

PAPER • OPEN ACCESS

## Kernel Based Fuzzy C-Means Clustering for Chronic Sinusitis Classification

To cite this article: Rezki Aulia Putri *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **546** 052060

View the [article online](#) for updates and enhancements.



**IOP | ebooks™**

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the **collection** - download the first chapter of every title for free.

# Kernel Based Fuzzy C-Means Clustering for Chronic Sinusitis Classification

**Rezki Aulia Putri<sup>1</sup>, Zuherman Rustam<sup>\*1</sup>, Jacob Pandelaki<sup>3</sup>**

<sup>1</sup>Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Indonesia, Kampus UI Depok, Depok 16424, Indonesia

<sup>3</sup>Department of Radiology, Cipto Mangunkusumo National General Hospital, DKI Jakarta 10430, Indonesia

\*Corresponding author: rustam@ui.ac.id

**Abstract.** Sinusitis is an inflammation of the sinus wall, a small cavity interconnected through the airways in the skull bones. It is located on the back of the forehead, inside the cheek bone structure, on both sides of the nose, and behind the eyes. Chronic sinusitis is caused by infection, growth of nasal polyps, or irregularities of the nasal septum. This condition can affect teenagers, adults, and even children. To classify sinusitis we use Kernel Based Fuzzy C-Means (FCM) Clustering Algorithm, which is the development of Fuzzy C-Means (FCM) Algorithm. FCM is one of the widely used clustering technique. FCM algorithm comprises of sample points used to make whole and sub vector spaces according to the size of the distance. However, when non-linear data is separated, the convergence is inaccurate and slow. To overcome this problem, a Kernel-Based Fuzzy C-Means algorithm that makes use of kernel functions as a substitute for Euclidean distance utilized. It maps out samples to high-dimensional space to increase the differences between cluster centres, so they can overcome FCM deficiencies and improve linear machine capabilities. Data was obtained from the laboratory of Radiology at Cipto Mangunkusumo National General Hospital, Indonesia, with a 100% accuracy.

## 1. Introduction

Sinusitis is the inflammation of the paranasal sinuses lining [1, 2, 3]. When it occurs with rhinitis, it is known as rhinosinusitis [2, 4]. Sinus is a small cavity that is connected through the airways inside the skull. Sinus is located in the back bone of the forehead, the inner structure of the cheekbones, both sides of the nose and behind the eyes [5]. There are some classifications of sinusitis, they are; acute sinusitis, subacute sinusitis, chronic sinusitis, and recurrent sinusitis [6]. Chronic sinusitis occurs if there is an inflammation of the paranasal sinuses for more than 12 weeks [2]. It can be caused by infection, sinus growth (nasal polyps) or nasal septal deviation [2, 3, 6]. This condition is common in teenagers and adults, but can also affect children [6].

Chronic sinusitis often occurs together with inflammation of the nasal airways. When it is preceded by rhinitis symptoms, the condition is known as chronic rhinosinusitis (CRS) [2]. CRS can manifest with and without nasal polyps, and allergic fungal rhinosinusitis [2]. Most often it occurs because of unresolved acute sinusitis [2].

Factors that can increase the risk of sinusitis in adults are fungal and dental infections, nose injury, enlarged glands adenoids, smoking (smokers), inhaling cigarette smoke (passive smoking), diving, swimming, and presence of strange objects in the nose [6].



When a person experiences at least 2 out of any of these 4 signs and symptoms, they might be prone to chronic sinusitis. These symptoms include inflammation of the nose, swollen around the eyes, cheeks, nose or forehead, thick colour liquid coming out of the nose or the fluid flowing from the back of the throat (postnasal drainage), lack of the sense of smell and taste in adults or coughing in children, and nasal obstruction which causes difficulty in breathing through the nose [5].

Chronic sinusitis is usually as a result of allergies or exposure to environmental pollutants [1]. To prevent recurrence, diligently wash your hands, drink more water, get an annual flu vaccine, avoid stress, consume nutritious foods, avoid allergens in the environment, and keep up oral health [5].

Some common treatments for chronic sinusitis are as follows [2]:

**Antimicrobial treatment.** The use of antibiotics depends on the type of sinusitis that the patient has. In using antibiotics, several things such as effectiveness, costs, and side effects should be considered [7]. Antibiotics are used for sinusitis caused by bacteria [6, 8]. Anti-microbial treatments are usually carried out for up to one and a half months or more and should not be stopped until the patient shows no further symptoms. This is because inadequate treatment will increase the possibility of relapse. Chronic sinusitis is more difficult to treat, because it is usually associated with nasal polyps, asthma, and respiratory diseases [9]. However, daily saline irrigation with topical corticosteroid therapy is considered the main therapy for its treatment [2].

**Symptomatic treatment.** Symptomatic treatment is performed when the symptoms of chronic sinusitis occur. Therapies such as humidification / vapor, maintain nutritional balance, abstinence from smoking, warm compresses, and fulfilment of body fluid requirements [1, 6].

**Surgical treatment.** Surgical treatment is used in addition to the use of medicines in cases. It is generally done for cases that are difficult to treat and have performed anatomical examinations. The purposes are to regulate mucosal resistance and restore sinus ventilation with the goal of rebuilding the mucociliary framework. Surgical efforts to rebuild the mucous cover are aggravated. The foundation of utilitarian endoscopic sinus surgery (FESS) simplifies the elimination of disease by rebuilding aeration and adequate drainage of the sinuses by establishing patent ostiomeatal complexes, eliminating severe polypoidosis, and not jeopardizing normal nose function [2].

There are several types of tests that the doctors may carry out to determine if a patient is diagnosed with sinusitis. These include [6]:

**Allergy test.** If you suspect that sinusitis is caused by a certain allergen, your doctor will carry out an allergic test through the skin. This type of test is safe and fast, and can help to decide the type of allergen [6].

**Nasal endoscopy.** Check by using an endoscope to see the inside of the sinus [6].

**Imaging test.** CT scans or MRI is used to get a detailed description of the area of the sinuses around the nose, such as inflammatory conditions or blockages that are difficult to detect with endoscopy [6].

In this experiment, we used data that are obtained from the Laboratory of Radiology at Cipto Mangunkusumo National General Hospital, Indonesia. To classify sinusitis in patients we made use of Kernel Based Fuzzy C-Means (FCM) Clustering Algorithm, which is the development of Fuzzy C-Means (FCM) Algorithm. FCM algorithm has the sample points that can be used to make vector space, which can further be divided into sub-spaces according to the size of the distance [10]. But, when non-linear data is separated, the convergence is inaccurate and slow [11]. To overcome this problem, a Kernel-Based Fuzzy C-Means algorithm that uses kernel functions as a substitute for Euclidean distance is used to map samples to high-dimensional space and to increase the differences between cluster centers, so they can overcome FCM deficiencies traditional and improve linear machine representation capabilities [12].

## 2. Methods

### 2.1. Clustering

With the development of technology, the number of existing data is becoming enormous, leading to what is known as Big Data. To solve problems related to this, machine learning is required [13].

Clustering is the process of entering data points with similar and dissimilar attributes into their different groups. Members of a cluster show similar characteristics with grouping based on proximity measurements or data point attributes [10].

### 2.2. Fuzzy C-Means Clustering

Among the types of clustering algorithms that exist, one of the commonly used is Fuzzy C-Means Clustering. This algorithm has the sample points that can be used to make vector space, which can further be divided into sub-spaces according to the size of the distance [10].

For data sets  $X = \{x_1, x_2, \dots, x_m\} \subseteq \mathbb{R}^d$ , defined  $n \times c$  Membership Matrix  $U = [u_{ij}]$ ,  $1 \leq i \leq n$ ,  $1 \leq j \leq c$ , and the Cluster Centre  $V = \{v_1, v_2, \dots, v_c\}$  where each object in  $V$  is an element of  $d$ -dimensional Euclidean Space [11, 14, 15].

The mathematical model of Fuzzy C-Means Clustering is as follows [11, 14, 15]:

$$J(U, V) = \min \sum_{i=1}^n \sum_{j=1}^c (u_{ij})^m d^2(x_i, v_j) \quad (1)$$

Subject to

$$\sum_{j=1}^c u_{ij} = 1, i = 1, 2, \dots, n \quad (2)$$

$$\sum_{i=1}^n u_{ij} > 0, i = 1, 2, \dots, c \quad (3)$$

$$u_{ij} \in [0, 1], j = 1, 2, 3, \dots, c \quad (4)$$

Where  $d$  is distance or dissimilarity function and  $m \in [1, \infty)$  is the fuzziness degree for partitioning.

Centre of the cluster is updated using [11, 14, 15]:

$$v_j = \frac{\sum_{i=1}^n u_{ij}^m x_i}{\sum_{i=1}^n u_{ij}^m}, j = 1, 2, \dots, c \quad (5)$$

The membership values are updated using [11, 14, 15]:

$$u_{ij} = \left( \sum_{j=1}^c \left( \frac{d(x_i, v_i^t)}{d(x_i, v_j^t)} \right)^{\frac{2}{m-1}} \right)^{-1}, 1 \leq i \leq n \quad (6)$$

FCM classification accuracy depends on the type of data. When non-linear data is separated, the convergence is inaccurate and slow [11]. To resolve this problem, the data set are transform to another feature space with higher dimensions so that the data can be separated [12].

### 2.3. Kernel Function

Kernel Function defined as  $k(x, y) = \langle \phi(x), \phi(y) \rangle$  where  $\langle \phi(x), \phi(y) \rangle$  is inner product.

$X = \{x_1, x_2, \dots, x_n\} \subset \mathbb{R}^d$ ,  $x_i \in \mathbb{R}^d$ . Where  $\phi : \mathbb{R}^d \rightarrow F$ ,  $\mathbb{R}^d$  is data space,  $F$  is feature space, and  $x, y$  are object in data space.

By using kernel function, distance between  $\phi(x)$  and  $\phi(y)$  defined as [11, 15, 16]:

$$d^2(\phi(x), \phi(y)) = \|\phi(x) - \phi(y)\|^2 \quad (7)$$

$$= \phi(x)^t \phi(x) - 2\phi(x)^t \phi(y) + \phi(y)^t \phi(y) \quad (8)$$

$$= k(x, x) - 2k(x, y) + k(y, y) \quad (9)$$

### 2.4. Kernel Based Fuzzy C-Means

By combining Fuzzy C-Means Clustering and Kernel Function, Kernel Based Fuzzy C-Means Clustering was obtained with the mathematical model as follows:

$$J(U, V) = \min \sum \sum (u_{ij})^m k(x_i, x_i) - 2k(x_i, v_j) + k(v_j, v_j) \quad (10)$$

Subject to

$$\sum_{j=1}^c u_{ij} = 1, i = 1, 2, \dots, n \quad (11)$$

$$\sum_{i=1}^n u_{ij} > 0, j = 1, 2, \dots, c \quad (12)$$

$$u_{ij} \in [0, 1], j = 1, 2, 3, \dots, c \quad (13)$$

The cluster center is updated using [14, 15]:

$$v_j = \frac{\sum_{i=1}^n u_{ij}^m x_i}{\sum_{i=1}^n u_{ij}^m}, j = 1, 2, \dots, c \quad (14)$$

The membership values updated using:

$$u_{ij} = \frac{\sum_{j=1}^c (K_{ii} - 2K_{ij} + K_{jj})^b}{\sum_{k=1}^c (K_{ii} - 2K_{ik} + K_{kk})^b}, 1 \leq i \leq n, 1 \leq j \leq c \quad (15)$$

Where  $b = -\frac{1}{m-1}$  and  $K_{ii} = k(x_i, x_i)$ ,  $K_{ij} = k(x_i, v_j)$ ,  $K_{jj} = k(v_j, v_j)$ ,  $K_{ik} = k(x_i, v_k)$ ,  $K_{kk} = k(x_k, v_k)$

For each iteration, the fuzziness degree will be different, namely [15]:

$$m = m_i + \frac{t}{T} (m_f - m_i) \quad (16)$$

Where  $m_i$  and  $m_f$  are the initial value and end value of  $m$ , respectively.

Kernel Based Fuzzy C-Means Algorithm [15]:

Input :  $X, c, m_i, m_f, \varepsilon, T$

Output:  $U$  dan  $V$

1. Intial condition:

$$V^0 = [v_1, v_2, \dots, v_c], v_j \in C_j$$

2. For  $t = 1$  to  $T$

3.  $m = m_i + \frac{t(m_f - m_i)}{T}$

4.  $b = -\frac{1}{m-1}$

5. Update membership

$$u_{ij} = \frac{\sum_{j=1}^c (K_{ii} - 2K_{ij} + K_{jj})^b}{\sum_{k=1}^c (K_{ii} - 2K_{ik} + K_{kk})^b}, 1 \leq i \leq n, 1 \leq j \leq c$$

6. Update cluster center

$$V^t = [v_1, v_2, \dots, v_c] \text{ where } v_j = \frac{\sum_{i=1}^n u_{ij}^m x_i}{\sum_{i=1}^n u_{ij}^m}, j = 1, 2, \dots, c$$

7. If  $\|V^t - V^0\| \leq \varepsilon$

STOP

else

$$t = t + 1$$

8. Go to 2

### 3. EXPERIMENT

#### 3.1. Experiment Dataset

Sinusitis data set is from the Laboratory of Radiology at Cipto Mangunkusumo National General Hospital, Indonesia (see Table 1).

**Table 1.** Sinusitis data set samples from the Laboratory of Radiology at Cipto Mangunkusumo National General Hospital, Indonesia.

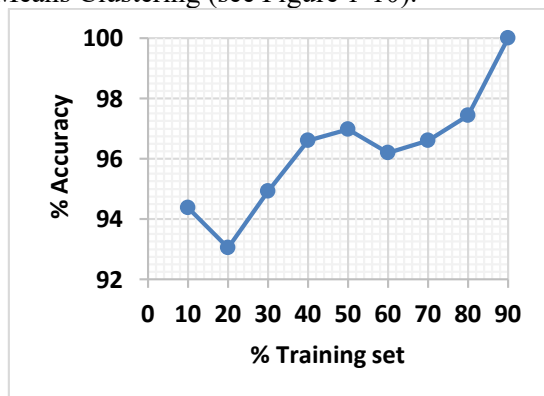
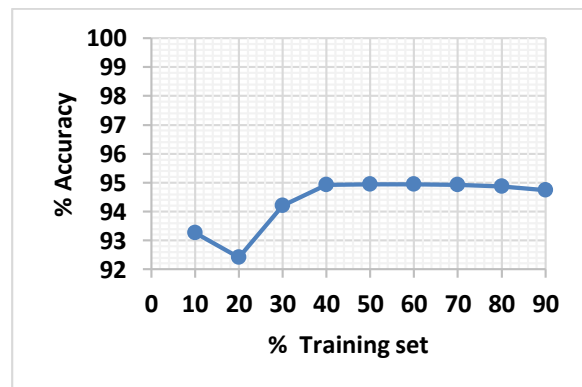
No	Gender	Age	Hounsfield Units (HU)	Air Cavity	Diagnosis
1	1	76	138	-1020	2
2	1	76	54	-1022	1
3	2	20	38	-967	1
4	2	20	24	-992	1
5	2	20	15	-987	1
6	2	20	23	-964	1

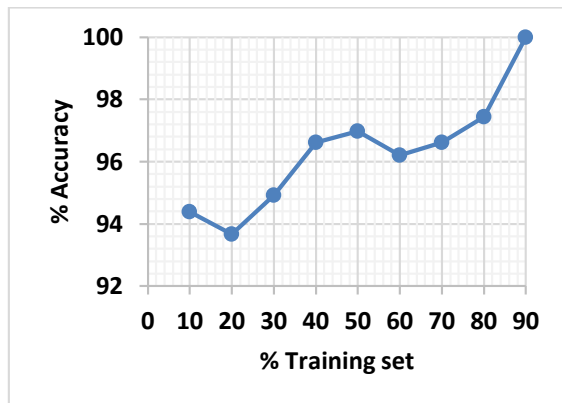
The data set has 4 attributes, 1 label's column, and 200 rows. In gender's columns, 1 refers to male, and 2 to female while in the diagnosis's columns, 1 refers to acute sinusitis, and 2 to chronic sinusitis. First, the data set is randomized with  $p\%$  taken as the training set, which is used to form the classification model. After that, the model is tested with  $(100 - p)\%$  used to examine the classification model. To get the accuracy, we following formula is used:

$$\% \text{ accuracy} = \frac{\sum \text{correct classification on testing set}}{\sum \text{all data set}} \times 100\%$$

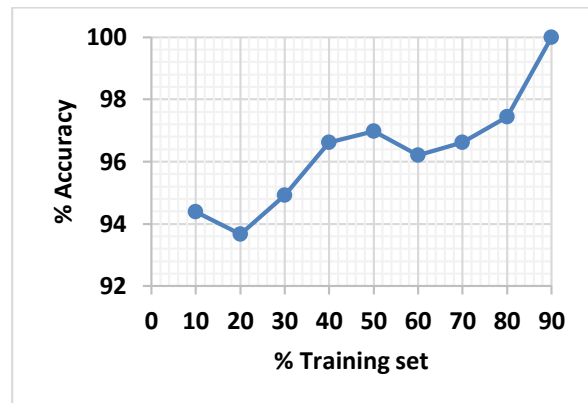
#### 4. RESULT AND DISCUSSION

The following charts are the result of chronic sinusitis classification by using Kernel Based Fuzzy C-Means Clustering (see Figure 1-10).

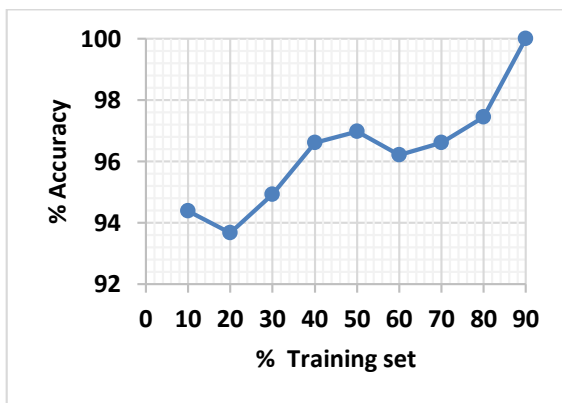
**Figure 1.** Accuracy of Kernel Based Fuzzy C-Means Clustering using  $\sigma=0.0001$ .**Figure 2.** Accuracy of Kernel Based Fuzzy C-Means Clustering using  $\sigma=0.001$ .



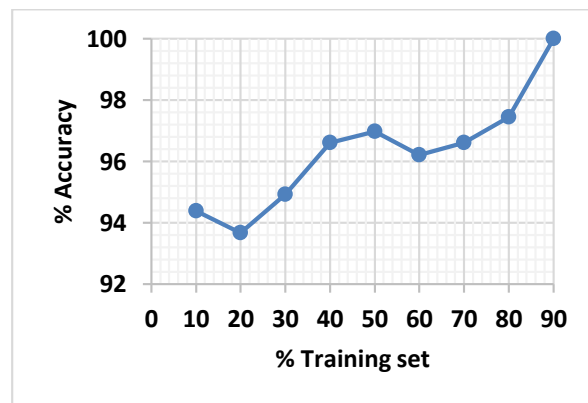
**Figure 3.** Accuracy of Kernel Based Fuzzy C-Means Clustering using  $\sigma=0.05$ .



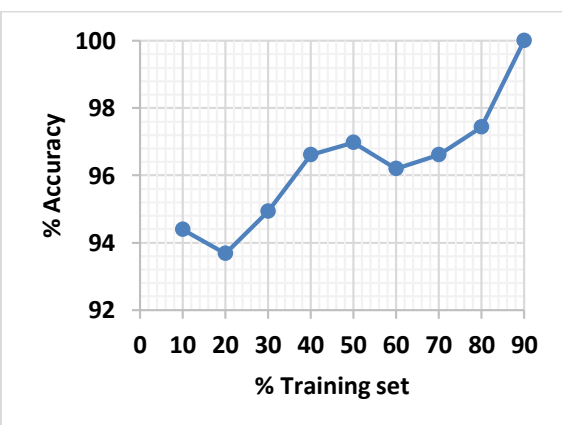
**Figure 4.** Accuracy of Kernel Based Fuzzy C-Means Clustering using  $\sigma=0.1$ .



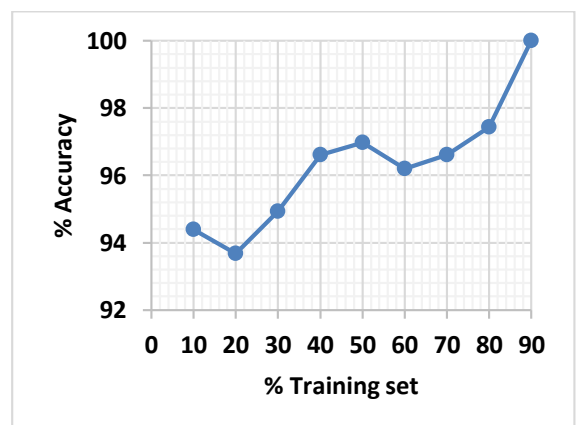
**Figure 5.** Accuracy of Kernel Based Fuzzy C-Means Clustering using  $\sigma=1$



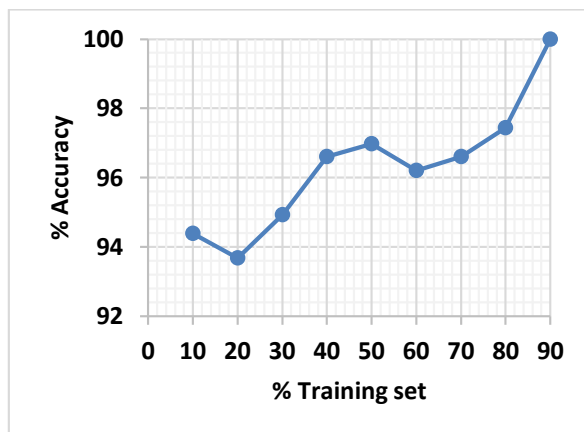
**Figure 6.** Accuracy of Kernel Based Fuzzy C-Means Clustering using  $\sigma=5$



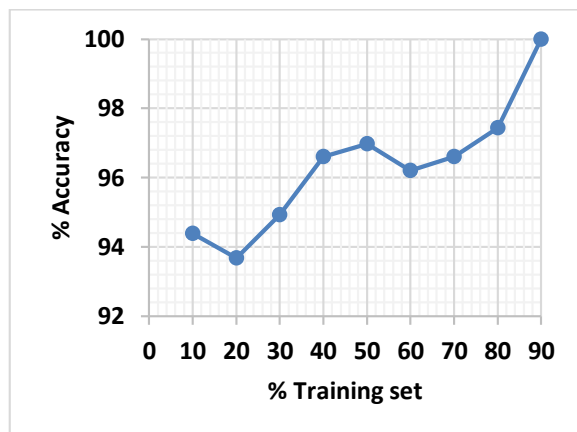
**Figure 7.** Accuracy of Kernel Based Fuzzy C-Means Clustering using  $\sigma=10$



**Figure 8.** Accuracy of Kernel Based Fuzzy C-Means Clustering using  $\sigma=50$

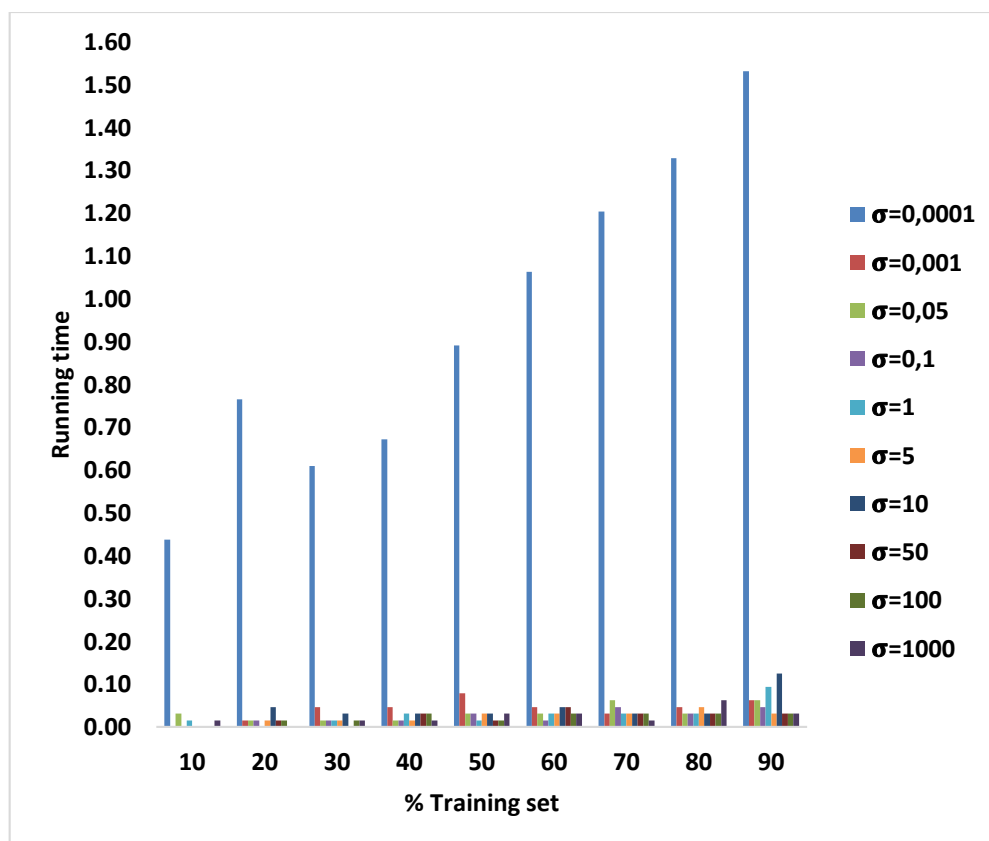


**Figure 9.** Accuracy of Kernel Based Fuzzy C-Means Clustering using  $\sigma=100$



**Figure 10.** Accuracy of Kernel Based Fuzzy C-Means Clustering using  $\sigma=1000$

From the charts above we can be seen that the best accuracy is 100%, by using 90% data training and using  $\sigma=0.0001$ ,  $\sigma=0.05$ ,  $\sigma=0.1$ ,  $\sigma=1$ ,  $\sigma=5$ ,  $\sigma=10$ ,  $\sigma=50$ ,  $\sigma=100$ ,  $\sigma=1000$ . The running time for Kernel Based Fuzzy C-Means shown as chart below (see Figure 11).

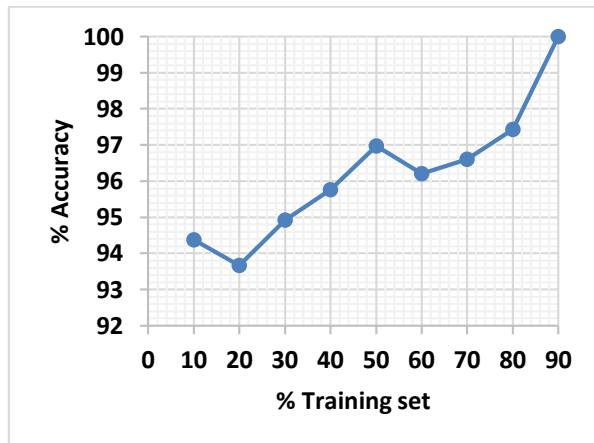


**Figure 11.** Running Time for Kernel Based Fuzzy C-Means

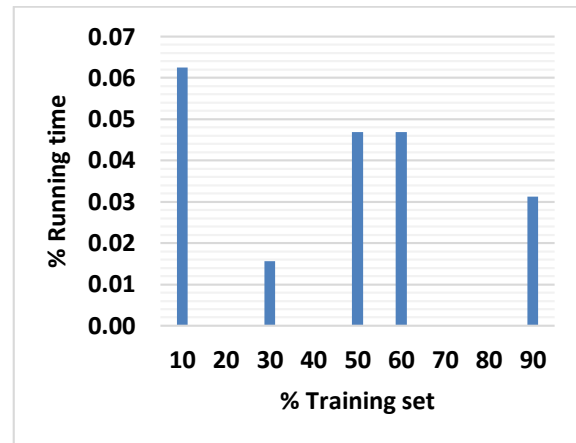
From the chart above, we can be seen that the fastest running time is close to 0, by using 10% data training and  $\sigma=0.1$ , 20% data training and  $\sigma=1$ , 10% data training and  $\sigma=5$ , 10% data training and  $\sigma=10$ , 10% data training and  $\sigma=50$ , 30% data training and  $\sigma=50$ , 10% data training and  $\sigma=50$ , 10% data training and  $\sigma=100$ , 20% data training and  $\sigma=1000$ .



By comparing the Fuzzy C-Means and Kernel Based Fuzzy C-Means, the accuracy and running time is seen in the charts below (see Figure 12&13).



**Figure 12.** Accuracy of Fuzzy C-Means Clustering



**Figure 13.** Running time of Fuzzy C-Means Clustering

From the charts above we can see that the best accuracy is 100% by using 90% data training. And the fastest running time is close to 0 by using 20%, 40%, 70%, and 80% data training. Information on the actual and prediction classification done by a classification system is shown as a confusion matrix (see Tables 2 and 3).

**Table 2.** Confusion matrix using kernel

		Actual	
		Class 1	Class 2
Prediction	Class 1	48.72 (TP)	0.00 (FP)
	Class 2	2.56 (FN)	48.72 (TN)

**Table 3.** Confusion matrix without kernel

		Actual	
		Class 1	Class 2
Prediction	Class 1	52.63 (TP)	0.00 (FP)
	Class 2	0.00 (FN)	47.37 (TN)

From the confusion matrixes above, other parameters can be computed that can also be used to compare the two methods. These are precision, recall, and f1 score. Precision is ability of a classification model to return only relevant instances, recall is the act of identify all relevant instances, while F1 score is single metric that combines recall and precision using the harmonic mean (see Table 4) [17].

**Table 4.** Table of Accuracy, precision, recall, and F1 score for Fuzzy C-Means and Kernel Based Fuzzy C-Means

	Formula	Fuzzy C-Means Clustering	Fuzzy Kernel Based C-Means Clustering
Accuracy	$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}$	1	0,9744
Precision	$\text{Precision} = \frac{TP}{TP+FP}$	1	1
Recall	$\text{Recall} = \frac{TP}{TP+FN}$	1	0,95

F1 Score	$2 \times \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$	1	0,974
----------	--	---	-------

From the table above, we can be see that Kernel Based Fuzzy C-Means has high values though not bigger that Fuzzy C-Means. It can be concluded that the weakness of this experiment lies in the researcher's inability to prove that the Kernel Based Fuzzy C-Means is better than Fuzzy C-Means in classifying chronic sinusitis.

## 5. CONCLUSION AND FUTURE WORK

From the figures, it can be concluded that Kernel Based Fuzzy C-Means has high accuracy with the best accuracy obtained at 100%, by using 90% data training and using  $\sigma=0.0001$ ,  $\sigma=0.05$ ,  $\sigma=0.1$ ,  $\sigma=1$ ,  $\sigma=5$ ,  $\sigma=10$ ,  $\sigma=50$ ,  $\sigma=100$ ,  $\sigma=1000$ . The fastest running time is close to 0, by using 10% data training and  $\sigma=0.1$ , 20% data training and  $\sigma=1$ , 10% data training and  $\sigma=5$ , 10% data training and  $\sigma=10$ , 10% data training and  $\sigma=50$ , 30% data training and  $\sigma=50$ , 10% data training and  $\sigma=50$ , 10% data training and  $\sigma=100$ , 20% data training and  $\sigma=1000$ . For the next step, a bigger data is used to illustrate that Kernel Based Fuzzy C-Means is better than Fuzzy C-Means.

## ACKNOWLEDGEMENT

This research was financially supported by University of Indonesia, with PIT.9 2019 research grant scheme (ID number NKB-0039/UN2.R3.1/HKP.05.00/2019).

## References

- [1] Almutairi AFN, Shafi RW, Albalawi SA, Basyuni MA, A Alzahrnai A, AA Alhaifi, Alshehri AA, Al-Gadouri MA, Alghamdi (2017). Acute and Chronic Sinusitis Causes and Management. The Egyptian Journal of Hospital Medicine (July 2017) Vol.68 (3), Page 1513-1519.
- [2] Head K, Chong LY, Piromchai P, Hopkins C, C Philpott, Schilder AG et al. (2016):Systemic and topical antibiotics for chronic rhinosinusitis. Cochrane Database Syst Rev. 4: CD011994
- [3] LY Chong, Head of K, Hopkins C, C Philpott, Glew S, Scadding G et al. (2016):Saline irrigation for chronic rhinosinusitis. Cochrane Database Syst Rev., 4: CD011995.
- [4] American Academy of Pediatrics (2001): Subcommittee on Management of Sinusitis and Committee on Quality Management. Clinical practice guideline: management of sinusitis. Pediatrics, 108(3):798-808.
- [5] <https://hellosehat.com/penyakit/sinusitis-kronis/> . Assessed in February 2019.
- [6] <https://www.alodokter.com/sinusitis> . Accessed in February 2019.
- [7] Slavin RG, Spector SL, Bernstein IL *et al.* (2005):The diagnosis and management of sinusitis: a practice parameter update. J Allergy Clin Immunol., 116(6):S13-47.
- [8] Falagas ME, Giannopoulou KP, Vardakas KZ, Dimopoulos G, Karageorgopoulos DE(2008): Comparison of antibiotics with placebo for treatment of acute sinusitis: a meta-analysis of randomised controlled trials. Lancet Infect Dis., 8(9):543-52
- [9] López-Chacón M, Mullol J, Pujols L(2015): Clinical and biological markers of difficult-to-treat severe chronic rhinosinusitis. Curr Allergy Asthma Rep. , 15 (5):19.
- [10] Prabakar HS, Prof Kumar S (2017). A Survey on Fuzzy C-Means Clustering Techniques. 2017 IJEDR Volume 5, Issue 4 ISSN: 2321-9939.
- [11] Z Rustam and AS Talita. *Fuzzy Kernel C-Means Algorithm for Intrusion Detection Systems*. Journal of Theoretical and Applied Information Technology. Vol.81. No.1. 10<sup>th</sup> November 2015.
- [12] Jinxian Lin,Shuangyang Zheng, “An optimizing search based on Kernel-based Fuzzy C-Means Clustering”, China.
- [13] Pérez-Ortega J, Almanza-Ortega NN, Romero D (2018). Balancing effort and benefit of K-

- means clustering algorithms in Big Data realms. PLoS ONE 13 (9): e0201874. <https://doi.org/10.1371/journal.pone.0201874>,
- [14] Z Rustam.and, D Zahras. *Comparison between Support Vector Machine and Fuzzy C-Means as Classifier for Intrusion Detection System*. 2018 Journal of Physics: Conference Series.
- [15] Z Rustam and F.Yaurita. *Insolvency Prediction in Insurance Companies Using Support Vector Machines and Fuzzy Kernel C-Means*. Journal of Physics: Conference Series. 2018.
- [16] Z Rustam and AS Talita. *Fuzzy Kernel K-Medoids Algorithm for Multiclass Multidimensional Data Classification*. Journal of Theoretical and Applied Information Technology. Vol.80.No.1.10<sup>th</sup> October 201.
- [17] <https://towardsdatascience.com/beyond-accuracy-precision-and-recall-3da06bea9f6c>. Accessed on March 3, 2019.