Optimization of Fuzzy Support Vector Machine (FSVM) Model in Multiple Metric Spaces

Dita Fadma Ristianti
Master of Mathematics Education
Ahmad Dahlan University
Yogyakarta, Indonesia
Dita1907050001@webmail.uad.ac.id

Sugiyrto Surono
Dept. Mathematic FAST
Ahmad Dahlan University
Yogyakarta, Indonesia
sugiyarto@math.uad.ac.id

Joko Eliyanto
Master of Mathematics Education
Ahmad Dahlan University
Yogyakarta, Indonesia
joko1907050003@webmail.uad.ac.id

2021 International Conference on Artificial Intelligence and Big Data Analytics (ICAIBDA) | 978-1-6654-0890-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICAIBDA53487.2021.9689703

***Abstract*— Fuzzy membership function was introduced into the Support Vector Machine (SVM) resulting in modifications. Selecting the correct membership function is an important step in the Fuzzy Support Vector Machine (FSVM) method. One of the general criteria for selecting fuzzy membership is determined by the distance between a point and its fixed center category. This study aims to develop the SVM method into Fuzzy SVM (FSVM) with several distance functions that are applied to the Early Stage Diabetes data which collects 520 data. The distance functions used include Euclid, Canberra distance, Minkowski distance, Chebyshev distance, Minkowski Chebyshev distance, and Bray-Curtis distance where this distance function is used to determine the best distance that can be seen from the results of accuracy, specificity, g-means which is best for viewing diabetes risk. The results of this comparison show that the FSVM method with several distance functions is more than the SVM method. Where the FSVM method at the Canberra distance with a penalty value of $C = 2^5$ is the best distance to see the risk of diabetes, based on the results of specificity = 100%, g-means = 86.91%, and accuracy = 85.26% is superior to the SVM method at the penalty value $C = 2^{10}$ with specificity = 69.36%, g-means = 77.31%, and accuracy = 79.49%. Although the FSVM method produces an evaluation value at sensitivity = 75.53%, it is lower than the SVM method with a sensitivity value = 86.17%.**

Keywords—Support Vector Machine (SVM), Fuzzy Support Vector Machine (FSVM), Membership Function, Metric

I. Introduction

The last ten years, machine learning methods have been developed to aid the classification without being bound by the assumptions, and to provide greater flexibility in data analysis, but still have the accuracy and ease of use are high. Machine learning methods that have been developed one Support Vector Machine (SVM) [1]. Vapnik said, [2] defined the Support Vector Machine (SVM) method as a new machine learning method. The SVM method finds an optimal global solution, by mapping the training data to a high-dimensional space, then in a high-dimensional space it will look for a classification that maximizes the margin between the two data classes [3]. The concept of SVM is an effort to find the best hyperplane, which is used as a separator between the two classes at the input[4]. SVM is one of the featured methods of machine learning because it has good performance in completing the classification and predict cases. The principle of SVM is to find the optimal classification model or set of separators from the classification data trained by the algorithm to divide the data set into two or more different classes. These classes can help predict classes based on new data [5].

However, in the application of SVM there are many distractions that could make the data sample is not ideal. Therefore, the Fuzzy membership functions are introduced into the SVM. FSVM is very effective in many real-time applications such as credit risk evaluation, text categorization and others [6] [7] [8] [9] [10]. The facts prove that FSVM is better than SVM in dealing with noise and can effectively eliminate the influence of noise on SVM [11]. The main problem in the FSVM model is the creation of appropriate memberships to minimize outlier effect data points.[12], [13], [14] and [15] selecting the correct membership function is an important step in the FSVM method. One common criterion for selecting Fuzzy membership is determined by the distance between the point and the central category and equipment[11],[16]. "Euclidean" distance is a common metric for FSVM. As an alternative method, several distance functions are proposed to measure the distance from each point to the center of the class, this distance function will be used to determine the best point.

Utilization data mining is not limited to science and technology, but in the world of healthcare data mining is often used to treat the buildup of medical data. SVM method can be used as a reference to predict and diagnose a particular type of disease using methods that can be applied. Diabetes is a disease in the form of a metabolic disorder characterized by blood sugar levels that exceed normal limits [17] which occurs because the pancreas does not produce enough insulin (a hormone that regulates blood sugar or glucose), or when the body cannot effectively use the insulin it produces[18]. Diabetes is not an infectious disease, but WHO data shows that the percentage of non-communicable diseases in 2004 which reached 48.30% was greater than the number of presentations of infectious diseases, which was 47.50%. Even non-communicable diseases are the number one cause of death in the world (63.50%) (Islam, Ferdousi, Rahman, & Bushra, 2020). (Garnita, Society, Studies, Society, & Indonesia, 2012). Many people with diabetes are not aware of the disease, especially, because of the lack of information in the community about diabetes symptoms. Symptoms of early characteristics of people with diabetes are often referred to as triaspoli (polyuria, polydipsia, and polyphagia). This study aims to develop the SVM method into Fuzzy SVM (FSVM) with several distance functions applied to Early Stage Diabetes data which collected 520 data from Sylhet Diabetes Hospital, Sylhet Bangladesh. The distance functions used include Euclid, Canberra distance, Minkowski distance, Chebyshev distance, Minkowski Chebyshev distance, and Bray-Curtis distance where this distance function is used to determine the best distance that can be seen from the results of accuracy, specificity, g-means which is best for viewing diabetes risk. This study also tried to experiment with developing the SVM method into FSVM using various distances. This is one of the

978-1-6654-0890-5/21/$31.00 ©2021 IEEE

novelty elements offered in this study compared to other studies. The results of the proposed method will compare the SVM method with Fuzzy SVM with several distance functions.

II. METHODS

A. Support Vector Machine (SVM)

Support vector machines (SVM) is a supervised learning method, first introduced by Vapnik in 1995 together with Bernhard Boser and Isabele Guyon [19]. [20] [6] Support Vector Machine (SVM) is a classification method that works by finding a hyperplane with optimum margins. Hyperplanea is a data dividing line between classes. Margin (m) is the distance between the hyperplane and the closest data in each class. The hyperplane can be represented as $\mathbf{w}^T x_i - b = 0$. Where x_i is the data, $y_i \in \{-1,+1\}$ is the class label of x_i, **w** is the weight vector of size *(px1)*, and **b** is the position of the plane relative to the center of the coordinates or better known as bias scalar value. The formula for the SVM optimization problem for linear classification is

$$\min \frac{1}{2}\|\mathbf{w}\|^2 + C\left[\sum_{i=1}^{n}\xi_i\right] \quad (1)$$

by combining the two functions separator for both classes, then it can be represented in the inequality as follows:

$$y_i\left(x_i^T\mathbf{w}+b\right)-1-\xi_i \geq 0$$
$$y_i\left(x_i^T\mathbf{w}+b\right) \geq 1-\xi_i$$

ξ is a slack variable ξ has been added to the model for classifying data that can not be separated linearly. Where C is the major parameters that determine the penalty due to errors in classification (misclassification) data.

To determine the optimal hyperplane above it is possible to change the shape of the primal into shape Quadratic Programming (QP). Thus the optimization problem can be solved by the Karush-Kuhn-Tucker (of the summit) and formulated into a formula lagrange

$$L = \frac{1}{2}\mathbf{w}^2 + \mathrm{C}\sum_{i=1}^{\mathrm{n}}\xi_i - \sum_{i=1}^{\mathrm{n}}\alpha_i\left[y_i\left(x_i^T\mathbf{w}+b\right)-1+\xi_i\right] - \sum_{i=1}^{\mathrm{n}}\mu_i\xi_i \quad (2)$$

where α_i dan μ_i are *Lagrange Multiplier*. By minimizing *L* with $\mathbf{w}$, b, and ξ,

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^{\mathrm{n}}\alpha_i y_i x_i = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^{n}\alpha_i y_i x_i$$
$$\frac{\partial L}{\partial b} = \sum_{i=1}^{\mathrm{n}}\alpha_i y_i x_i = 0$$
$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \mu_{\mathrm{i}} = 0 \Rightarrow \alpha_i - \mu_{\mathrm{i}} = \mathrm{C}$$

with $\xi_i \geq 0, \alpha_i \geq 0, \mu_i \geq 0, \ \alpha_i\left[y_i\left(x_i^T\mathbf{w}+b\right)-1+\xi_i\right]=0$, $\mu_i\xi_i = 0$ Thus obtained the dual problem

$$\max_{\alpha} \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n}\alpha_i\alpha_j y_i y_j x_i^T x_j \quad (3)$$

where $\alpha = (\alpha_i, \ldots, \alpha_n)$ is a non-negative Lagrange multiplier vector. By completing the above quadratic optimization α_i so that obtained $w = \sum_{i=1}^{n}\alpha_i y_i x_i$. Based on KKT conditions, is term bias

$$b = y_i - \sum_{i=1}^{n}\alpha_i y_i x_i \quad (4)$$

can also be computed for any supporting vector (observation that the corresponding α_i is greater than zero). The sample point x_i is classified based on the sign of its classification function as follows,

$$f(x) = sign\left(\mathbf{w}^{\mathrm{T}}(x_i)+b\right) \quad (5)$$

For the non-linear separable in feature space, kernel function $K(x_i, x_j) = \Phi(x_i)^T\,\Phi(x_j)$ is used to find hyperplane in a higher dimensional space, where $\Phi(x_i)$ is a non-linear mapping function.

B. Fuzzy Support Vector Machine (FSVM)

In the classification of soft intervals, the value of parameter C should not be too large or too small to ensure the effect of the classifier [11]. Training given S, where dimana $S = \{(\mathrm{x}_i, y_i, s_i)\}_{i=1}^{N}$, x_i is a sample of size n, $y_i \in \{+1,-1\}$ stating grade (+1 for positive class and -1 for negative class), and s_i is the fuzzy membership. So, the objective function is written as follows,

$$\min \frac{1}{2}\|\mathbf{w}\|^2 + C\left[\sum_{i=1}^{n}s_i\xi_i\right] \quad (6)$$

by combining the two functions separator for both classes, then it can be represented in the inequality as follows:

$$y_i\left(\mathbf{\Phi}x_i^T\mathbf{w}+b\right)-1-\xi_i \geq 0$$
$$y_i\left(\mathbf{\Phi}x_i^T\mathbf{w}+b\right) \geq 1-\xi_i$$

where **w** is the vector weighting on local decisions, *b* stated bias, $\mathbf{\Phi}x_i$ a nonlinear function that maps x_i into space features high dimensional in which areas a better decision can be found, C is a regularization parameter chosen beforehand to control the trade-off between margins classification and misclassification costs. Non-negatif variables ξ_i is slack variable states of x_i on SVM, while $s_i\xi_i$ is a error size with different weights according to s_i.

To solve quadratic optimization, the Lagrange Equation is as follows,

$$L=\frac{1}{2}\mathbf{w}^2+C\sum_{i=1}^{n}s_i\xi_i-\sum_{i=1}^{n}\alpha_i\left[y_i\left(x_i^T\mathbf{w}+b\right)-1+\xi_i\right]-\sum_{i=1}^{n}\mu_i\xi_i \tag{7}$$

where α_i and μ_i are *Lagrange Multiplier*. By minimizing L with to $\mathbf{w}, b$, and $s_i\xi_i$:

$$\frac{\partial L}{\partial \mathbf{w}}=\mathbf{w}-\sum_{i=1}^{n}\alpha_i y_i x_i=0\rightarrow \mathbf{w}=\sum_{i=1}^{n}\alpha_i y_i x_i$$

$$\frac{\partial L}{\partial b}=\sum_{i=1}^{n}\alpha_i y_i=0$$

$$\frac{\partial L}{\partial \xi_i}=s_iC-\alpha_i-\mu_i=0\rightarrow \alpha_i-\mu_i=s_iC$$

$$L=\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n}\alpha_i\alpha_i y_i y_i x_i^T x_i \tag{8}$$

C. Fuzzy Membership Function u for FSVM

Ding Xiaokang [11] explains that FSVM models adopting the conventional method of calculating membership, which determines the class centers by averaging all of the samples. By using the distance from each sample point to the center of the class as d_i, then the membership function can be expressed as:

$$s_i=\begin{cases}1-\dfrac{d_{i+}}{r_{i+}+\beta}, y_i=+1\\[2ex] 1-\dfrac{d_{i-}}{r_{i-}+\beta}, y_i=-1\end{cases} \tag{9}$$

$$s_i=\begin{cases}1-\dfrac{\left\|x_i^+-x_{cen}^+\right\|}{max\left(\left\|x_i^+-x_{cen}^+\right\|\right)+\beta}\rightarrow y_i=+1\\[2ex] 1-\dfrac{\left\|x_i^+-x_{cen}^+\right\|}{max\left(\left\|x_i^+-x_{cen}^+\right\|\right)+\beta}\rightarrow y_i=-1\end{cases} \tag{10}$$

Where δ is positive value used to avoid s to zero, while d represents the Euclidean distance from each sample to the class center.

β = constant to avoid $s_i=0$

$d_{i+}=\left\|x_i^+-x_{cen}^+\right\|$

$d_{i-}=\left\|x_i^--x_{cen}^-\right\|$

$r_{i+}=\max d_{i+}$

$r_{i-}=\max d_{i-}$

x_{cen}^+ = positive sample center

x_{cen}^- = negative sample center

x_i^+ = labelled sample $y_i=1$

x_i^- = labelled sample $y_i=-1$

This function indicates that the closer to the center of the class, the greater the value of membership, and the smaller the contrary.

D. Metric

1. Minkowski Distance

The Minkowski distance is a generalization of the distance matrix, defined as follows:

$$d_{\min}(x,y)=\left(\sum_{i=1}^{d}\left|x_i-y_i\right|^r\right)^{\frac{i}{r}}, r\geq 1 \tag{11}$$

where r is a Minkowski parameter, at Euclidean (r = 2) and Manhattan (r = 1) distances. Metric conditions are met as long as p is equal to or greater than 1[21].

2. Chebyshev Distance

The Chebyshev distance is the variance of the Minkowski distance where,

$$\begin{aligned}&p\rightarrow\infty\\ &d_{cbc}(x,y)=\max_{k=1}^{n}\left|x_i-y_i\right|\end{aligned} \tag{12}$$

where x_i and y_i are nilai the values of x and y in dimension n [21]

3. Canbera Distance

The Canbera distance is given as follows:

$$d(x,y)=\sum_{i=1}^{n}\frac{|x_i-y_i|}{|x_i|+|y_i|} \quad (13)$$

Canberra distances can perform very well, significantly better than the most used Manhattan and Euclidean distances, as shown [22] This distance tests the sum of the series of fractional differences between the coordinates of a pair of vectors [23].

4. Minkowski Chebyshev Distance

Rodriguez [24] brings up a new distance, namely the combination of the Minkowski and Chebyshev distances. The combination of the Minkowski and Chebyshev distances is shown in the following definition:

$$d_{(w_1,w_2,d)}\left(\bar{x},\bar{y}\right)=w_1 d_{mkw}\left(\bar{x},\bar{y}\right)+w_2 d_{cheb}\left(\bar{x},\bar{y}\right) \quad (14)$$

Or

$$d_{(w_1,w_2,r)}\left(\bar{x},\bar{y}\right)=w_1\left(\sum_{i=1}^{d}|x_i-y_i|^r\right)^{\frac{i}{r}}-\max_{k=1}^{n}|x_i-y_i|, 1\le k\le n \quad (15)$$

where x_i and y_i are the value to $-i$ on two vectors $\bar{x}$ and $\bar{y}$, and vice versa on the dimension n

5. Bray-Curtis Distance

The Bray-Curtis distance, sometimes also called the Sorensen distance, is commonly used in ecology and environmental sciences. This distance view space as a lattice that is similar to the distance of a city block. The Bray-Curtis distance has the nice property that if all coordinates are positive, the value is between zero and one. If both objects are at zero coordinates, the Bray-Curtis distance is not specified. [23]

$$d(x,y)=\sum_{i=1}^{n}\frac{|x_i-y_i|}{x_i+y_i} \quad (16)$$

where,

$d =$ distance between x and y

$x =$ cluster center data

$y =$ data on attributes

E. Classification Evaluation

The actual data and the predicted data from the classification model are presented using a cross tabulation (Confusion matrix), which contains information about the actual data class represented in the row matrix and the predicted data class in the column[19].

Accuracy is an evaluation matrix that is very important to assess the performance of an overall classification results. The higher the classification accuracy of classification techniques also means that the performance is getting better. [25] explained that the evaluation of the performance of a classifier in the imbalance class can be measured using the G-mean. Sensitivity is a performance measure to measure the positive class or the accuracy of the positive class. Specificity is a performance measure to measure the negative class or the accuracy of the negative class.

Table 1. Confusion Matrix

Actual	**Predicted**	
	Positive	**Negative**
Positive	TP	FN
Negative	FP	TN

Information:

TP : *True Positive* (the number of correct predictions in the positive class)

FP : *False Positive* (the number of wrong predictions in the positive class)

FN : *False Negative* (the number of incorrect predictions in the negative class)

TN : *True Negative* (the number of correct predictions in the negative class)

Accuracy

Accuracy assesses the overall effectiveness of the algorithm by estimating the correct value of the class label. The Accuracy Value is stated as follows

$$Accuracy=\frac{TN+TP}{TN+TP+FN+FP} \quad (17)$$

Sensitivity (SE)

Sensitivity is a performance measure to measure the positive class or the accuracy of the positive class. The sensitivity value states how many positive class samples are correctly labeled. The sensitivity value is stated as follows.

$$Sensivity=\frac{TP}{TP+FN} \quad (18)$$

Specificity (SP)

Specificity is a performance measure to measure the negative class or the accuracy of the negative class. The specificity value states how many samples of the negative class are correctly labeled. The specificity value is stated as follows.

$$Specifity = \frac{TN}{TN+TP} \qquad (19)$$

G-means (GM)

[26] said that the g-mean value was used to evaluate the performance of the algorithm on imbalanced data problems. G-means is the product of the prediction accuracy for both classes which includes accuracy in the positive class (sensitivity) and accuracy in the negative class (specificity). This value shows the balance between the classification performance of the majority and minority classes. poor performance in positive sample prediction will result in a low G-means value as well as for the negative class. The g-means value is expressed as follows.

$$g-mean = \sqrt{Sensivity\ x\ Specifity} \qquad (20)$$

III. RESULT AND DISCUSSION

In this study the type of data used is secondary data obtained from the official page through https://archive.ics.uci.edu/ml/datasets.php. Data collected in the article amounted to 520 using questionnaires data taken directly from the patient's Hospital ethical standards institutions in which research is conducted and ethical approval was obtained from the Hospital Diabetes Sylhet, Bangladesh Sylhet. The factors that influence the risk of diabetes are 16 as the x_i variable and the y variable as the class label of the x_i variable with members {1,-1}, where 1 is for the class that is not at risk of developing diabetes and -1 for the class that is at risk of developing the disease diabetes. The steps in conducting the analysis in this study are as follows

a) Exploration to see the characteristics of the data.

b) Divide the data into training and testing data.

c) SVM classification on the training data and evaluate the classification performance on the test data.

d) FSVM classify the training data using Euclidean metrics, Canberra, Minkowski, Chebyshev, Minkowski, Chebyshev, and Bray-Cutris and evaluate the classification performance on the test data.

- Calculates Euclid, Canberra, Minkowski, Chebyshev, Minkowski-Chebyshev, and Bray-Cutris matrices from data points to class center.
- Calculate the value of membership function

e) Comparing the performance of SVM and FSVM classification with several matrix models to see the best classification results.

Before the SVM modeling data is divided into training and testing. In this study, the data used amounted to 520 cases divided into training data of 70%, namely 364 cases and testing data of 30%, namely 156 cases. This SVM method uses a polynomial kernel with different C penalty values to see the best accuracy results.

Table 2. Classification

MODEL	C	SE	SP	GM	Accuracy
SVM	2^1	84,04%	67,74%	75,45%	77,56%
	2^5	74,47%	62,90%	68,44%	69,87%
	$\mathbf{2^{10}}$	**86,17%**	**69,36%**	**77,31%**	**79,49%**
FSVM-1	2^1	72,34%	100%	85,68%	83,97%
	2^5	72,34%	100%	85,05%	83,33%
	2^{10}	71,28%	100%	84,43%	82,69%
FSVM-2	2^1	74,47%	100%	86,30%	84,62%
	$\mathbf{2^5}$	**75,53%**	**100%**	**86,91%**	**85,26%**
	2^{10}	71,28%	100%	84,43%	82,69%
FSVM-3	2^1	73,40%	100%	85,68%	83,97%
	2^5	72,34%	100%	85,05%	83,33%
	2^{10}	71,28%	100%	84,43%	82,69%
FSVM-4	2^1	72,34%	100%	85,05%	83,33%
	2^5	71,28%	100%	84,43%	82,69%
	2^{10}	71,28%	100%	84,43%	82,69%
FSVM-5	2^1	72,34%	100%	85,05%	83,33%
	2^5	72,34%	100%	85,05%	83,33%
	2^{10}	71,28%	100%	84,43%	82,69%
FSVM-6	2^1	72,34%	100%	85,05%	83,33%
	2^5	72,34%	100%	85,05%	83,33%
	2^{10}	71,28%	100%	84,43%	82,69%

Table 3. Classification Performance The results of the SVM classification performance at different C penalty values resulted in the values of sensitivity, specificity, G-means, and accuracy. On the SVM classification, it can be seen that the best classification performance evaluation is given by a penalty value of $C = 2^{10}$ with an evaluation value of sensitivity 86.170%, specifity 69.355%, G-means 77.307% and accuracy 79.487%. FSVM Classification at the Euclidean Distance (FSVM-1), the best classification performance evaluation results are given by a penalty value of $C = 2^1$ with the same evaluation values, namely sensitivity 72.340%, specifity 100%, G-means 85.676% and accuracy 83.974. FSVM Classification at the Canberra Distance (FSVM-2), it can be seen that the best classification performance evaluation results are given by a penalty value of $C = 2^5$ with an evaluation value of 75.532% sensitivity, 100% specificity, 86.909% G-means and 85.256% accuracy. FSVM Classification at the Minkowski Distance (FSVM-3), it can be seen that the best classifi ation performance evaluation results are given by a penalty value of $C = 2^1$ with an evaluation value of 73.404% sensitivity, 100% specificity, 85.676% G-means and 83.974% accuracy. Furthermore, the results of the evaluation of the FSVM classification at the Chebyshev distance will c be given.

FSVM Classification at the Chebyshev Distance (FSVM-4), it can be seen that the best classification performance evaluation results are given by a penalty value of $C = 2^1$ with evaluation values of 72.340% sensitivity, 100% specificity, 85.053% G-means and 83.333% accuracy. Furthermore, the results of the evaluation of the FSVM classification at the Minkowski-Chebyshev distance will be given. FSVM Classification at the Minkowski-Chebyshev Distance (FSVM-5), it can be seen that the best classification performance evaluation results are given by a penalty value of $C = 2^1$ with an evaluation value of 72.340% sensitivity, 100% specificity, 85.053% G-means and 83.333% accuracy. Furthermore, the results of the evaluation of the FSVM classification at the Bray-Curtis distance will be given. Similarly in the FSVM-4 and FSVM-5, the results of the best classification performance evaluation in FSVM-6 are given by a penalty value of $C = 2^1$ with the same evaluation values, namely sensitivity 72.340%, specifity 100%, G-means 85.053% and accuracy 83.333%.

From the performance of SVM and FSVM classification with several distance functions, it can be seen that the results of Fuzzy SVM give the best results to see the risk of diabetes. It can be seen in Table 2. the sensitivity value (SE) of the SVM method is superior with the highest percentage of 86.17% at $C = 2^1$ while the FSVM method with several distance functions gives the highest percentage of 75.53% at $C = 2^5$. However, in terms of specificity (SP), g-means (GM), and accuracy for all C penalty values, the FSVM method with several distance functions is very superior to the SVM method. The specificity value (SP) of the FSVM method with several distance functions gives an average percentage result of 100% while the SVM method has the highest specificity (SP) value with a percentage of 69.36% at $C = 2^{10}$, the value of g-means (GM) method FSVM with several distance functions gives the highest percentage of 86.91% at C= 2^5 while the SVM method has the highest g-means (GM) $C = 2^5$alue with a percentage of 77.31%. The accuracy value of the FSVM method with several distance functions gives the highest percentage of 85.256% at $C = 2^5$.

IV. CONCLUSION

In this paper, a method for developing SVM into FSVM has been presented with several distance functions including Euclid distance, Euclid distance, Canberra distance, Minkowski distance, Chebyshev distance, Minkowski Chebyshev distance, and Bray-Curtis distance where this distance is used to determine the best distance that can be seen from the results. the best accuracy, sensitivity, specificity, g-means. We applied the FSVM method with multiple distance functions to the Early Stage Diabetes data. The results of this comparison show that the FSVM method with several distance functions is more than the SVM method. Although the sensitivity (SE) value of the SVM method is superior, for the value of specificity (SP), g-means (GM), and accuracy on all C penalty values, the FSVM method with several distance functions is very superior to the SVM method.

ACKNOWLEDGMENT

Thank you to the leaders of Ahmad Dahlan University (UAD) and the UAD Postgraduate Program, and the UAD Mathematics Education Study Program who have facilitated this research. Thank you also said to reviewers who have provided input and comments so that this article becomes more qualified.

REFERENCES

[1] B. Scholkopf and A. Smola, Learning with Kernel :Support Vector Machines, Regulerizatiom, Optimization, and Beyond. 2002

[2] Vladimir. N. Vapnik., "The Nature of Statistical Learning," *Theory*. 1995

[3] L. Evans, N. Lohse, and M. Summers, "A fuzzy-decision-tree approach for manufacturing technology selection exploiting experience-based information," *Expert Syst. Appl.*, vol. 40, no. 16, pp. 6412–6426, 2013

[4] W. Purnami, A. M. Regresi, and L. Ordinal, "Perbandingan Klasifikasi Tingkat Keganasan Breast Cancer Dengan Menggunakan Regresi Logistik Ordinal Dan Support Vector Machine (SVM)," *J. Sains Dan Seni Its*, vol. 1, no. 1, 2012

[5] W. Purnami, A. M. Regresi, and L. Ordinal, "Perbandingan Klasifikasi Tingkat Keganasan Breast Cancer Dengan Menggunakan Regresi Logistik Ordinal Dan Support Vector Machine (SVM)," *J. Sains Dan Seni Its*, vol. 1, no. 1, 2012

[6] Y. Wang, S. Wang, and K. K. Lai, "A new fuzzy support vector machine to evaluate credit risk," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 6, pp. 820–831, 2005

[7] Z. Zhiwang, G. Gao, and Yong Shi, "Credit risk evaluation using multi-criteria optimization classifier with kernel, fuzzification and penalty factors," *Eur. J. Oper. Res.*, vol. 1, no. 237, pp. 335–348, 2014

[8] T.-Y. Wang and H.-M. Chiang, "Fuzzy support vector machine for multi-class text categorization," *Inf. Process. Manag.*, vol. 4, no. 43, pp. 914–929, 2007

[9] M. Perez, D. M. Rubin, L. E. Scott, T. Marwala, and W. Stevens, "A hybrid fuzzy-SVM classifer, applied to gene expression profling for automated leukaemia diagnosis," in *Conference: EduPath ConferenceAt: Cape Town, South Africa*, 2008, pp. 3–5

[10] N. Ö. Özcan and F. Gürgen, "Fuzzy support vector machines for ECG arrhythmia detection," *Proc. - Int. Conf. Pattern Recognit.*, pp. 2973–2976, 2010

[11] X. K. Ding, X. J. Yang, J. Y. Jiang, X. L. Deng, J. C. Cai, and Y. Y. Ji, "Optimization and analysis on fuzzy SVM for objects classification," *J. Inf. Hiding Multimed. Signal Process.*, vol. 9, no. 6, pp. 1421–1429, 2018

[12] X. Jiang, Z. Yi, and J. C. Lv, "Fuzzy SVM with a new fuzzy membership function," *Neural Comput. Appl.*, vol. 15, no. 3–4, pp. 268–276, 2006

[13] A. Shilton and D. T. H. Lai, "Iterative fuzzy support vector machine classification," *IEEE Int. Conf. Fuzzy Syst.*, pp. 1391–1397, 2007

[14] W. M. Tang, "No TitleFuzzy SVM with a New Fuzzy Membership Function to Solve the Two-Class Problems," *Neural Process. Lett.*, vol. 3, no. 34, pp. 209–219, 2011

[15] H. Li, F. Qi, and S. Wang, "A comparison of model selection methods for multi-class support vector machines," *Lect. Notes Comput. Sci.*, vol. 3483, no. IV, pp. 1140–1148, 2005

[16] C. F. Lin and S. De Wang, "Fuzzy support vector machines," *IEEE Trans. Neural Networks*, vol. 13, no. 2, pp. 464–471, 2002

[17] Kementrian kesehatan republik indonesia, InfoDATIN, Pusat Data dan Informasi Kementrian Kesehatan RI. 2020

[18] Kementrian kesehatan republik indonesia, InfoDATIN, Pusat Data dan Informasi Kementrian Kesehatan RI. 2018

[19] J. Han and M. Kamber, Data Mining Consepts And Techniques 3 edition. San Fransisco. 2012

[20] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. 1981

[21] B. Tirozi, D. B., and E. F, Introduction To Computational Neurobiology And Clustering. 2007

[22] M. Kokare, B. N. Chatterji, and P. K, "Biswas. Comparison of similarity metrics for texture image retrieval," *TENCON 2003*, vol. 2, pp. 571–575, 2003

[23] Jelínek. Jiří, Metrics in Similarity Search. 2007

[24] E. O. Rodrigues, "Combining Minkowski and Chebyshev: New Distance Proposal and Survey of Distance Metrics Using k-Nearest Neighbours Classifier," 2018

[25] K. Lokanayaki and A. Malathi, "A Prediction for Classification of Highly Imbalanced Medical Dataset Using Databoost.IM with SVM," 2014

[26] Q. Li, Y. Wang, and S. H. Bryant, "A novel method for mining highly imbalanced high-throughput screening data in PubChem," *Bioinformatics*, vol. 25, no. 24, pp. 3310–3316, 2009