

Comparison Support Vector Machine and Fuzzy Possibilistic C-Means based on the kernel for Knee Osteoarthritis data Classification

Zuherman Rustam[#], Jacob Pandelaki[^], Dea Aulia Utami⁺, Rahmat Hidayat^{*\$}, Azizul Azhar Ramli^{*1}

[#]Department Mathematics, University of Indonesia, Kampus UI Depok, Depok, 16424, Indonesia
E-mail: rustam@ui.ac.id

[^]Department Radiology, FKUI/RSCM, Kampus UI Salemba, Jakarta, 10430, Indonesia
E-mail: jacobp@gmail.com

⁺Department Mathematics, University of Indonesia, Kampus UI Depok, Depok, 16424, Indonesia
E-mail: deaauliautami@sci.ui.ac.id

^{*}Faculty of Computer Science and Information Technology, University Tun Hussein Onn Malaysia, Malaysia
E-mail: ¹azizulr@uthm.edu.my

^{\$}Department of Information Technology, Politeknik Negeri Padang, Indonesia
E-mail: rahmat@pnp.ac.id

Abstract— Osteoarthritis is a chronic joint disease that occurs when the protective cartilage that cushions the ends of bones wears down over time and fails to be repaired. The common form of the disease is knee osteoarthritis while it can affect all body parts with joints, such as hands, ankles, hips, and spine. The major cause of knee osteoarthritis is the continuous depletion of its cartilage. During the diagnosis, machine learning is used because early prevention is necessary for proper treatment. This study, therefore, considers classification methods of Support Vector Machine (SVM) and clustering methods using fuzzy clusterings such as Fuzzy C-Means (FCM), Fuzzy Possibilistic C-Means (FPCM), and Fuzzy Possibilistic C-Means based on kernel (FPCMK) to analyze of knee osteoarthritis. SVM is a machine learning technique that works based on the principle of structural risk minimization (SRM) to obtain the best hyperplane to separate two or more classes in input space. Otherwise, the fuzzy clustering is to determine the value of a distance and to know and measure the similarity of each object to be observed. FPCMK uses the kernel Radial Base Function (RBF) in the fuzzy clustering method. The kernel function is applicable for handling non-separable data problems. This method will be compared to the level of the measured parameter; their accuracy, recall, precision, and f1 score. The greatest level of accuracy is generated from SVM with an accuracy value of 86.7%, then followed by FPCMK with an accuracy value of 85.5%.

Keywords— fuzzy C-Means clustering; fuzzy possibilistic C-Means; FPCMK; osteoarthritis knee; support vector machine.

I. INTRODUCTION

Osteoarthritis is a chronic joint disease of the cartilage that often occurs in older adults [1]. It occurs because of the inability of the joint to repair the damage itself or also known as the slow degenerative joint process [1]. Factors leading to the occurrence of osteoarthritis include overweight, getting older, joint injuries, joint cartilage that experiences disability, and certain activities that cause joints to be damaged [2]. This disease occurs in all parts of the body that have joints, such as hands, ankles, knees, hips, and spine [2] where knee osteoarthritis is the most common form [2]. Knee osteoarthritis is a disorder caused when the

cartilage continuously depletes until it finally runs out, allowing bones to rubs against themselves, causing cysts on the edge of the bone [3]. These cases are rampant in Japan, America, Vietnam, Japan, and Indonesia [3].

Therefore, early prevention of knee osteoarthritis is needed. The primary technique of prevention is the use of clustering and classification through machine learning methods for patients to receive proper and early treatment. The clustering methods considered in this paper are Fuzzy C-Means (FCM), Fuzzy Possibilistic C-Means (FPCM), and Fuzzy Possibilistic C-Means based on Kernel (FPCMK) while, the classification method is Support Vector Machine (SVM).

There are several previous studies regarding the methods of FCM, FPCM, FPCMK, and SVM. [4] Comparing fuzzy clustering algorithms for feature extraction in the vineyard showed the FCM method is the best technique based on the speed of performance compared to the PCM, FPCM, and Robust Fuzzy Possibilistic C-Means (RFCM). [5] Proposed kernel-based fuzzy and possibilistic c-means clustering. The results show that the Kernel Fuzzy C-Means (KFCM) and Kernel Possibilistic C-Means (KPCM) are resistant to the FCM and PCM methods when outlier data.

The research proposed a fuzzy c-means fuzzy swarm for fuzzy clustering problems [6]. It was found that the combination of FCM and Fuzzy Particle Swarm Optimization (FPSO) was more efficient than FCM and FPSO without being combined. [7] It provides a complete explanation of Support Vector Machines (SVM) that can be used for the classification of uncertain data. SVM uses kernel configurations to produce better results in classification. In this paper, therefore, breast cancer data is used for four types of SVM kernel methods, which include linear, polynomial, sigmoid, and radial kernels.

II. MATERIAL AND METHOD

A. Fuzzy C-Means Clustering

In 1981, Jim Bezdek introduced Fuzzy C-Means, which is a clustering data technique where each data point in a group is determined by its degree of membership. The basic concept of the FCM is to determine the center of the cluster that will mark the average location of each cluster [8]. Each data point in each cluster has a degree of membership. The degree of membership is the distance between the data points provided and the cluster center. However, in the initial conditions, the cluster center and the degree of membership are not accurate. Therefore, the center of the cluster and the degree of membership are corrected repeatedly to ensure they are in the right location [8]. The output of the FCM method is not a fuzzy inference system, but the degree of the cluster center and the degree of membership for each data.

The FCM algorithm divides available data from its finite element to generate clusters based on the given criteria. This repetitive improvement is based on the objective function given in the equation below [8]:

$$J(\mathbf{U}, \mathbf{V}, \mathbf{X}, c, m) = \sum_{i=1}^c \sum_{k=1}^N u_{ik}^m \|x_k - v_i\|^2 \quad (1)$$

With the constraint function $\sum_{i=1}^c u_{ik} = 1$ Where N is the amount of data, c is the number of clusters, \mathbf{V} is the center of the cluster, \mathbf{U} is the membership function, \mathbf{X} is the data to be clustered, m is the fuzzy ($m > 1$) and $\|x_k - v_i\|^2$ is the distance between data points with the cluster center. The value of the degree of membership in the FCM method is

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{\|x_k - v_i\|}{\|x_k - v_j\|} \right)^{2/(m-1)}}, \quad (2)$$

The updated i-cluster center is

$$v_i = \frac{\sum_{k=1}^N u_{ik}^m x_k}{\sum_{k=1}^N u_{ik}^m}, \quad i = 1, 2, \dots, c \quad (3)$$

Where the iteration termination criteria are as follows:

$$\Delta = \|\mathbf{V}^t - \mathbf{V}^{t-1}\| < \varepsilon \quad (4)$$

With \mathbf{V}^t as the center of the cluster in the t-iteration and \mathbf{V}^{t-1} as the center of the cluster in the previous iteration.

B. Fuzzy Possibilistic C-Means (FPCM)

Fuzzy Possibilistic C-Means is an algorithm development from the Fuzzy C-Means and Possibilistic C-Means method. In the FCM algorithm, the value of the u_{ik} degree of membership value is affected by all data to be clustered and all cluster centers [9]. Meanwhile, in the Possibilistic C-Means (PCM) algorithm, the typicality value τ_{ik} is affected by all data to be clustered and the cluster center to k [9]. The objective functions of FPCM are as follows [9]:

$$J(\mathbf{U}, \mathbf{V}, \mathbf{X}, c, m, n, \tau) = \sum_{i=1}^c \sum_{k=1}^n (u_{ik}^m + \tau_{ik}^\eta) d^2(x_k, v_i), \quad m, \eta > 1 \quad (5)$$

with the constraints $\sum_{i=1}^c u_{ik} = 1, \forall k \in \{1, 2, \dots, n\}$, $\sum_{k=1}^n \tau_{ik} = 1, \forall i \in \{1, 2, \dots, c\}$. Where n is the number of sample data, c is the number of clusters, m is the fuzzy degree, η is the possibilistic degree, x_k is the k data, v_i is the center value in the i -cluster, u_{ik} is member value x_k in the i -cluster, and τ_{ik} is the typicality value x_k in the i cluster.

The value of the degree of membership in the FPCM method is

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{\|x_k - v_i\|}{\|x_k - v_j\|} \right)^{1/(m-1)}} \quad (6)$$

The value of the typicality of the FPCM method is

$$\tau_{ik} = \frac{\left(\frac{1}{\|x_k - v_i\|} \right)^{1/(\eta-1)}}{\sum_{l=1}^n \left(\frac{1}{\|x_l - v_i\|} \right)^{1/(\eta-1)}} \quad (7)$$

The updated i-cluster center is

$$v_i = \frac{\sum_{k=1}^N (u_{ik}^m + \tau_{ik}^\eta) x_k}{\sum_{k=1}^N (u_{ik}^m + \tau_{ik}^\eta)} \quad (8)$$

With the iteration termination criteria as follows:

$$\Delta = \|\mathbf{V}^t - \mathbf{V}^{t-1}\| < \varepsilon \quad (9)$$

where \mathbf{V}^t is the center of the cluster in the t-iteration and \mathbf{V}^{t-1} is the center of the cluster in the previous iteration.

C. Fuzzy Possibilistic C-Means Based on the Kernel (FPCMK)

Fuzzy Possibilistic C-Means based on the Kernel is a generalized algorithm from the Fuzzy Possibilistic C-Means method. FPCMK uses the kernel Radial Base Function (RBF) function in the FPCM method [10]. The kernel function handles problems that are linear to be applied in handling nonlinear problems by using the function ϕ which is a non-linear mapping from the input space to the feature space [10].

Defined kernel functions:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) \quad (10)$$

The kernel distance is defined as follows:

$$\begin{aligned} d^2(\mathbf{x}_i, \mathbf{x}_j) &= \|\phi(\mathbf{x}_i) - \phi(\mathbf{x}_j)\|^2 \\ &= \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_i) - 2\phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) + \phi(\mathbf{x}_j)^T \phi(\mathbf{x}_j) \\ &= K(\mathbf{x}_i, \mathbf{x}_i) - 2K(\mathbf{x}_i, \mathbf{x}_j) + K(\mathbf{x}_j, \mathbf{x}_j) \end{aligned}$$

With $K(x_k, v_i)$ is the Radial Base Function kernel function shown below.

$$K(x_k, v_j) = \exp\left(-\frac{\|x_k - v_j\|^2}{\sigma^2}\right) \quad (11)$$

The objective function of FPCMK is as follows [10]:

$$J(U, V) = 2 \sum_{i=1}^c \sum_{k=1}^n (u_{ik}^m + \tau_{ik}^\eta) (1 - K(x_k, v_i)), \quad m, \eta > 1 \quad (12)$$

with the constraints $\sum_{i=1}^c u_{ik} = 1, \forall k \in \{1, 2, \dots, n\}$, $\sum_{k=1}^n \tau_{ik} = 1, \forall i \in \{1, 2, \dots, c\}$. Where n is the number of sample data, c is the number of clusters, m is the fuzzy degree, η is the possibilistic degree, x_k is the k data, v_i is the center value in the i -cluster, u_{ik} is member value x_k in the i -cluster, and τ_{ik} is the typicality value x_k in the i cluster.

The value of the degree of membership in the FPCMK method is as presented below.

$$u_{ik} = \frac{\frac{1}{(2-2K(x_k, v_i))} \frac{1}{(m-1)}}{\sum_{j=1}^c \frac{1}{(2-2K(x_k, v_j))} \frac{1}{(m-1)}} \quad (13)$$

The value of the typicality of the FPCMK method is

$$\tau_{ik} = \frac{\frac{1}{(2-2K(x_k, v_i))} \frac{1}{(\eta-1)}}{\sum_{l=1}^n \frac{1}{(2-2K(x_l, v_i))} \frac{1}{(\eta-1)}} \quad (14)$$

The updated i -cluster center is

$$v_i = \frac{\sum_{j=1}^n u_{ik}^m K(x_k, v_i) x_k}{\sum_{j=1}^n u_{ik}^m K(x_k, v_i)} \quad (15)$$

Where the iteration termination criteria is as follows:

$$\Delta = \|\mathbf{V}^t - \mathbf{V}^{t-1}\| < \varepsilon \quad (16)$$

With \mathbf{V}^t is the center of the cluster in the t -iteration and \mathbf{V}^{t-1} is the center of the cluster in the previous iteration.

D. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a machine learning technique that was first introduced by Vapnik in 1992. SVM works based on the principle of structural risk minimization (SRM) to obtain the best hyperplane that can separate two or many classes in input space [11–13]. The best hyperplane as a separator of two classes can be found by calculating the margin of the hyperplane and looking for the maximum point of the margin [11–13]. Margin is the distance between the hyperplane and the closest data from each class. The data closest to the hyperplane are called the support vector [11–13].

Suppose a dataset is denoted as $\mathbf{x}_i \in R^n$ which has a label denoted as $y_i \in \{-1, +1\}$ for $i = 1, 2, \dots, n$ where n is the amount of data. The formula for the hyperplane is between class -1 and +1 if it is assumed that the two classes are completely separated by a dimension n hyperplane as follows

$$H_1 = \mathbf{w}^T \mathbf{x}_i + b = 0 \quad (17)$$

Where H_1 represents the hyperplane, \mathbf{w} is the normal field and b is the optimal hyperplane bias. The optimal hyperplane function is obtained by looking for the weight parameter \mathbf{w} and the bias b parameter in the function below which is known as the decision function [12, 13].

$$f(x) = \mathbf{w}^T \mathbf{x}_i + b \quad (18)$$

In getting the best hyperplane, a hyperplane will be found that has the largest margin value. The biggest margin can be found by maximizing the value of the distance between the hyperplane and the closest point of each class [12]. Therefore, forming:

$$\max |d_+ - d_-| = \max \left(\left| \frac{(1-b)}{\|\mathbf{w}\|} - \frac{(-1-b)}{\|\mathbf{w}\|} \right| \right) = \max \left(\frac{2}{\|\mathbf{w}\|} \right) \quad (19)$$

with condition

$$\max(\mathbf{w}^T \mathbf{x}_i + b) = -1, \text{ for } y_i = -1, \quad i = 1, 2, \dots, n \quad (20)$$

$$\min(\mathbf{w}^T \mathbf{x}_i + b) = 1, \text{ for } y_i = +1, \quad i = 1, 2, \dots, n \quad (21)$$

The problem of maximizing the margin $|d_+ - d_-|$ equivalent to minimizing the value of $\|\mathbf{w}\|$, hence written as a primal optimization problem as below [11–13]

$$\min \frac{1}{2} \|\mathbf{w}\|^2 \quad (22)$$

Subject to

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad i = 1, 2, \dots, n \quad (23)$$

III. RESULTS AND DISCUSSION

A. Data

The study used 41 data on knee osteoarthritis patients who performed examinations at Dr. Cipto Mangunkusumo (RSCM)'s Hospital. The data was divided into three classes based on the grade of osteoarthritis. Based on Kallgren and Lawrence, there are five gradings to assess the severity of knee osteoarthritis, as shown in Table I below.

TABLE I
DEFINITION OF GRADE

Grade	Definition of Grade
0	No picture of osteoarthritis appears radiologically.
1	Suspect, there is narrowing in the joint gap and possibly osteophytes.
2	Has definite osteophytes and there may be a narrowing of the joint gap on the side of weight-bearing.
3	Multiple osteophytes, narrowing of the definite joint gap, sclerosis, and possible bone deformity.
4	Large osteophytes, narrowing of the joint gap, severe sclerosis, and definitive bone deformity.

Based on the grading, data on knee osteoarthritis patients are divided into 3 classes, which includes class 0 which is the class for patients with grade 1 with 22 patients, class 1 which is the class for patients with grade 2 with 81 patients, and class 2 which is the class for patients with grade 3 with 38 patients.

This data has four features that influence the severity of knee osteoarthritis, and its features are explained in table II below.

TABLE II
DEFINITION OF FEATURE

Name of Feature	Definition of Grade
X1	The average thickness of cartilage in the medial femur
X2	The average thickness of cartilage in the lateral femur
X3	The average thickness of cartilage in the medial tibia
X4	The average thickness of cartilage in the lateral tibia

B. Results and Analysis

In this study, data on knee osteoarthritis patients used were divided into two types; there are 70% training data and 30% testing data. Data validation is used with 5-fold cross-validation to divide the dataset into five parts. In evaluating

the performance of each method, parameters of accuracy, recall, precision, and f1 score will be used where the greater the values, the better the method in clustering and classification. Table III shows the formula for calculating these values.

TABLE III
THE FORMULA OF EACH PARAMETER

Name of Parameters	Formula
Accuracy	$\frac{(TN + TP)}{(FN + TP + FP + TN)} \times 100\%$
Recall	$\frac{TP}{(FN + TP)} \times 100\%$
Precision	$\frac{TP}{(FP + TP)} \times 100\%$
F1 Score	$2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\%$

This study will compare FCM, FPCM, FPCMK, and SVM in the classification of knee osteoarthritis. Table IV, V, VI, VII and Figure 1 show the performance evaluation of each method based on the parameters of accuracy, recall, precision, and f1 score.

TABLE IV
ACCURACY FOR EACH METHOD

Method	Accuracy
FCM	79.5%
FPCM	81.9%
FPCMK	85.5%
SVM	86.7%

From Table IV, the best value of accuracy is using the SVM method. The accuracy of the SVM method is 86.7% and the smallest value of accuracy is 79.5% using the FCM method.

TABLE V
RECALL FOR EACH METHOD

Method	Recall
FCM	76.7%
FPCM	85.3%
FPCMK	88.1%
SVM	90.7%

The best value of Recall from Table V is 90.7% using the SVM method. The smallest value of Recall is using the FCM method with 76.7% value.

TABLE VI
PRECISION FOR EACH METHOD

Method	Precision
FCM	82.5%
FPCM	74.4%
FPCMK	84.1%
SVM	84.8%

From Table VI, the best value of precision is using the SVM method. The precision of the SVM method is 84.8% and the smallest value of precision is 74.4% using the FPCM method.

TABLE VII
F1-SCORE FOR EACH METHOD

Method	Accuracy
FCM	79.5%
FPCM	79.4%
FPCMK	86.05%
SVM	87.65%

The best value of f1-score from Table VII is 87.65% using SVM method. The smallest value of f1-score is using the FPCM method with 79.4% value.

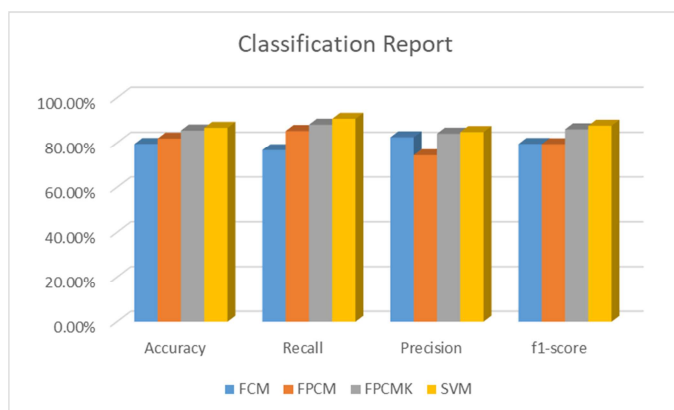


Fig. 1 Classification report for each method

From Table IV, V, VI, VII, and Figure 1, the SVM is better than FCM, FPCM and FPCMK methods based on the value of accuracy, recall, precision, and f1 score, at 86.7%, 90.7%, 84.8%, and 87.65% respectively. The FPCMK is the best compared to FCM and FPCM based on the value of accuracy, recall, precision, and f1 score, at 85.5%, 88.1%, 84.1%, and 86.05% respectively.

IV. CONCLUSIONS

This study uses classification methods such as Support Vector Machine and fuzzy clustering, including Fuzzy C-Means (FCM), Fuzzy Possibilistic C-Means (FPCM) and

Fuzzy Possibilistic C-Means based on the Kernel (FPCMK) in the early classification of knee osteoarthritis patients to obtain the best treatment. The comparison was made for the four methods, and SVM became the best for the classification of knee osteoarthritis. For the fuzzy clustering method, it is preferable to add the RBF kernel function to the FPCM. Measurement of the performance of these methods uses the following parameters, such as accuracy, recall, precision, and f1 score.

ACKNOWLEDGMENT

This research was financially supported by the Indonesian Ministry of Research and Higher Education, with a PDUPT 2019 research grant scheme (ID number NKB-1621/UN2.R3.1/HKP.05.00/2019). The authors are grateful to the Department of Radiology, Cipto Mangunkusumo Hospital, for their kindness in providing the Osteoarthritis knee datasets. All reviewers involved in the improvement of this article are also appreciated

REFERENCES

- [1] Gale Encyclopedia of Medicine, (2008), Key facts retrieved from <http://medical.dictionnaire.thefreedictionary.com/osteoarthritis> on 22 August 2018.
- [2] Lawrence RC, Felson DT, Helmick CG, Arnold LM, Choi H, Deyo RA, et al. *Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part II. Arthritis Rheum.* (2008), 58(1), 26-35.
- [3] Arden N, Blanco F, Cooper C, Guermazi A, Hayashi D, Hunter D, et al. Introduction: Historical and Current Perspective on Osteoarthritis. *Atlas of Osteoarthritis* Springer Healthcare Ltd.; 2015. p.10-9
- [4] S.R.Kannan, R.Devi, S.Ramathilagam, T.P.Hong, A.Ravikumar, "Robust fuzzy clustering algorithms in analyzing high dimensional cancer databases", 2015, *Journal Applied Soft Computing* vol.35, pp. 199-213.
- [5] N.R.Pal, K.Pal, J.M.Keller, J.C.Bezdek, "A Possibilistic Fuzzy c-Means Clustering Algorithm", 2005, *IEEE Transactions on Fuzzy Systems*, vol.13, No.4.
- [6] C.Zhang, Y.Zhou, T.Martin, "Similarity-Based Fuzzy and Possibilistic c-means Algorithm", 2008, *Proceedings of the 11th Joint Conference on Information Sciences*.
- [7] K.J.Dsouza, Z.A.Ansari, "Experimental Exploration of Support Vector Machine for Cancer Cell Classification", 2017, *IEEE International Conference on Cloud Computing in Emerging Markets*.
- [8] Z. Rustam and D. Zahras. *Comparison between Support Vector Machine and Fuzzy C-means as a classifier for the intrusion detection system*. 2nd International Conference on Statistics, Mathematics, Teaching, and Research. (2018).
- [9] Pal, N. R., Kal, P., dan Bezdek, J. C. 1997. A Mixed C Means Clustering Model. *IEEE*, 11- 21.
- [10] Z. Rustam and F. Yaurita. *Insolvency prediction in insurance companies using Support Vector Machines and Fuzzy Kernel C-Means*. 2nd International Conference on Statistics, Mathematics, Teaching, and Research. (2018).
- [11] Z. Rustam and N. P. A. A. Ariantari. *Support Vector Machines for Classifying Policyholders Satisfactorily in Automobile Insurance*. 2nd International Conference on Statistics, Mathematics, Teaching, and Research. (2018).
- [12] Wang, L. (2005). *Support Vector Machines: Theory and Applications*. New York: Springer, 177.
- [13] Xue, H., Yang, Q., & Chen, S. (2009). *SVM: Support Vector Machines*. United States of America: Taylor & Francis Group.
- [14] Wu Xindong and Kumar Vipin. *The Top Ten Algorithms in Data Mining* (Taylor and Francis Group, LLC, 2009), 127-150.