Kernel Based Comparison between Fuzzy C-Means and Support Vector Machine for Sinusitis Classification

R A Putri 1*, Z Rustam 2, J Pandelaki 3, N Salmi 4

1,2,4 Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Indonesia, Kampus UI Depok, Depok 16424, Indonesia
3 Department of Radiology, Cipto Mangunkusumo National General Hospital, DKI Jakarta 10430, Indonesia

*Email: 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 side of the nose, and behind the eyes. Sinusitis is caused by infection, growth of nasal polyps, allergies, and others. This condition can effect adults, teenagers, and even children. To classify sinusitis, we used Kernel Based Fuzzy C-Means, which is the development of Fuzzy C-Means (FCM). FCM algorithm groups data using Euclidean distance. However, when non-linear data is separated, the convergence is inaccurate and need a long-running time. To overcome this problem, a Kernel Based Fuzzy C-Means that use kernel functions as a substitute for Euclidean distance. It maps objects from data space to a higher dimension feature space, so they can overcome FCM deficiencies. Beside we used Kernel Based Support Vector Machine to do the same thing, that separate the data set by hyperplane. From the result of both methods, we will compare both of them to get the best method for the data set. Data that is used is sinusitis data set obtained from the laboratory of radiology at Cipto Mangunkusumo National General Hospital, Jakarta. From the experiment we got 100% accuracy of Kernel Based Fuzzy C-Means and 100% accuracy of Kernel Based Support Vector Machine using the same parameter sigma for the kernel.

Keywords: Kernel function, Fuzzy C-means, support vector machines

1. Introduction
Sinusitis is an inflammation of one or more of the paranasal sinuses. It is classified into four types, namely acute sinusitis, sub acute sinusitis, chronic sinusitis, and recurrent sinusitis [1]. Sinusitis is caused by viral infections, allergies, smoking, diabetes mellitus, swimming, diving, climbing heights, dental infections, the use of breathing tubes and food canisters, asthma, immune deficiency, sinus surgery, and others [2]. Children under 15 years of age and adults aged 25 to 64 years are most susceptible to sinusitis [3]. To find out whether a person has sinusitis or several tests cannot be performed, one of them is a Computed Tomographic (CT) that shows the existing abnormality, and guides therapist intervention [4].

In this research, the authors used numeric data from the results of CT scan of patients suffering from sinusitis. The data will be grouped based on the type of sinusitis using the Kernel-Based Fuzzy
C-Means and Kernel Based Support Vector Machine method. Kernel Based Fuzzy C-Means is the
development of the Fuzzy C-Means method. The accuracy of the FCM depends on the type of data
used. If the data to be separated is non-linear, then the accuracy will be small and the running time
required will be long. Therefore, a tool is needed that can map data to feature space whose dimensions
are higher than data space. With this step, it is expected that the mapped data can be separated linearly
so that accuracy will be higher and running time will be faster [5]. In this study, the author uses the
kernel function as a tool that can solve the above problems.

Beside we also used Kernel Based Support Vector Machine to classify the data set. Support Vector
Machine (SVM) separate the data set to their perspective classes using hyperplane. The optimal
hyperplane in question is a field that separates data into in its class, and it is perpendicular
to its closest pattern. Pattern is the points that describe a data set [6].

2. Methods
The methods used in this paper are Kernel Based Fuzzy C-Means (FCM) and Kernel Based Support
Vector Machine (SVM) using sinusitis data set.

2.1. Kernel Function
Kernel function is defined as [5]:

\[
k(x, y) = \langle \phi(x), \phi(y) \rangle \tag{1}
\]

\[
= \phi(x)^T \phi(y) \tag{2}
\]

Where \( \langle \phi(x), \phi(y) \rangle \) is inner product and \( x, y \in R^d \).

Using kernel function, the distance between \( \phi(x) \) and \( \phi(y) \) is defined as [5]:

\[
d^2(x, y) = \|\phi(x) - \phi(y)\|^2 \tag{3}
\]

\[
= \langle \phi(x), \phi(x) \rangle - 2\langle \phi(x), \phi(y) \rangle + \langle \phi(y), \phi(y) \rangle \tag{4}
\]

\[
= k(x, x) - 2k(x, y) + k(y, y) \tag{5}
\]

There are several types of kernels, one of which is Gaussian Radial Basis Function (RBF) [5]:

\[
k(x, y) = \exp\left(\frac{-\|x-y\|^2}{2\sigma^2}\right) \tag{6}
\]

Where \( \sigma > 0 \) is parameter.

2.2. Fuzzy C-Means
Fuzzy C-Means (FCM) is a grouping that minimizes objective functions. The purpose of FCM is to
form clusters from data points that have membership values close to existing classes or clusters. The
objective functions for Fuzzy C-Means are [5,7]:

\[
J(U, V) = \min \sum_{i=1}^{N} \sum_{j=1}^{c} u_{ij} m d^2(x_i, v_j) \tag{7}
\]

Subject to:

\[
\sum_{j=1}^{c} u_{ij} = 1 \tag{8}
\]

Where:
- \( N \) is the number of data
- \( C \) is the number of clusters
- \( u_{ij} \) is membership degree \( i \) data in \( j \) cluster
- \( [U_{ij}] = \begin{bmatrix} U_{11} & \cdots & U_{1C} \\ \vdots & \ddots & \vdots \\ U_{N1} & \cdots & U_{NC} \end{bmatrix} \), with \( 1 \leq i \leq N, and 1 \leq j \leq c. \)
• \( m \) is fuzziness degree, \( m > 1 \).
• \( d(x_i, v_j) = \|x_i - v_j\| \) is a distance between \( x_i \) (i-th data set) with cluster center \( v_j \).

To get Kernel Based Fuzzy C-Means, we can change the distance in Fuzzy C-Means with kernel function. So we get the formula such as:

\[
J(U, V) = \min 2 \sum_{i=1}^{N} \sum_{j=1}^{c} (u_{ij})^m \left(1 - k(x_i, v_j)\right)
\]

(9)

Where [8]:

\[
u_{ij} = \left(\sum_{k=1}^{c} \left(\frac{1-k(x_i,v_j)}{1-k(x_i,v_k)}\right)^{\frac{1}{m-1}}\right)^{-1}
\]

(10)

And [8]:

\[
v_j = \frac{\sum_{i=1}^{N} (u_{ij})^m (k(x_i,v_j)) x_i}{\sum_{i=1}^{N} (u_{ij})^m (k(x_i,v_j))}
\]

(11)

Kernel Based Fuzzy C-Means Algorithm:

1. Determine:
   a. Number of training set
   b. Number of cluster \( (c \geq 2) \).
   c. Fuzziness degree \( (m) \)
   d. \( \varepsilon \)
   e. Parameter \( \sigma \)
   f. Initial \( u^0 = [u_{ij}] \) and \( V^0 = \{v_1, v_2, ..., v_c\} \)

2. Cluster center \( (v_j) \): \( V^t = \{v_1, v_2, ..., v_c\} \)

\[
v_j = \frac{\sum_{i=1}^{N} (u_{ij})^m (k(x_i,v_j)) x_i}{\sum_{i=1}^{N} (u_{ij})^m (k(x_i,v_j))} \quad , \quad j = 1,2, ..., c
\]

3. Update membership degree

\[
u_{ij} = \left(\sum_{k=1}^{c} \left(\frac{1-k(x_i,v_j)}{1-k(x_i,v_k)}\right)^{\frac{1}{m-1}}\right)^{-1}, \quad i = 1,2, ..., N \quad \text{dan} \quad j = 1,2, ..., c
\]

4. If \( |V^t - V^{t-0}| < \varepsilon \) stop, if no back to iteration 2.

2.3. Support Vector Machine

The concept of the Support Vector Machine (SVM) is to form a field or optimal hyperplane that separates data into their perspective classes. The optimal hyperplane in question is a field that separates data into its class, and it is perpendicular to its closest pattern. Pattern is the points that describe a data set [6].

Off all the hyperplane, an optimal hyperplane will be searched it is perpendicular to the pattern. To find the optimal hyperplane, the maximum margin will be sought. Margin is the distance between hyperplane and the closest pattern to each class. Pattern that is located closest to optimal hyperplane is called a support vector [6].

The optimal hyperplane formed is illustrated in the figure below (see Figure 1).
Figure 1. Optimal hyperplane created from maximum margin (Source: Sayad, 2010)

Suppose there is as many data as N. For example \((x_i, y_i)\) as the data set with \(i = 1,2,..., N\). Features that are input data can be written as \(x_i \in \mathbb{R}^n\), while the class label that is output can be written as \(y_i \in \{-1,1\}\). Hyperplane to be formed can be written with equation bellow [9]:

\[
y(x) = w^T x + b
\]

(12)

Given \(w\) is a vector of weight parameter values, and \(b\) is a bias scalar value.

To separate data set doing based on this formula [9]:

\[
w^T x + b \geq 1, y_i = +1
\]

(13)

\[
w^T x + b \leq 1, y_i = -1
\]

(14)

The optimization problem of SVM is minimized [7]:

\[
\frac{1}{2} \|w\|^2
\]

(15)

s.t

\[
y_i (w^T x + b) \geq 1, \forall i = 1,2,...,N
\]

(16)

and the decision function is [7]:

\[
f(x) = w \cdot x + b
\]

(17)

Where [10]:

\[
w = \sum_{i=1}^{n} a_i^* y_i x_i
\]

(18)

\[
b = \frac{1}{N_x} \sum_{i \in S}(y_i - \sum_{m \in S} a_m y_m x_m)
\]

(19)

3. Experiment

3.1. Experiment dataset

This research used sinusitis data set from Department of Radiology, National General Hospital, Indonesia. In this data, there are two types of sinusitis, they are acute sinusitis and chronic sinusitis presented in Table 1 (see Table 1).

In gender’s feature, 1 refers to male, and 2 to female while in the diagnosis’s columns, 1 refers to acute sinusitis, and 2 to chronic sinusitis. The data set contains 200 rows.

First, the data set is split to training set and testing set. \(p\)% of the data set 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. From the result of testing set we will get confusion matrix that used to get some evaluation such as accuracy, recall, and others.
Table 1. Sinusitis data set

| No | Gender | Age (year) | Hounsfield Units (HU) | Air Cavity | Diagnosis |
|----|--------|------------|-----------------------|------------|-----------|
| 1  | 1      | 76         | 138                   | -1020      | 2         |
| 2  | 1      | 76         | 54                    | -1022      | 1         |
| 3  | 2      | 20         | 38                    | -967       | 1         |
| ...| ...    | ...        | ...                   | ...        | ...       |
| 199| 1      | 62         | 93                    | -1009      | 2         |
| 200| 2      | 62         | 79                    | -575       | 2         |

4. Result and Discussion
The following table is the result of sinusitis classification by using Kernel Based Fuzzy C-Means using some sigma (see Table 2).

Table 2. Result of kernel based fuzzy c-means

| Training set | $\sigma = 0.5$   | $\sigma = 0.1$   | $\sigma = 1$     | $\sigma = 5$     |
|--------------|------------------|------------------|------------------|------------------|
| 10%          | 95,50562         | 95,50562         | 95,50562         | 95,50562         |
| 20%          | 93,03797         | 93,03797         | 93,03797         | 93,03797         |
| 30%          | 95,65217         | 95,65217         | 95,65217         | 95,65217         |
| 40%          | 96,61017         | 96,61017         | 96,61017         | 96,61017         |
| 50%          | 96,9697         | 96,9697         | 96,9697         | 96,9697         |
| 60%          | 96,20253        | 96,20253        | 96,20253        | 96,20253        |
| 70%          | 96,61017        | 96,61017        | 96,61017        | 96,61017        |
| 80%          | 97,4359         | 97,4359         | 97,4359         | 97,4359         |
| 90%          | 100             | 100             | 100             | 100             |

From the table above, we know that the highest accuracy of Kernel Based Fuzzy C-Means got by using 90% training set and $\sigma = 0.5, \sigma = 0.1, \sigma = 1,$ and $\sigma = 5.$

The following table is the result of sinusitis classification by using Kernel Based Support Vector Machine using some sigma (see Table 3).

Table 3. Result of kernel based support vector machine

| Training set | $\sigma = 0.5$   | $\sigma = 0.1$   | $\sigma = 1$     | $\sigma = 5$     |
|--------------|------------------|------------------|------------------|------------------|
| 10%          | 0.906            | 0.967            | 0.894            | 0.761            |
| 20%          | 0.981            | 0.963            | 0.9875           | 0.931            |
| 30%          | 0.986            | 0.971            | 0.993            | 0.964            |
| 40%          | 0.967            | 0.975            | 0.967            | 0.917            |
| 50%          | 0.99             | 0.99             | 0.98             | 0.95             |
| 60%          | 0.9625           | 0.975            | 0.9625           | 0.925            |
| 70%          | 0.983            | 0.95             | 0.967            | 0.967            |
| 80%          | 0.975            | 0.925            | 0.9              | 0.975            |
| 90%          | 1                | 1                | 0.95             | 0.95             |
From the table above, we know that the highest accuracy of Kernel Based Support Vector Machine got by using 90% training set and $\sigma = 0.5$ and $\sigma = 0.1$.

From the experiment, we also get the confusion matrix of Kernel Based Fuzzy C-Means and Kernel Based Support Vector Machine (see Table 4 and Table 5).

**Table 4. Confusion matrix of kernel based fuzzy c-means**

| Actual | Prediction | | |
|---|---|---|
| Class 1 | Class 2 |
| Class 1 | 52.63 % | 0 % |
| Class 2 | 0 % | 47.37 % |

**Table 5. Confusion matrix of kernel based support vector machine**

| Actual | Prediction | |
|---|---|---|
| Class 1 | Class 2 |
| Class 1 | 50 % | 0 % |
| Class 2 | 0 % | 50 % |

From table 4 and table 5 above, we get some evaluation of both method, they are accuracy, precision, recall, specificity, and F1 score. Accuracy is defined as the probability of correctly predicting classes of data. Sensitivity is defined as the true positive proportion of the total number of positive events. Specificity is defined as the true negative proportion of the total number of negative events. Precision is a measure of true positive ratio compared to overall positive predictions. F1-score is a harmonic average of precision and sensitivity [11]. To see the evaluation result of these methods (see Figure 2).

**Figure 2. Comparison of kernel based fcm and kernel based svm**
From Figure 2 above, we know that Kernel Based FCM and Kernel Based SVM have the same evaluation value if we use 90% training data with $\sigma = 0.5$ and $\sigma = 0.1$. They have 100% of accuracy, 100% of precision, 100% of recall, 100% of specificity, and 100% of F1 score. So that, Kernel Based FCM and Kernel Based FCM have the same performance in classify sinusitis data set.

5. Conclusion and Future Work
From this experiment, we conclude that Kernel Based FCM and Kernel Based SVM have the same performance in classify sinusitis data set. They got 100% of accuracy, 100% of precision, 100% of recall, 100% specificity, and 100% of F1 score, by using 90% training set with $\sigma = 0.5$ and $\sigma = 0.1$. It can happen because the data set used is not too non-linear.

So that, the future work of this experiment is doing the experiment using the bigger data set, more non-linear, have many more features, another kernel function, and using many more sigma. So that, we can see the significant differences between these methods.

6. Acknowledgement
This research supported financially by the Ministry of Research and Higher Education Republic of Indonesia (KEMENRISTEKDIKTI) with PTUPT 2020 research grant scheme.

References
[1] Leung, R.S., & Katial, R. (2008). The diagnosis and management of acute and chronic sinusitis. Primary Care: Clinics in Office Practice, 35, 11-24
[2] Almutairi, A.F.N., Shafi, R.W., Albalawi, S.A., Basyuni, M.A., Alzahrani, A.A., Alhaifi, A.A., . . . Alghamdi, A.S. (2017). Acute and chronic sinusitis cause and management. The Egyption Journal of Hospital Medicine, 68(3), 1513-1519
[3] Battisti, A.S., Pangia, A. (2018). Sinusitis. October 27, 2018. https://www.ncbi.nlm.nih.gov/books/NBK470383/
[4] April, M.M., Zinreich, S.J., Baroody, M.F., & Naclerio, R.M. (1993). Chronic sinusitis. Laryngoscope, 103.
[5] Rustam, Z., & Yaurita, F. (2018). Insolvency prediction in insurance companies using support vector machines and fuzzy kernel c-means. Journal of Physics: Conference Series, 1028(1), [012118].
[6] Srivastava, D. K. & Bhambhu, L. (2005). Data classification using Support Vector Machines. Journal of Theoretical and Applied Information Technology.
[7] Rustam, Z., & Zahras, D. (2018). Comparison between support vector machine and fuzzy c-means as classifier for intrusion detection system. Journal of Physics: Conference Series, 1028(1), [012227].
[8] Graves, D., & Pedrycz, W. (2009). Kernel-based fuzzy c-means clustering and fuzzy clustering: A comparative experimental study. Fuzzy Sets and Systems, 161, 522-543.
[9] Karatzoglou, A., Meyer, D., & Varsha, J. (2006). Support Vector Machines in R. Journal of Statistical Softwatre, 15(9).
[10] Rampisela., T. V., & Rustam, Z. (2018). Classification of Schizophrenia Data Using Support Vector Machine (SVM). IOP Conf. Series: Journal of Physics: Conf. Series 1108 (2018) 012044.
[11] Galdi, P., & Tagliafeci, R. (2018). Data mining: accuracy and error measures for classification and prediction.