Science and Technology Indonesia
e-ISSN:2580-4391 *p*-ISSN:2580-4405
Vol. 8, No. 1, January 2023

Research Paper

Check for updates

The Risk Cluster in Type 2 Diabetes Mellitus Based on Risk Parameters Using Fuzzy C-Means Algorithm

Marhamah[1], Sugiyarto Surono[2*], Endang Darmawan[1]

[1]*Department of Pharmacology and Clinical Pharmacy, Faculty of Pharmacy, Ahmad Dahlan University, Yogyakarta, 55164, Indonesia*
[2]*Department of Mathematics, Faculty of Applied Science and Technology, Ahmad Dahlan University, Yogyakarta, 55191, Indonesia*
**Corresponding author*: Sugiyarto@math.uad.ac.id

Abstract
The prevalence of type 2 diabetes mellitus increases every year. In the long term, type 2 diabetes mellitus can lead to complications of other diseases. This study aimed to analyze the risk cluster for type 2 diabetes mellitus based on risk parameters using the Fuzzy C-Means algorithm. The benefit of analyzing the risk cluster as an initial screening to prevent the occurrence of type 2 diabetes mellitus. This study used 905 subjects' data consisting of 562 males and 343 females. After the data preprocessing, the optimal number of clusters was determined using a Fuzzy C-Means algorithm process. Subsequently, the Pearson correlation test was conducted to determine the correlation between the risk parameters of type 2 diabetes mellitus and the cluster results. The study resulted in 2 risk clusters, subjects in cluster 1 were older than 60 years (34.1%), had a family history of type 2 diabetes mellitus (62.7%), had hypertension (55.4%), routinely took medicines (73.5%), undertook physical activity for less than half an hour (40.5%), and had a high blood pressure level (53.5%). The Pearson correlation test found that age, regular medication use, hypertension and blood pressure level all seem to have significant correlations with cluster outcomes. The risk cluster of type 2 diabetes mellitus was separated into two clusters using Fuzzy C-Means algorithm, namely the high-risk cluster and the low-risk cluster.

Keywords
Type 2 Diabetes Mellitus, Cluster Analysis, Fuzzy C-Means Algorithm, Pearson Correlation

Received: 10 August 2022, Accepted: 10 November 2022
https://doi.org/10.26554/sti.2023.8.1.17-24

1. INTRODUCTION

Diabetes mellitus is a chronic condition that occurs due to increase blood glucose levels, resulting from insufficient production of insulin hormones or insulin that does not work optimally due to damage to pancreatic beta cells (Sarría-Santamera et al., 2020). The risk factors are lack of physical activity, family history of diabetes mellitus, certain racial/ethnic groups, gestational diabetes history, hypertension history, high density lipoprotein (HDL) 150 mg/dL, smoking, prediabetes history, cardiovascular disease history, and age >45 years without the previously mentioned risk factors (Association, 2021). The symptoms include polyuria, polyphagia, polydipsia, and unexplained weight loss (Virani et al., 2021).

Type 2 diabetes mellitus has been identified as the world's leading cause of mortality (Pham et al., 2020). In 2019, the International Diabetes Federation estimated that 463 million adults aged 20-79 years had diabetes, and this number was predicted to reach 578.4 million and 700.2 million in 2030 and 2045 worldwide, respectively (Saeedi et al., 2019). Type 2 diabetes mellitus could cause complications such as heart attack, stroke, neuropathy, nephropathy, retinopathy, and kidney failure (Mezil, 2021).

Early screening for the risk of type 2 diabetes mellitus should be conducted to prevent the disease. Previous studies had predicted the risk of type 2 diabetes mellitus using classification techniques (Logistic Regression, K-Nearest Neighbor, Support Vector Machine, Decision Tree Classification, and Random Forest Classification) (Tigga and Garg, 2020). This study used a different approach-cluster analysis-to identify the type 2 diabetes mellitus risk group. The ability of cluster analysis to categorize substantial amounts of data and a variety of variables was one of its benefits (Muslimatin, 2011). Cluster analysis divides similar objects into a group and differs from others based on distance measures (Madhulatha, 2012). One of the methods in cluster analysis was the Fuzzy C-Means algorithm, a clustering method in which the level of data presented was determined by the degree of membership introduced, the membership value allows it to be between 0 and 1 (Ramya, 2018). Fuzzy C-Means clustering have been used to predict the level of diabetes mellitus, so that it can assist medical per-

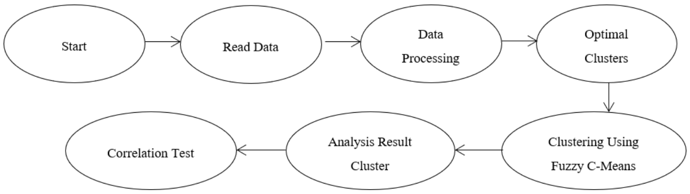

Figure 1. Flow Chart of Methods

sonnel in determining the right treatment therapy for patients (Jamuna, 2020). Kunwar et al. (2019) used Fuzzy C-Means method to assist clinical decision-making regarding kidney failure which showed there were 7 clusters. The existence of this grouping is an effort to prevent early kidney failure so that timely treatment can be given and reduce the risk of death. Subsequently, Sanakal and Jayakumari (2014) performed a diabetes mellitus prognosis by comparing Fuzzy C-Means algorithm with Support Vector Machine (SVM), showing the Fuzzy C-Means algorithm accuracy of 94.3%, but was 59.5% for SVM. Lomo et al. (2021) used Fuzzy C-Means clustering to identify patients with type 2 diabetes mellitus, based on demographic factors, blood glucose levels, the patient's surviving condition, and medication. It revealed that there were three clusters, each representing a patient's condition that helped individuals with type 2 diabetes mellitus live longer. Therefore, this study aimed to perform a risk cluster analysis of type 2 diabetes mellitus by identifying the characteristics of the risk parameters using the Fuzzy C-Means algorithm. Compared to earlier studies, more and different variables were used in this cluster analysis study to assess the risk of type 2 diabetes mellitus. Some variables were used to test the risk of type 2 diabetes mellitus, as used by the Association (2021).

2. EXPERIMENTAL SECTION

2.1 Materials

This study used secondary data obtained from kaggle datasets of diabetes owned by Tigga with 952 subjects aged >18 years old, female and male (Tigga and Garg, 2020). They were asked to answer a questionnaire containing questions about their health, lifestyle, and family background, as shown in Table 1 (Tigga and Garg, 2020). After the blank data were removed, 905 subjects were obtained, consisting of 562 males and 343 females.

2.2 Methods

The data were analyzed using the Python program with the followed stages (See Figure 1) :

1. Read data The data is read used pandas library in Python.
2. Data preprocessing

   a. Data Conversion from categorical to numerical
   Data in the categorical form should first be converted into numerical form. The categorical data contained in the parameters of age, gender, family history of diabetes, hypertension, physical activity, smoking, alcohol consumption, regularly taking medication, consumption of fast food, stress, blood pressure level, gestational diabetes history, and frequency of urination.

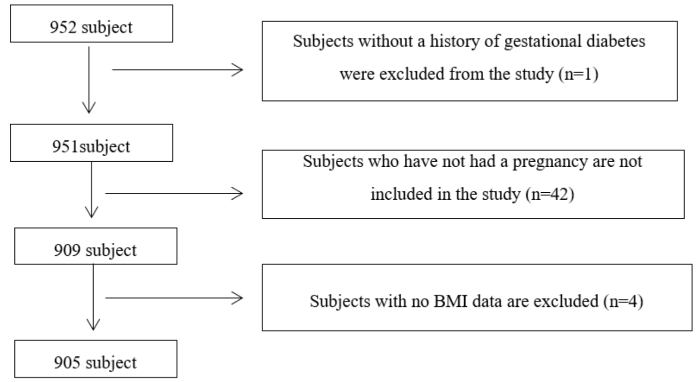

Figure 2. Flow Chart of Research Subject Selection

   b. Blank data Elimination
   Blank data were excluded from this study, yielding 905 final results (See Figure 2).

   c. Data normalization
   Data normalization was conducted using Z-score to make each parameter had a distribution in the same range. The data normalization formula is as follows (Gökhan et al., 2019):

$$Z = \frac{(X - \mu)}{\sigma} \quad (1)$$

   Where:
   x = observed value
   μ = average
   σ = standard deviation
   Z = Z-score (Raw Value)

   d. The determination of the optimal number of clusters
   The elbow method was used to analyze and determine the optimal dataset cluster. It began by plotting the values resulting from the function of the number of clusters and marking them at the elbow of the curve. Subsequently, the curve provided information about the number of clusters used. For example, when the value of the first and second gave the angle on the graph or the value that had decreased the most, the number can be used (Kurniawan et al., 2020).

3. Data clustering using the Fuzzy C-Means algorithm.
4. Analyzing the results of clusters to determine the characteristics or distribution of each parameter from each obtained cluster.
5. Doing a correlation test on the cluster results to measure the correlation between the parameters used in the cluster and the cluster results. The interpretation of Pearson correlation coefficient is shown in Table 2 (Hinkle et al., 2003).

Table 1. Risk Parameters of Type 2 Diabetes Mellitus of the Subjects (Tigga and Garg, 2020)

Parameters	Category
Age	< 40 years
	40-49 years
	50-59 years
	>60 years
Gender	Male
	Female
Family history	Yes/No
Hypertension	Yes/No
Physical activity	None
	Less than half an hour
	More than half an hour
	One hour or more
Smoking	Yes/No
Alcohol consumption	Yes/No
Routinely taking drugs	Yes/No
Fast food consumption	Occasionally
	Often
	Very Often
	Always
Stress	Not at all
	Sometimes
	Very often
	Always
Blood pressure level	Low/Normal/High
Gestational diabetes history	Yes/No
Urination frequency	Not much/Quite often
Body Mass Index (BMI)	Numeric (15-45 kg/m^2)
Sleep time	Numeric (4-11 hours)
Deep sleep	Numeric (0-11 hours)
Pregnancies	Numeric (0-4)

Table 2. The Interpretation of Pearson Correlation Coefficient (Hinkle et al., 2003)

Size of Correlation	Interpretation
.90 to 1.00 (-.90 to -1.00)	Very high positive (negative) correlation
.70 to .90 (-.70 to -.90)	High positive (negative) correlation
.50 to .70 (-.50 to -.70)	Moderate positive (negative) correlation
.30 to .50 (-.30 to -.50)	Low positive (negative) correlation
.00 to .30 (.00 to -.30)	Negligible correlation

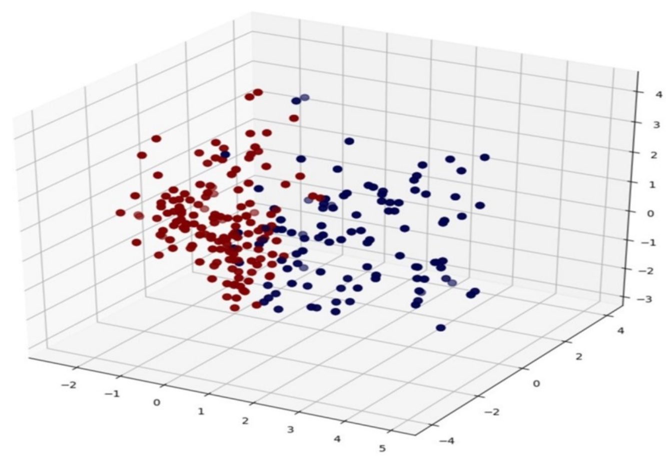

Figure 3. The Results of the Cluster of Subjects at Risk of Type 2 Diabetes Mellitus Using the Fuzzy C- Means Algorithm (Cluster 1 is Depicted in Blue and Cluster 2 is Depicted in Red)

3. RESULT AND DISCUSSION

3.1 Strand Geometry

The risk parameters of the subject in this study are described in Table 1. Figure 3 shows the data distribution plot of risk cluster results, cluster 1 is depicted in blue and cluster 2 is depicted in red. The optimal number of risk clusters of type 2 diabetes mellitus obtained was 2 clusters. Cluster 1 consisted of 370 subjects, while cluster 2 consisted of 535 subjects. The risk cluster results using the Fuzzy C-Means method can be seen in Table 3 and Table 4. Cluster 1 was dominated by subjects with 34.1% aged > 60 years, with a family history of type 2 diabetes mellitus (62.7%), who have hypertension (55.4%), who regularly take drugs (73.5%), who did physical activity for less than half an hour (40.5%), and who have a high blood pressure level (53.5%). In addition, Cluster 1 predominantly consisted of subjects with an average body mass index (BMI) of 27.84 kg/m^2 and the average sleep time was 6.75 hours.

Cluster 2 was dominated by subjects under the age of 40 (77.0%), who did not have a family history of type 2 diabetes mellitus (62.4%), had no history of hypertension (97.2%), did not frequently take medicines (90.5%), did more than half an hour of physical activity (31.0%), and had a normal blood pressure level category (92.7%). Furthermore, the average sleep time was 7.09 hours and the BMI was 23.9 kg/m2. The Pearson correlation test showed the highest correlation value with the cluster results on the age parameter (-0.67), regularly taking medication (-0.66), hypertension (-0.60), and blood pressure level (-0.59).

Our study showed that cluster 1 had a significant probability of acquiring diabetes mellitus, because many of the subjects in this cluster had diabetes mellitus risk parameters. The risk parameters were being aged >60 years, having a family history of diabetes, having hypertension, taking medication regularly, doing physical activity for less than half an hour, having high blood pressure levels, having an average BMI of 27.84 kg/m^2, and sleeping less than 6.75 hours.

In this study, subjects aged more than 60 years old were in the cluster with a high-risk of developing diabetes mellitus based on Fuzzy C-Means clustering. These results are consistent with previous studies which reported that the prevalence of diabetes mellitus increased at 55- 64 years of age and decreased after passing this range. Therefore, aged people were at high risk of developing diabetes mellitus (Association, 2021). Asiimwe et al. (2020) showed that the age group of 61–65 years was strongly influenced by diabetes mellitus. The meta-analysis results also showed an increase in the likelihood of diabetes mellitus at age 40 years compared with those aged <40 years (Asamoah-Boaheng et al., 2019). Aging can increase the occurrence of chronic inflammation. In addition, there is an increase in the concentration of free fatty acids in the blood due to impaired lipid metabolism in the elderly. This can lead to insulin resistance (Ismail et al., 2021).

In several studies, a family history of diabetes mellitus has been linked to an increased risk of developing type 2 diabetes. A cross sectional study conducted by Zhang et al. (2015) showed that there was a significant increase in the prevalence of diabetes mellitus in study participants who had two generations of first-degree relatives with diabetes with a history of diabetes by 32.7%, then only one generation of first-degree relatives with a history of diabetes by 20.1% and no first-degree relatives with diabetes at 8.4%. The first-degree relatives of diabetes mainly had β-cell dysfunction, and that the higher the family history risk category, the more severe the β-cell dysfunction. Gopalakrishnan and Geetha (2017) reported that almost 68.8% of patients with diabetes mellitus had a family history of diabetes in which mothers with diabetes had a greater influence than fathers with diabetes. The meta-analysis conducted by Asamoah-Boaheng et al. (2019) showed that a family history of diabetes mellitus had a significant relationship with the occurrence of diabetes mellitus in a person.

Based on hypertension parameters, cluster 1 was dominated by subjects with a history of hypertension. Study showed that the prevalence of hypertension in patients with type 2 diabetes mellitus is 59.5% higher in the 50-60-years age group (Akalu and Belsti, 2020). Wei et al. (2011) reported a significant relationship between the incidence of type 2 diabetes mellitus and hypertension compared to normal blood pressure in white. The study by Kim et al. (2015) showed that diabetes mellitus was higher in initial subjects with pre-hypertension and hypertension than those with normal blood pressure and subjects who had normal blood pressure at the beginning of the examination later. After eight years, prehypertension or hypertension was at a significantly higher risk of developing diabetes mellitus in subjects compared to those with controlled blood pressure.

Wang et al. (2018) reported that moderate levels of physical activity (30 minutes/3 days) had a significant relationship with the occurrence of type 2 diabetes mellitus. The study by Sim-

Table 3. The Results of the Cluster of Subjects at Risk of Type 2 Diabetes Mellitus Using the Fuzzy C- Means Algorithm Based on Categorical Parameters

Parameters	Cluster Results Cluster 1 n = 370	%	Cluster 2 n = 535	%	Total n = 952	Pearson correlation (r)
Age						
<40 years	50	13.5	412	77.0	462	-0.67
40-49 years	70	18.9	83	15.5	153	
50-59 years	124	33.5	24	4.5	148	
>60 years	126	34.1	16	3.0	142	
Gender						
Male	207	55.9	355	66.4	562	0.11
Female	163	44.1	180	33.6	343	
Family history						
Yes	232	62.7	201	37.6	433	0.25
No	138	37	334	62.4	472	
Hypertension						
Yes	205	55.4	15	2.8	220	-0.60
No	165	44.6	520	97.2	685	
Physical activity						
None	67	18.1	62	11.6	129	0.15
Less than half an hour	150	40.5	167	31.2	317	
More than half an hour	85	23.0	166	31.0	251	
One hour or more	65	18.4	140	26.2	208	
Smoking						
Yes	40	10.8	66	12.3	106	0.023
No	330	89.2	469	87.7	799	
Alcohol consumption						
Yes	99	26.8	88	16.4	187	-0.13
No	271	73.2	447	83.6	718	
Routinely taking drugs						
Yes	272	73.5	51	9.5	323	-0.66
No	98	26.5	484	90.5	582	
Fast food consumption						
Occasionally	290	78.4	343	64.1	633	0.1
Often	44	11.9	132	24.7	176	
Very Often	20	5.4	32	6.0	52	
Always	16	4.3	28	5.2	44	
Stress						
Not at all	39	10.5	92	17.2	131	-0.34
Sometimes	168	45.4	362	67.7	530	
Very often	20	23.0	73	13.6	158	
Always	78	21.1	8	1.5	86	
Blood pressure level						
Low	0	0	27	5.0	27	-0.59
Normal	172	46.5	496	92.7	668	
High	198	53.5	12	2.2	210	
Gestational diabetes history						
Yes	10	2.7	4	0.7	14	-0.078
No	360	97.3	531	99.3	891	
Urination frequency						
Not much	208	56.2	434	81.1	642	-0.27
Quite often	162	43.8	101	18.9	263	

Table 4. The Results of the Cluster of Subjects at Risk of Type 2 Diabetes Mellitus Using the Fuzzy C- Means Algorithm Based on Numeric Parameters

Parameters	Cluster 1 (n=370) (mean)	Cluster 2 (n=535) (mean)	Total subject (n=905)	Pearson correlation (r)
BMI (kg/m^2)	27.8	23.9	905	-0.37
Sleep time (hours)	6.75	7.09	905	0.13
Deep sleep (hours)	5.05	5.88	905	0.22
Pregnancies	0.60	0.23	905	-0.19

bolon et al. (2020) showed that individuals with less physical activity were more likely to suffer from type 2 diabetes mellitus than very active individuals. Suppose a person's physical activity was lacking; an imbalance in the amount of energy in the body occurred since the energy consumed was greater than the amount expended. The energy that entered the body and went unused was stored mainly as adipose tissue, triggering insulin resistance, causing type 2 diabetes mellitus. Furthermore, lack of physical activity coupled with the consumption of carbohydrates, proteins, and fats were obesity factors that resulted in an increase in free fatty acids. Therefore, it reduced the translocation of glucose transporters to the membrane plasma and causes insulin resistance. So, subjects with less physical activity were more at high-risk of developing type 2 diabetes mellitus.

In this study, subjects taking medicine regularly were categorized as having a higher risk of developing diabetes mellitus. Diabetes mellitus was sometimes associated with comsumption some drugs. Some drugs could increase the occurrence of type 2 diabetes mellitus associated with reduced insulin production, reduced insulin sensitivity, or the occurrence of both conditions. One example of a drug-associated with the risk of diabetes mellitus was glucocorticoids, with a mechanism that could reduce insulin production and sensitivity (Repaske, 2016). Based on the results of a study involving subjects with autoimmune diseases, the overall risk of developing diabetes after one year was 0.9% when glucocorticoids were not received. In contrast, the risk increased to 2.1% for prednisolone <5 mg (daily dose) when glucocorticoids were administered and 5.0% for prednisone 25 mg (daily dose) (Wu et al., 2020). Furthermore, the thiazide drug class could also increase the risk by reducing insulin production (Repaske, 2016). There was an increase in fasting blood glucose in hypertensive patients given thiazide diuretics compared to those that did not receive thiazide diuretics. However, with low doses of thiazide diuretics (25 mg/day) given thiazide diuretics (hydrochlorothiazide or chlorthalidone), there were smaller changes in fasting blood glucose compared with higher doses in hypertensive patients Zhang and Zhao (2016) Another study showed that taking antibiotics within 90 days had a higher risk of developing diabetes mellitus than those who did not use them. The use of five or more classes of antibiotics was also at higher risk than those using only one class of antibiotics (Park et al., 2021). Patients are expected to discontinue the medication when diabetes mellitus is associated with the consumption of drugs (Repaske, 2016).

The subjects with a high blood pressure level in this study were included in the high-risk of diabetes mellitus. The results of a prospective study showed that the systolic blood pressure level in the standard, prehypertension, mild hypertension and moderate/severe hypertension groups had a risk of 19%, 30%, 31%, and 49% of having diabetes mellitus during the follow-up period without being influenced by risk factors such as BMI and others (Stahl et al., 2012).

The higher average BMI indicates having a higher risk of developing diabetes mellitus. Pinidiyapathirage et al. (2013) reported that a high BMI is associated with an increase in fasting plasma glucose levels. The study by Liyanage (2018) showed a significant relationship between BMI and diabetes mellitus. The majority (51%) of study participants who were overweight experienced diabetes mellitus, followed by 29.40% of normal weight and 7.80% of underweight. According to Tang et al. (2021) the risk ratio for diabetes mellitus was 2.13 times for BMI 22.5- <25.0 kg/m^2, 2.14 times for BMI 25.0- <27.5 kg/m^2, 3.17 times for BMI 27.5- <30.0 kg/m^2, and 3.14 times for BMI ≥30,0 kg/m^2.

Cluster 1 has a higher risk of developing type 2 diabetes mellitus on sleep time and deep sleep parameters. The meta-analysis results showed that the risk of type 2 diabetes was low at 7-8 hours of sleep per day. Sleep durations shorter and longer than usual were associated with a significantly increased risk of type 2 diabetes mellitus. A decrease in one hour was associated with 9%, and an increase of one hour was associated with a 14% increase in the risk of type 2 diabetes mellitus (Shan et al., 2015). Furthermore, lack of sleep also affect leptin and ghrelin hormones. The hormone leptin helps provide a feeling of fullness, while ghrelin helps increase appetite. Insufficient sleep duration decreases leptin hormone and increases ghrelin. Furthermore, lack of sleep can reduce peptide tyrosine-tyrosine and glucagon like peptide-1 hormones that help suppress hunger; hence, these changes increase appetite, cause obesity, and interfere with blood glucose control (Sakamoto et al., 2018).

We found that the diabetes mellitus risk parameters correlated with the obtained risk cluster were age, regularly taking medication, hypertension, and blood pressure. The obtained Pearson correlation coefficients between the risk parameters and the risk clusters are categorized as moderate correlation (Hinkle et al., 2003).

According to Steyn et al. (2004) the risk parameters of type 2 diabetes mellitus are divided into 2, namely modifiable and non-modifiable risk parameters. We can start prevention as early as possible to reduce the modifiable risk parameters of the occurrence or development of type 2 diabetes mellitus, such as changing lifestyle (Uusitupa et al., 2019). These efforts include regular physical activity such as brisk walking for 30 minutes, which when done five times a week will help increase insulin sensitivity in the body (Association, 2021; Steyn et al., 2004). Then maintain an ideal body weight by managing meal portions and eating foods that have balanced nutrition, namely fruits, vegetables, and whole grains that can make you full longer, accompanied by physical activity, because obesity can increase the risk of type 2 diabetes mellitus (Uusitupa et al., 2019). Another effort is to lead a healthy lifestyle by not smoking and not consuming alcohol because both will reduce insulin sensitivity. Too little or too much sleep is also not good because it can increase the risk of type 2 diabetes mellitus, so sufficient sleep duration is needed, which is 7-8 hours per day and is also beneficial in improving concentration, memory, and mood. In addition, as an effort to prevent and detect the risk of diabetes mellitus, it can also be done by controlling blood glucose, blood pressure, and lipid profile at the nearest health care facility (Association, 2021).

This study had some limitations. First, the data used was secondary data obtained from Kaggle. As a result, some subjects' data were incomplete, and they were eventually dropped from the study. Second, we could not confirm more detailed information from the subject, for example, the risk parameter, which stated that the subject regularly consumes drugs. We did not know the type and amount of drugs consumed by the subjects.

4. CONCLUSION

The Fuzzy C-Means clustering algorithm was used to cluster the results of 905 subjects, yielding clusters 1 (high risk cluster) and 2 (low risk cluster). Cluster 1 was more likely to acquire diabetes mellitus due to the presence of more prevalent diabetes risk parameters. Subjects who have risk parameters that are included in the high-risk cluster need to be aware and alert to prevent the occurrence of type 2 diabetes mellitus. Further research is needed to determine whether changing the risk parameters of type 2 diabetes mellitus through lifestyle changes can reduce the risk of developing type 2 diabetes mellitus.

5. ACKNOWLEDGMENT

The authors are grateful to the Faculty of Pharmacy and computational math lab department of mathematics FAST UAD for their support.

REFERENCES

Akalu, Y. and Y. Belsti (2020). Hypertension and its Associated Factors Among Type 2 Diabetes Mellitus Patients at Debre Tabor General Hospital, Northwest Ethiopia. *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*, **13**; 1621

Asamoah-Boaheng, M., O. Sarfo-Kantanka, A. B. Tuffour, B. Eghan, and J. C. Mbanya (2019). Prevalence and Risk Factors for Diabetes Mellitus Among Adults in Ghana: A Systematic Review and Meta-analysis. *International Health*, **11**(2); 83–92

Asiimwe, D., G. O. Mauti, and R. Kiconco (2020). Prevalence and Risk Factors Associated With Type 2 Diabetes in Elderly Patients Aged 45-80 Years at Kanungu District. *Journal of Diabetes Research*, **2020**; 1–5

Association, A. D. (2021). Improving Care and Promoting Health in Populations: Standards of Medical Care in Diabetes-2021. *Diabetes Care*, **44**(1); S7–S14

Gökhan, A., C. O. Güzeller, and M. T. Eser (2019). The Effect of The Normalization Method Used in Different Sample Sizes on the Success Of Artificial Neural Network Model. *International Journal of Assessment Tools in Education*, **6**(2); 170–192

Gopalakrishnan, S. and A. Geetha (2017). Study on the Impact of Family History of Diabetes Among Type 2 Diabetes Mellitus Patients in an Urban Area of Kancheepuram District, Tamil Nadu. *International Journal Community Medicine and Public Health*, **4**; 4151–4156

Hinkle, D. E., W. Wiersma, and S. G. Jurs (2003). *Applied Statistics for the Behavioral Sciences*. Houghton Mifflin College Division

Ismail, L., H. Materwala, and J. Al Kaabi (2021). Association of Risk Factors With Type 2 Diabetes: A Systematic Review. *Computational and Structural Biotechnology Journal*, **19**; 1759–1785

Jamuna, . K. K. M., S. (2020). Prediction of Diabetes and Clustering Based on its Levels Using Fuzzy C Means Algorithm. *International Journal of Scientific and Technology*, **9**(2); 3222–3225

Kim, M. J., N. K. Lim, S. J. Choi, and H. Y. Park (2015). Hypertension is an Independent Risk Factor For Type 2 Diabetes: the Korean Genome and Epidemiology Study. *Hypertension Research*, **38**(11); 783–789

Kunwar, V., A. S. Sabitha, T. Choudhury, and A. Aggarwal (2019). Chronic Kidney Disease Using Fuzzy C-means Clustering Analysis. *International Journal of Business Analytics (IJBAN)*, **6**(3); 43–64

Kurniawan, M., R. R. Muhima, and S. Agustini (2020). Comparison of Clustering K-means, Fuzzy C-means, and Linkage For Nasa Active Fire Dataset. *International Journal of Artificial Intelligence and Robotics (IJAIR)*, **2**(2); 34

Liyanage, L. (2018). Diabetes Mellitus and its Risk Factors. *EPRA International Journal of Multidisciplinary Research (IJMR)*, **4**(9); 114–117

Lomo, S. I., E. Darmawan, and S. Sugiyarto (2021). Cluster analysis of Type Ii Diabetes Mellitus Patients with the Fuzzy C-means Method. *Annals of Mathematical Modeling*, **13**(1); 32–39

Madhulatha, T. S. (2012). An Overview on Clustering Methods. *IOSR Journal of Engineering*, **2**(4); 719–725

Mezil, . A. B. A., S. A. (2021). Complication of Diabetes Mellitus. *Annals of the Romanian Society for Cell Biology*, **25**(3); 1546–156

Muslimatin, B. (2011). *Perbandingan Metode K-means dan Metode Fuzzy C-means untuk Clustering Data: Studi Kasus Pada Harga Saham Harian Pada Pt. Astra, Tbk*. Ph.D. thesis, Universitas Islam Negeri Maulana Malik Ibrahim

Park, S. J., Y. J. Park, J. Chang, S. Choi, G. Lee, J. S. Son, K. H. Kim, Y. H. Oh, and S. M. Park (2021). Association Between Antibiotics Use and Diabetes Incidence in A Nationally Representative Retrospective Cohort Among Koreans. *Scientific Reports*, **11**(1); 1–10

Pham, T. B., T. T. Nguyen, H. T. Truong, C. H. Trinh, H. N. T. Du, T. T. Ngo, and L. H. Nguyen (2020). Effects of Diabetic Complications on Health-related Quality of Life Impairment in Vietnamese Patients with Type 2 Diabetes. *Journal of Diabetes Research*, **2020**; 1–8

Pinidiyapathirage, M., A. Kasturiratne, U. Ranawaka, D. Gunasekara, N. Wijekoon, K. Medagoda, S. Perera, F. Takeuchi, N. Kato, and T. Warnakulasuriya (2013). The Burden of Diabetes Mellitus and Impaired Fasting Glucose in an Urban Population of Sri Lanka. *Diabetic Medicine*, **30**(3); 326–332

Ramya, T. B. (2018). Disease Prediction System Using Fuzzy C-means Algorithm. *International Journal of Engineering Research & Technology (IJERT)*, **6**(3); 1–4

Repaske, D. R. (2016). Medication Induced Diabetes Mellitus. *Pediatric Diabetes*, **17**(6); 392–397

Saeedi, P., I. Petersohn, P. Salpea, B. Malanda, S. Karuranga, N. Unwin, S. Colagiuri, L. Guariguata, A. A. Motala, and K. Ogurtsova (2019). Global and Regional Diabetes Prevalence Estimates For 2019 and Projections For 2030 and 2045: Results From The International Diabetes Federation Diabetes Atlas. *Diabetes Research and Clinical Practice*, **157**; 107843

Sakamoto, R., T. Yamakawa, K. Takahashi, J. Suzuki, M. M. Shinoda, K. Sakamaki, H. Danno, H. Tsuchiya, M. Waseda, and T. Takano (2018). Association of Usual Sleep Quality and Glycemic Control in Type 2 Diabetes in Japanese: A Cross Sectional Study. Sleep and Food Registry in Kanagawa (SOREKA). *PloS one*, **13**(1); 0191771

Sanakal, R. and T. Jayakumari (2014). Prognosis of Diabetes Using Data Mining Approach-fuzzy C Means Clustering and Support Vector Machine. *International Journal of Computer Trends and Technology*, **11**(2); 94–98

Sarría-Santamera, A., B. Orazumbekova, T. Maulenkul, A. Gaipov, and K. Atageldiyeva (2020). The Identification of Diabetes Mellitus Subtypes Applying Cluster Analysis Techniques: A Systematic Review. *International Journal of Environmental Research and Public Health*, **17**(24); 9523

Shan, Z., H. Ma, M. Xie, P. Yan, Y. Guo, W. Bao, Y. Rong, C. L. Jackson, F. B. Hu, and L. Liu (2015). Sleep Duration and Risk Of Type 2 Diabetes: A Meta-analysis of Prospective Studies. *Diabetes Care*, **38**(3); 529–537

Simbolon, D., A. Siregar, and R. A. Talib (2020). Physiological Factors and Physical Activity Contribute to The Incidence of Type 2 Diabetes Mellitus in Indonesia. *Kesmas: Jurnal Kesehatan Masyarakat Nasional (National Public Health Journal)*, **15**(3); 120–127 (In Indonesia)

Stahl, C. H., M. Novak, G. Lappas, L. Wilhelmsen, L. Björck, P.-O. Hansson, and A. Rosengren (2012). High-normal Blood Pressure and Long-term Risk Of Type 2 Diabetes: 35-year Prospective Population Based Cohort Study of Men. *BMC Cardiovascular Disorders*, **12**(1); 1–8

Steyn, N. P., J. Mann, P. Bennett, N. Temple, P. Zimmet, J. Tuomilehto, J. Lindström, and A. Louheranta (2004). Diet, Nutrition and The Prevention of Type 2 Diabetes. *Public Health Nutrition*, **7**(1a); 147–165

Tang, M., Y. Zhou, A. Song, J. Wang, Y. Wan, and R. Xu (2021). The Relationship Between Body Mass Index and Incident Diabetes Mellitus in Chinese Aged Population: A Cohort Study. *Journal of Diabetes Research*, **2021**; 1–8

Tigga, N. P. and S. Garg (2020). Prediction of Type 2 Diabetes Using Machine Learning Classification Methods. *Procedia Computer Science*, **167**; 706–716

Uusitupa, M., T. A. Khan, E. Viguiliouk, H. Kahleova, A. A. Rivellese, K. Hermansen, A. Pfeiffer, A. Thanopoulou, J. Salas-Salvadó, and U. Schwab (2019). Prevention of Type 2 Diabetes by Lifestyle Changes: A Systematic Review and Meta-analysis. *Nutrients*, **11**(11); 2611

Virani, S. S., A. Alonso, H. J. Aparicio, E. J. Benjamin, M. S. Bittencourt, C. W. Callaway, A. P. Carson, A. M. Chamberlain, S. Cheng, and F. N. Delling (2021). Heart Disease and Stroke Statistics—2021 Update: A Report From the American Heart Association. *Circulation*, **143**(8); e254–e743

Wang, Q., X. Zhang, L. Fang, Q. Guan, L. Gao, and Q. Li (2018). Physical Activity Patterns and Risk of Type 2 Diabetes and Metabolic Syndrome In Middle-aged and Elderly Northern Chinese Adults. *Journal of Diabetes Research*, **2018**; 1–8

Wei, G. S., S. A. Coady, D. C. Goff Jr, F. L. Brancati, D. Levy, E. Selvin, R. S. Vasan, and C. S. Fox (2011). Blood Pressure and the Risk Of Developing Diabetes in African Americans and Whites: Aric, Cardia, and the Framingham Heart Study. *Diabetes Care*, **34**(4); 873–879

Wu, J., S. L. Mackie, and M. Pujades-Rodriguez (2020). Glucocorticoid Dose Dependent Risk of Type 2 Diabetes in Six Immune-mediated Inflammatory Diseases: A Population-based Cohort Analysis. *BMJ Open Diabetes Research and Care*, **8**(1); 001220

Zhang, J., Z. Yang, J. Xiao, X. Xing, J. Lu, J. Weng, W. Jia, L. Ji, Z. Shan, and J. Liu (2015). Association Between Family History Risk Categories and Prevalence of Diabetes in Chinese Population. *PLoS One*, **10**(2); e0117044

Zhang, X. and Q. Zhao (2016). Association of Thiazide-Type Diuretics With Glycemic Changes in Hypertensive Patients: A Systematic Review and Meta-Analysis of Randomized Controlled Clinical Trials. *The Journal of Clinical Hypertension*, **18**(4); 342–351