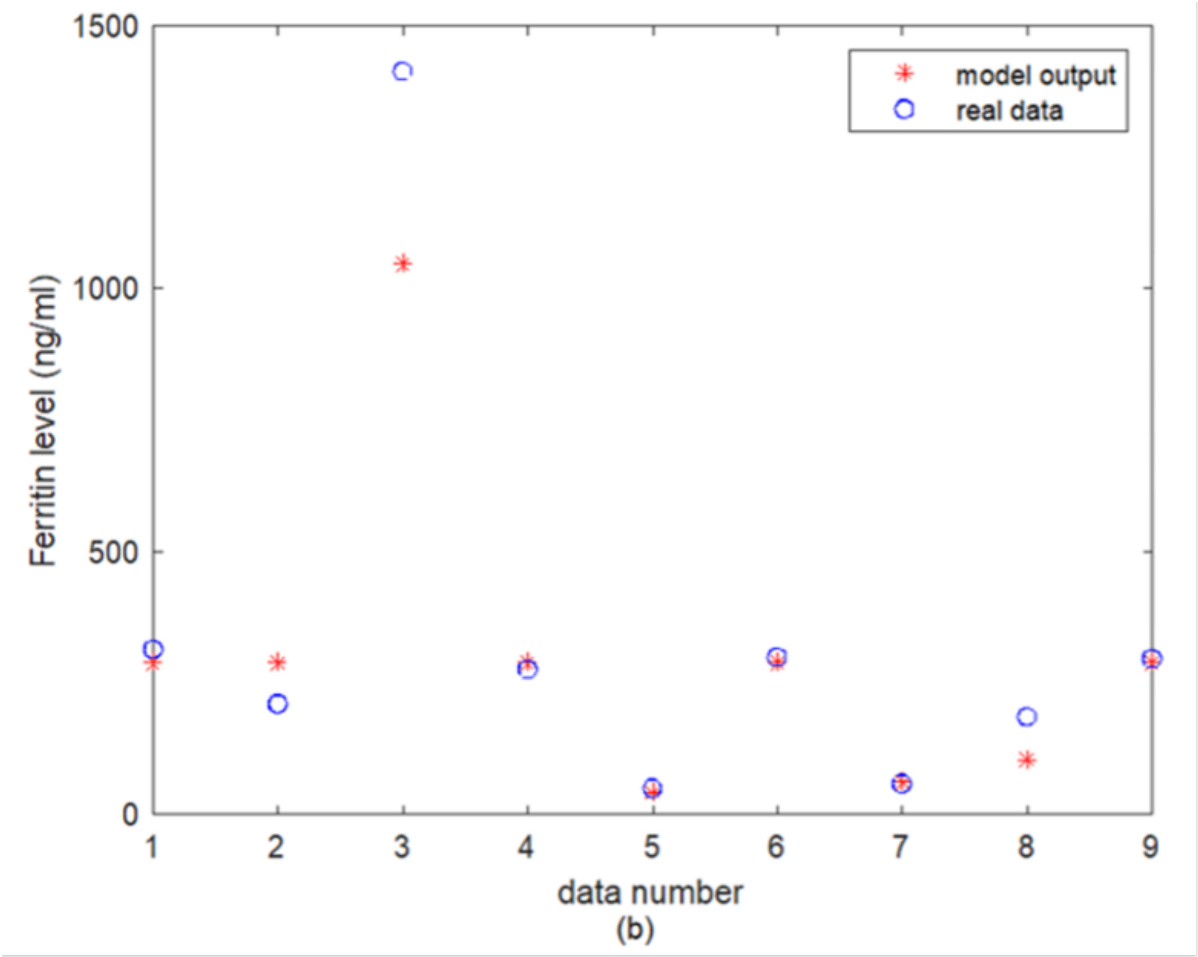

Figure 14. Model output vs real data: (b) data testing

TABLE III. The Results of Testing Ferritin Levels (Model Output VS Real Data)

Testing data number	**Feritin level (real data)**	**FSC**			**K-Means**		
		Cluster Number	Model Output	Accuracy/ Similarity	Cluster Number	Model Output	Accuracy/ Similarity
1	313.80	3	288.30	92%	6	135.95	43%
2	210.20	3	288.30	63%	6	135.95	65%
3	1,413.00	5	1,047.00	74%	7	1,208.00	85%
4	275.20	3	288.30	95%	6	135.95	49%
5	49.10	6	41.40	84%	1	449.72	11%
6	298.40	3	288.30	97%	6	135.95	46%
7	57.50	4	60.20	95%	6	135.95	42%
8	185.30	2	103.50	56%	1	449.72	41%
9	295.60	3	288.30	98%	3	224.18	76%
Average				**83.74%**		**50.91%**	

height, weight, blood pressure, examination of hemoglobin levels, and comorbidities. Once the estimated fertinine level is known, the clinician can make a decision about the next action to be taken so that the quality of life of CKD sufferers will be better maintained.

There are several advantages presented in the results of this study. First, the input data used in this model is easy to obtain on a daily basis, so effort and costs are not required. This condition is ideal for use in underdeveloped or developing countries [29], [3]. Second, the implementation of fuzzy clustering allows a data point to be a member of several clusters with different degrees of membership [30]. This is in line with the concept of grouping the conditions of CKD patients on hemodialysis. Third, the clustering process is carried out adaptively, where the number of clusters is not determined at the start but is obtained based on the learning process by finding the largest density value in the data sample. In this way, the cluster center that occurs is not new data but is one of the data samples that has the highest density. Fourth, by knowing the prediction of iron reserves in the form of ferritin levels quickly, clinicians can quickly make decisions on what therapy to take, especially when CKD patients have very low or very high ferritin levels.

There are still a number of things that require further research, especially related to the identification and analysis of outlier data. In this study the identification of outlier data was still carried out with the help of visualization. Such modern algorithms can be considered for further research [20], [31], [32], [33].

4. Conclusions and Future Work

Data can be efficiently clustered into eight clusters using FCS, which has an influence range of 0.5, an accept ratio of 0.75, and a reject ratio of 0.25. Every cluster has particular characteristics. Ferritin can be predicted by using the cluster center generated by FCS as a reference. The test results show an average level of similarity of 83.74%. This outcome is significantly higher than the 50.91% application of cluster centers that resulted from clustering the data using the K-Means method.

One of the implications of handling outliers in FSC is the formation of quite a large number of clusters. Because of this, optimization needs to be carried out so that, apart from being able to overcome outliers, the algorithm used also produces an optimal number of clusters. Further research needs to be conducted to enable the thorough identification and evaluation of outlier data.

References

[1] S. Kusumadewi, L. Rosita, and E. Wahyuni, "Selection of aggregation function in fuzzy inference system for metabolic syndrome," *Int J Adv Sci Eng Inf Technol*, vol. 12, no. 5, p. 2140, 2022.

[2] S. Kusumadewi, L. Rosita, and E. G. Wahyuni, "Development of a modified certainty factor model for prediction of metabolic syndrome," *International Journal of Innovative Computing, Information and Control*, vol. 18, no. 5, pp. 1463–1475, 2022.

[3] J. Sundström, J. Bodegard, A. Bollmann, M. G. Vervloet, P. B. Mark, A. Karasik, T. Taveira-Gomes, M. Botana, K. I. Birkeland, M. Thuresson *et al.*, "Prevalence, outcomes, and cost of chronic kidney disease in a contemporary population of 2· 4 million patients from 11 countries: The careme ckd study," *The Lancet Regional Health–Europe*, vol. 20, 2022.

[4] H. L. Chang, C. C. Wu, S. P. Lee, Y. K. Chen, W. Su, and S. L. Su, "A predictive model for progression of ckd," *Medicine*, vol. 98, 2019.

[5] S. Eddy, L. H. Mariani, and M. Kretzler, "Integrated multi-omics approaches to improve classification of chronic kidney disease," *Nature Reviews Nephrology*, vol. 16, no. 11, pp. 657–668, 2020.

[6] C. George, J. B. Echouffo-Tcheugui, B. G. Jaar, I. G. Okpechi, and A. P. Kengne, "The need for screening, early diagnosis, and prediction of chronic kidney disease in people with diabetes in low- and middle-income countries—a review of the current literature," *BMC medicine*, vol. 20, no. 1, p. 247, 2022.

[7] R. M. Hanna, E. Streja, and K. Kalantar-Zadeh, "Burden of anemia in chronic kidney disease: beyond erythropoietin," *Advances in therapy*, vol. 38, no. 1, pp. 52–75, 2021.

[8] N. Ueda and K. Takasawa, "Impact of inflammation on ferritin, hepcidin and the management of iron deficiency anemia in chronic kidney disease," *Nutrients*, vol. 10, no. 9, p. 1173, 2018.

[9] E. K. Batchelor, P. Kapitsinou, P. E. Pergola, C. P. Kovesdy, and D. I. Jalal, "Iron deficiency in chronic kidney disease: updates on pathophysiology, diagnosis, and treatment," *Journal of the American Society of Nephrology*, vol. 31, no. 3, pp. 456–468, 2020.

[10] A. Singh, D. Coyne, W. Shapiro, A. Rizkala, D. S. Group *et al.*, "Predictors of the response to treatment in anemic hemodialysis patients with high serum ferritin and low transferrin saturation," *Kidney international*, vol. 71, no. 11, pp. 1163–1171, 2007.

[11] A. Gafter-Gvili, A. Schechter, and B. Rozen-Zvi, "Iron deficiency anemia in chronic kidney disease," *Acta haematologica*, vol. 142, no. 1, pp. 44–50, 2019.

[12] T. Kuragano, N. Joki, H. Hase, K. Kitamura, T. Murata, S. Fujimoto, A. Fukatsu, T. Inoue, Y. Itakura, and T. Nakanishi, "Low transferrin saturation (tsat) and high ferritin levels are significant predictors for cerebrovascular and cardiovascular disease and death in maintenance hemodialysis patients," *PLoS One*, vol. 15, no. 9, p. e0236277, 2020.

[13] T. Ferguson, P. Ravani, M. M. Sood, A. Clarke, P. Komenda, C. Rigatto, and N. Tangri, "Development and external validation of a machine learning model for progression of ckd," *Kidney International Reports*, vol. 7, no. 8, pp. 1772–1781, 2022.

[14] J. Peng, K. Yang, H. Tian, Y. Lin, M. Hou, Y. Gao, X. Zhou, Z. Gao, and J. Ren, "The mechanisms of qizhu tangshen formula in the treatment of diabetic kidney disease: Network pharmacology, machine learning, molecular docking and experimental assessment," *Phytomedicine*, vol. 108, p. 154525, 2023.

[15] H. Ilyas, S. Ali, M. Ponum, O. Hasan, M. T. Mahmood, M. Iftikhar, and M. H. Malik, "Chronic kidney disease diagnosis using decision tree algorithms," *BMC nephrology*, vol. 22, no. 1, p. 273, 2021.

[16] S. Kusumadewi, L. Rosita, and E. G. Wahyuni, "Fuzzy linear regression based on a hybrid of fuzzy c-means and the fuzzy inference system for predicting serum iron levels in patients with

chronic kidney disease," *Expert Systems with Applications*, vol. 227, p. 120314, 2023.

[17] ——, "Implementation of fuzzy associative memory toward optimizing a neural network model to predict total iron binding capacity," *Biomedical Signal Processing and Control*, vol. 86, p. 105297, 2023.

[18] F. Ridzuan and W. M. N. W. Zainon, "Diagnostic analysis for outlier detection in big data analytics," *Procedia Computer Science*, vol. 197, pp. 685–692, 2022.

[19] L. Lyutikova, "Logical analysis of data for outliers detection," *Procedia Computer Science*, vol. 169, pp. 330–336, 2020.

[20] L. Lyutikova and E. Shmatova, "Using a logical derivative to analyze data outlier," *Procedia Computer Science*, vol. 169, pp. 304–309, 2020.

[21] M. Rubio-Rivas and X. Corbella, "Clinical phenotypes and prediction of chronicity in sarcoidosis using cluster analysis in a prospective cohort of 694 patients," *European journal of internal medicine*, vol. 77, pp. 59–65, 2020.

[22] K. Marks-Garber, T. Bdolah-Abram, and S. Nusair, "Cluster analysis based clinical profiling of idiopathic pulmonary fibrosis patients according to comorbidities evident prior to diagnosis: a single-center observational study," *European journal of internal medicine*, vol. 80, pp. 18–23, 2020.

[23] J. Bousquet, B. Sousa-Pinto, J. M. Antó, R. Amaral, L. Brussino, G. Canonica, A. Cruz, B. Gemicioglu, T. Haahtela, M. Kupczyk *et al.*, "Identification by cluster analysis of patients with asthma and nasal symptoms using the mask-air® mhealth app," *Pulmonology*, vol. 29, no. 4, pp. 292–305, 2023.

[24] N. G. Seresht, R. Lourenzutti, and A. R. Fayek, "A fuzzy clustering algorithm for developing predictive models in construction applications," *Applied Soft Computing*, vol. 96, p. 106679, 2020.

[25] S. Zeng, S.-M. Chen, and M. O. Teng, "Fuzzy forecasting based on linear combinations of independent variables, subtractive clustering algorithm and artificial bee colony algorithm," *Information Sciences*, vol. 484, pp. 350–366, 2019.

[26] U. M. Rao, Y. Sood, and R. Jarial, "Subtractive clustering fuzzy expert system for engineering applications," *Procedia Computer Science*, vol. 48, pp. 77–83, 2015.

[27] L.-H. Yang, F.-F. Ye, J. Liu, Y.-M. Wang, and H. Hu, "An improved fuzzy rule-based system using evidential reasoning and subtractive clustering for environmental investment prediction," *Fuzzy sets and systems*, vol. 421, pp. 44–61, 2021.

[28] S. F. Kane, C. Roberts, and R. Paulus, "Hereditary hemochromatosis: rapid evidence review," *American family physician*, vol. 104, no. 2, pp. 263–270, 2021.

[29] V. A. Luyckx, D. Z. Cherney, and A. K. Bello, "Preventing ckd in developed countries," *Kidney international reports*, vol. 5, no. 3, pp. 263–277, 2020.

[30] P. Nerurkar, A. Shirke, M. Chandane, and S. Bhirud, "Empirical analysis of data clustering algorithms," *Procedia Computer Science*, vol. 125, pp. 770–779, 2018.

[31] K. Domino, "Multivariate cumulants in outlier detection for financial data analysis," *Physica A: Statistical Mechanics and its Applications*, vol. 558, p. 124995, 2020.

[32] Z. Meng, Y. Xiong, W. Huang, L. Qin, X. Jin, and H. Yan, "Appscalpel: Combining static analysis and outlier detection to identify and prune undesirable usage of sensitive data in android applications," *Neurocomputing*, vol. 341, pp. 10–25, 2019.

[33] T. I. Souza, A. L. Aquino, and D. G. Gomes, "A method to detect data outliers from smart urban spaces via tensor analysis," *Future Generation Computer Systems*, vol. 92, pp. 290–301, 2019.

Linda Rosita She graduated from Gadjah Mada University in 2007 with a Specialist I degree in Clinical Pathology and in 2020 with a Specialist II degree in Hematology. She is currently working as lecturer in Faculty of Medicine, Universitas Islam Indonesia (UII), Yogyakarta, Indonesia. Her main research interests include clinical pathology and laboratory. She has 199 citations, with Google Scholar at H Index at 7. He is a member of the European Hematology Association since 2021. Apart from being an academic, he is also active as a doctor at the "JIH" hospital and the Yogyakarta PDHI hospital.

Sri Kusumadewi She received the Doctor (Dr) in Computer Science from Gadjah Mada University Yogyakarta, 2008, and the master degree in electrical engineering from Gadjah Mada University Yogyakarta, 1997. She is currently working as lecturer in Informatics Department, Universitas Islam Indonesia (UII), Yogyakarta, Indonesia. Her research interest lies in decision support systems, especially in the domain of medical informatics. With a Google Scholar H Index of 27, she has received over 11,713 citations. Many of her publications concentrated on clinical decision support systems, and various decision support system fields. She has also written dozens of books on the topic of fuzzy systems, medical informatics, and multi-attribute decision making.

Nyoman Kertia He is currently working as lecturer in Department of Internal Medicine, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia. His H Index is 10 on Google Scholar, and he has 369 citations.

Barkah Djaka Purwanto He is currently working as lecturer in Department of Internal Medicine, Faculty of Medicine Universitas Ahmad Dahlan, Bantul, Yogyakarta, Indonesia. His H Index is 2 on Google Scholar, and he has 19 citations.

Elyza Gustri Wahyuni She received the Master of Computer Science (M.Cs.) from Gadjah Mada University Yogyakarta. She is currently working as lecturer in Informatics Department, Universitas Islam Indonesia (UII), Yogyakarta, Indonesia. Her main research interests include medical informatics, decision support and intelligent systems. She has an H-Index of 8 on Google Scholar, and she has 314 citations.