Application of Fuzzy Kernel C-Means in face recognition to identify look-alike faces

Zuherman Rustam\textsuperscript{1,a} and Ridhani Faradina\textsuperscript{2,b}

\textsuperscript{1,2}Department of Mathematics, Universitas Indonesia, Depok, 16424, Indonesia.

E-mail: \textsuperscript{a}rustam@ui.ac.id, \textsuperscript{b}ridhani.faradina@sci.ui.ac.id

Abstract. Machine learning has been rapidly evolving and continuously developing. Many problems can be solved by machine learning. One of those is face recognition. There are many application of face recognition. One application that will be discussed in this research is face recognition to identify look-alike faces. Such application can be useful when dealing with huge data. Machine learning method that will be used is Fuzzy Kernel C-Means. This method is the modification of previous method Fuzzy C-Means. The kernel used is Radial Basis Function Kernel. Each image is characterised by its features. It is believed that reducing the number of features can also reduce the cost. Therefore, a feature selection method called Chi-Square was also used. Solving this problem in face recognition using Fuzzy Kernel C-Means resulted in quite high accuracies.

1. Introduction
The data in this world are so huge and various such that the world needs a solution to process these data. However, one solution is never enough. One way we can do is to classify the data into classes. This is where machine learning comes in. Machine learning is a part of artificial intelligence that focuses on computers to 'learn' and act or give decisions like humans [1].

The main idea of machine learning is to replace a machine designed to do a job with specific instructions into a set of algorithms. The algorithms are trained to use the data and to gather information to be studied which will later be formed into a model. A method in machine learning called Fuzzy C-Means with the implementation of kernel will be used in this research. This method is often called as Fuzzy Kernel C-Means.

One application of Fuzzy Kernel C-Means to a real-world problem is face recognition. High accuracy and short running time are usually the main aim for researchers to obtain when constructing face recognition algorithms. There have been several face recognition problem that have been discussed by other researchers, such as face recognition in general [2], face recognition for the purpose of gender classification [3], and face recognition in the area for look-alike faces [4,5,6].

The problem that will be discussed in this research is look-alike faces identification. It is useful to cut the cost of searching for missing person, criminal or wanted person. Unqualified faces can be removed and the process is left with look-alike face images. Removing the unqualified data can result in saving more time. While the data in the real-world are likely to be huge, removing a huge amount of data and continuing with only the qualified data can reduce the cost [4]. The database that is used in this research is the Look-Alike Face (LAF) database that was made by Lamba et al. in 2011 [5].
The basic reasons of look-alike faces are described in three types \([5,6]\). The first reason is biological, the second reason is situation, and the third reason being plastic surgery or any other artificial handling to the face. Biological reason includes those people who are related by blood, particularly identical twins who happen to very likely look almost the same. Situation reason depends on things such as the lighting and pose of a picture being taken \([6]\). For plastic surgery, it is common for people who had plastic surgery to have similar shape in those commonly changed spots on the face such as nose, chin, or lips.

The two main stages performed for face recognition are feature selection or extraction and classification \([7]\). A feature selection or extraction stage is performed to reduce the dimensions of the image without losing important information in the image. In this research, feature selection will be used along with classification. The feature selection method used in this research is Chi-Square.

This paper is arranged as follows: The methods used in this research, which include Chi-Square as the feature selection method, Fuzzy Kernel C-Means as the classification method and confusion matrix as the performance evaluation tool to calculate the accuracy are explained in section 2. The data used in this research are described in section 3. Experiment results with the explanations are given in section 4 and the conclusions are given in section 5.

2. Method

This section discusses about the methods used for this research. First, Chi-Square as the feature selection method is explained briefly. Second, the idea and fundamental formulation for kernel and Fuzzy Kernel C-Means are explained. Lastly, the general form of confusion matrix is explained.

2.1. Chi-Square Feature Selection

Chi-Square feature selection uses the basic concept of looking for a feature connection with a class and then rank those connections. It uses statistical theory to test the independence of a feature with its class \([8]\). Each feature will calculate its Chi-Square value respectively, denoted by \(\chi^2\). The formula used to compute \(\chi^2\) is \([9]\):

\[
\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{(n_{ij} - \mu_{ij})^2}{\mu_{ij}}
\]

where \(j\) is the \(j^{th}\) class, \(i\) is the \(i^{th}\) feature, \(n_{ij}\) is the feature value \(i\) in class \(j\) and \(\mu_{ij}\) is defined as:

\[
\mu_{ij} = \frac{n_{ij}n_{i*}}{n}
\]

where \(n_{j*} = \sum_{i=1}^{k} n_{ij}\) and \(n_{i*} = \sum_{j=1}^{m} n_{ij}\).

2.2. Kernel

Any great machine learning methods can produce high accuracy for data that can be separated linearly in general. The problem in real world is that such linearly separable data is almost impossible to find. One technique that can be used as the solution to trick the problem of non-linear data is by implementing kernel \([10]\). The kernel function can be considered as a link between non-linear data and the algorithms. Kernel function will map data in the input space to a new feature space that has a higher dimension such that the data can be separated linearly in the new feature space \([11]\).

Suppose \(x_1, x_2, ... , x_n \in \mathbb{R}^n\) is a set of the original data in \(\mathbb{R}^n\). There is a function \(\phi\) that maps the data in \(\mathbb{R}^n\) to a new feature space with a higher dimension \(F\).

\[
\phi: \mathbb{R}^n \rightarrow F
\]

The kernel function is defined as follows \([12]\):

\[
K(x, y) = \phi(x) \cdot \phi(y)
\]
and let the data that have been mapped to a new feature space be called as kernelised data. Thus, the
distance between two kernelised data is defined as:
\[
d(x, y) = \|\phi(x) - \phi(y)\| = \sqrt{K(x, x) - 2K(x, y) + K(y, y)}
\] (4)

The distance between two kernelised data which is stated in equation (4) will help to simplify
further process of finding the necessity of any method that will be implemented by kernel. There are
several types of kernel, but a type of kernel has been commonly used for Fuzzy Kernel C-Means is the
Radial Basis Function (RBF) kernel. The RBF kernel is defined as [12]:
\[
K(x, y) = e^{\frac{-\|x-y\|^2}{2\sigma^2}}
\] (5)

2.3. Fuzzy Kernel C-Means
The Fuzzy C-Means method is a clustering method that was introduced by Bezdek in 1981 [10]. A
modification of the Fuzzy C-Means method by implementing the kernel has been then developed. The
modified method is called Fuzzy Kernel C-Means. The implementation of kernel is done so that the
algorithm can correctly classify the data which cannot be linearly separated [13].

Few examples of previous researches using Fuzzy C-Means are its application in cancer
classification [14] and intrusion detection system [15]. While the examples of previous researches
using Fuzzy Kernel C-Means are for intrusion detection system [16], gliomatosis cerebri classification
[17], and insolvency prediction in insurance companies [18].

The objective functions of Fuzzy Kernel C-Means are as follows [19]:
\[
f(V, U, X, c, m) = \sum_{i=1}^{c} \sum_{k=1}^{N} (u_{ik})^m \|\phi(x_k) - \phi(v_i)\|^2
\] (6)

subject to:
\[
0 \leq u_{ik} \leq 1
\] (7)
\[
\sum_{i=1}^{c} u_{ik} = 1
\] (8)
\[
0 \leq \sum_{i=1}^{c} u_{ik} \leq N
\] (9)
\[
i = 1, 2, ..., c
\]
\[
k = 1, 2, ..., N
\]

where \(c\) is the number of clusters which needs to be more or equal to two, \(N\) is the number of data,
\(m\) is the fuzziness degree (\(m > 1\)), \(u_{ik}\) is the membership of \(x_k\) in class \(i\), \(X = \{x_1, x_2, ..., x_N\}\) is the
set of data that are to be clustered, and \(V = \{v_1, v_2, ..., v_c\}\) is the set of cluster centers.

The optimal condition is obtained when the objective function is minimised. One way to find the
solution is to use Lagrange multiplier. Suppose given \(\lambda\) as the Lagrange multiplier, then it is possible
to establish a Lagrange function for Fuzzy Kernel C-Means as follows:
\[
L = \sum_{i=1}^{c} \sum_{k=1}^{N} (u_{ik})^m \|\phi(x_k) - \phi(v_i)\|^2 + \lambda(1 - \sum_{i=1}^{c} u_{ik})
\] (10)

By solving the derivatives of Lagrange equation in (10) above and substituting the RBF kernel
formula, we would get \(u_{ik}\) as the membership of \(x_k\) in class \(i\) and \(v_i\) as the cluster center for class \(i\) as follows [20]:
\[
u_{ik} = \left(\sum_{j=1}^{c} \frac{(1-K(x_k, v_j))^{m-1}}{1-K(x_k, v_j)}\right)^{-1}
\] (11)
\[
v_i = \frac{\sum_{k=1}^{N} (u_{ik})^m K(x_k, v_i) x_k}{\sum_{k=1}^{N} (u_{ik})^m K(x_k, v_i)}
\] (12)
2.4. Confusion matrix

When conducting a scientific experiment, a performance evaluation is needed so that the results are described systematically. The performance evaluation tool used in this research is confusion matrix. A confusion matrix shows the performance results of a machine learning algorithm in the form of a special matrix measuring $n \times n$, where $n$ is the number of classes [21]. By using confusion matrix, the accuracy can be seen clearly.

Confusion matrix can be used for binary classification as well as multiclass classification. For binary classification, there would be only two classes at stake. The general form for confusion matrix in binary classification is shown in table 1.

| True Situation | Class 1 | Class 2 |
|----------------|---------|---------|
| Class 1        | $f_{11}$ | $f_{12}$ |
| Class 2        | $f_{21}$ | $f_{22}$ |

From table 1, $f_{11}$ is called true positive, $f_{12}$ is called false negative, $f_{21}$ is called false positive, and $f_{22}$ is called true negative [21]. True positives and true negatives are labeled for data that are correctly classified, while false positives and false negatives are labeled for data that are incorrectly classified. For multiclass classification, table 1 can be expanded until $m$ classes which is shown in table 2.

| True Situation | Class 1 | Class 2 | ... | Class $m$ |
|----------------|---------|---------|-----|-----------|
| Class 1        | $f_{11}$ | $f_{12}$ | ... | $f_{1m}$ |
| Class 2        | $f_{21}$ | $f_{22}$ | ... | $f_{2m}$ |
| Class $m$      | $f_{m1}$ | $f_{m2}$ | ... | $f_{mm}$ |

If we want to know the accuracy, we have to calculate how many data are correctly classified and divide it by the total data being classified. So, all we need is the results in the diagonal of confusion matrix and the total data. The diagonal of confusion matrix would consist of the true positives and the true negatives. Formula used to calculate the accuracy is shown in equation (13).

$$
\text{accuracy} = \frac{\text{True Positives + True Negatives}}{\text{Total amount of data being classified}} \\
= \frac{\sum_{i=1}^{m} f_{ii}}{\text{Total value in the diagonal of confusion matrix}} \\
= \frac{\sum_{i=1}^{m} f_{ii}}{\sum_{k=1}^{m} \sum_{i=1}^{m} f_{ki}} 
$$

(13)
3. Data
Every experiments need data. Due to the aim of this research is to identify look-alike faces, a database called Look-Alike Face database (LAF) was used. This database was created by Lamba et al. in 2011 [5]. Five hundred face images were used in this research. Those 500 face images consist of 50 different people who each have 5 images of their genuine face and 5 look-alike images.

The genuine face images are used for training dataset and look-alike face images are used for testing dataset. The images in Look-Alike Face database have variations in colour and size, meaning that they are not intentionally made in uniform surroundings. Some images are in colour form (RGB) and some are in grayscale form. All the images are modified into grayscale form for this research. The modification that has been done is only changing the colour images into grayscale images such that all of the images are in the same level of system. By doing this kind of modification first, the computational time that is required would be lower.

After all the images in the database are converted into grayscale form, the images are then transformed into row vectors. For an image with the size of $n \times n$, we will get a row vector with the size of $1 \times n^2$. This step works by reading each pixel in the image and putting the gray level as an element for the row vector. These elements are the features of the image that will be used for further steps. For this research, each image is $125 \times 125$ pixels in size, which means that it has 15625 features.

4. Experiment Results
Implementation of Fuzzy Kernel C-Means for face recognition to identify look-alike faces as proposed in this research was conducted and the results are shown in table 3 and table 4 below. Table 3 shows the result when the feature selection method which is Chi-Square was used, whereas table 4 shows the result when no feature selection method was used. The experiments were conducted on a 32GB RAM All-in-One computer with Intel® Core™ i7 processor. Different device may result in different accuracy and running time.

As seen in table 3, the experiments resulted in various accuracies ranging from 42% to 74%. The highest accuracy obtained was 74% when all of the features were used. The parameter for RBF kernel used in the experiment in table 3 is $\sigma = 0.05$. Although the running time was quite high, due to the number of features used. Based on this experiment, as the number of features increases, the accuracy also increases.

The next experiment is to use all of the features and conduct face recognition to identify look-alike faces using Fuzzy Kernel C-Means but with various different $\sigma$ as the parameter of the RBF kernel. The results are summarised in table 4. The highest accuracy is still the same as in table 3, which is 74%. Using three different parameter all resulted in 74%, but the lowest running time was the one using $\sigma = 0.05$, so that would be the best result for table 4.

| Number of Features | Accuracy (%) | Running time (seconds) |
|--------------------|--------------|------------------------|
| 1000               | 42.00        | 55.26563               |
| 2000               | 46.00        | 77.29688               |
| 3000               | 61.20        | 586.3594               |
| 4000               | 66.00        | 620.00                 |
| 5000               | 67.20        | 724.4219               |
| 6000               | 65.60        | 813.0938               |
| 7000               | 69.20        | 908.4375               |
| 8000               | 70.40        | 1071.438               |

Table 3. Results obtained using Fuzzy Kernel C-Means with Chi Square Feature Selection
Face images can actually be interpreted in so many ways. This research used only one way to interpret the face images which is by obtaining the gray level of all the pixels in the image. Other researchers may call this as interpreting the image from its texture. At the time of the experiment in this research was conducted, another research for face recognition to identify look-alike faces using images from the same database but different methods by other researches has already been executed. The comparison of accuracies obtained by different classification methods are shown in table 5. It can be seen that Fuzzy Kernel C-Means for face recognition to identify look-alike faces has produced higher accuracy than other methods from previous research stated in table 5. Researchers are continuing to develop better algorithms for many cases in face recognition including in identifying look-alike faces.

**Table 5. Comparison of accuracies**

| Research                       | Classification Method                              | Accuracy (%) |
|--------------------------------|---------------------------------------------------|--------------|
| Proposed in this research      | Fuzzy Kernel C-Means                              | 74.00        |
| Other research [6]             | Open-Set Sparse Representation                     | 53.81        |
|                                | Open-Set Collaborative Representation              | 57.21        |
|                                | Open-Set Sparse Representation and Collaborative Representation | 60.92 |

5. Conclusion

Face recognition to identify look-alike faces using Fuzzy Kernel C-Means has been proposed in this research. Fuzzy Kernel C-Means is a modification from the previous Fuzzy C-Means method with

---

**Table 4. Results obtained using Fuzzy Kernel C-Means with various σ as the parameter of the Radial Basis Function kernel**

| σ     | Accuracy (%) | Running time (seconds) |
|-------|--------------|------------------------|
| 0.0001 | 74.00        | 2717.234               |
| 0.001  | 74.00        | 2606.984               |
| 0.05   | 74.00        | 1537.234               |
| 0.1    | 73.60        | 191.5313               |
| 1      | 73.60        | 28.64063               |
| 5      | 73.60        | 28.00                  |
| 10     | 73.60        | 27.73438               |
| 50     | 73.60        | 28.65625               |
| 100    | 73.60        | 28.26563               |
| 1000   | 73.60        | 27.875                 |
kernel implementation. There are lots of different forms of kernel, but this research used Radial Basis Function (RBF) kernels. In addition, Chi-Square method is also used as a feature selection method. The data used are 500 face images obtained from the Look-Alike Face database (LAF).

Face recognition to identify look-alike faces can be useful in real-world problem where there are huge amount of data by taking the look-alike face images and eliminating unrelated face images. From the results that are obtained, the highest accuracy using Fuzzy Kernel C-Means with Radial Basis Function kernel is 74% when all of the features from the images were used. This result is obtained for this research using a particular device, so keep in mind that using different devices may result in different accuracy and running time.

References
[1] Gori M 2018 Machine learning: A constraint-based approach (Cambridge, Massachusetts: Morgan Kaufmann).
[2] Li J, Zhao B, Zhang H and Jiao J 2009 Face recognition system using SVM classifier and feature extraction by PCA and LDA combination 2009 International Conference on Computational Intelligence and Software Engineering 1-4.
[3] Shan C 2012 Learning local binary patterns for gender classification on real-world face images Pattern recognition letters 33 431-7.
[4] Rustam Z and Faradina R 2018 Face recognition to identify look-alike faces using support vector machine Journal of Physics: Conference Series 1108 012071.
[5] Lamba H, Sarkar A, Vatsa M, Singh R and Noore A 2011 Face recognition for look-alikes: A preliminary study 2011 International Joint Conference on Biometrics (IJCB) 1-6.
[6] Moine J, Faiez K, Moine H and Safai A M 2017 Open-set face recognition across look-alike faces in real-world scenarios. Image and Vision Computing 57 1-14.
[7] Wei J, Jian-qi Z and Xiang Z 2011 Face recognition method based on support vector machine and particle swarm optimization Expert Systems with Applications 38 4390-3.
[8] Sarkar S D and Goswami S 2013 Empirical study on filter based feature selection methods for text classification International Journal of Computer Applications 81 38-43.
[9] Rachburee N and Punlumjeak W 2015 A comparison of feature selection approach between greedy, IG-ratio, chi-square, and MRMR in educational mining 2015 7th International Conference on Information Technology and Electrical Engineering (ICITTEE) 420-4.
[10] Bezdek J C 1981 Pattern recognition with fuzzy objective function algorithms (New York: Plenum Press).
[11] Bishop C M 2006 Pattern Recognition and Machine Learning (New York: Springer).
[12] Liu L, Shen B and Wang X 2014 Research on kernel function of support vector machine Advanced Technologies, Embedded and Multimedia for Human-centric Computing 827-34.
[13] Wu Z D, Xie W X and Yu J P 2003 Fuzzy c-means clustering algorithm based on kernel method 5th International Conference on Computational Intelligence and Multimedia Application 49-54.
[14] Rachman A A and Rustam Z 2016 Cancer classification using Fuzzy C-Means with feature selection Mathematics, Statistics, and Their Applications (ICMSA), 2016 12th International Conference on 31-4.
[15] Rustam Z and Zahras D 2018 Comparison between support vector machine and fuzzy c-means as classifier for intrusion detection system Journal of Physics: Conference Series 1028 012227.
[16] Rustam Z and Talita A S 2015 Fuzzy kernel c-means algorithm for intrusion detection systems Journal of Theoretical and Applied Information Technology 81 161.
[17] Wulan A, Jannati M V, Rustam Z and Fauzan A A 2016 Application Kernel Modified Fuzzy C-Means for gliomatosis cerebri Mathematics, Statistics, and Their Applications (ICMSA), 2016 12th International Conference on 35-8.
[18] Rustam Z and Yaurita F 2018 Insolvency prediction in insurance companies using support vector machines and fuzzy kernel c-means *Journal of Physics: Conference Series* **1028** 012118.

[19] Miyamoto S, Ichihashi H and Honda K 2008 *Algorithms for fuzzy clustering* (Heidelberg: Springer).

[20] Graves D and Pedrycz W 2010 Kernel-based fuzzy clustering and fuzzy clustering: A comparative experimental study *Fuzzy sets and systems* **161** 522-43.

[21] Kohl M 2012 Performance measures in binary classification *International Journal of Statistics in Medical Research* **1** 79-81.