

RESEARCH ARTICLE

Influence of fuzzy index parameter on FSVM classifier performance

P. Sridevi*, M. Punniyamoorthy and B. Senthil Arasu

Department of Management Studies, National Institute of Technology, Tiruchirappalli, Tamil Nadu, India.

Revised: 22 February 2017; Accepted: 25 May 2017

Abstract: Support vector machine (SVM), a machine learning algorithm used extensively for pattern analysis and recognition, is found sensitive to outliers and noise. Fuzzy support vector machine (FSVM) has been used in many applications as a most prominent technique by researchers to overcome the sensitivity issue faced by SVM, and for its good generalisation performance. In this research, a method to justify the performance of FSVM classifier by showing the influence of fuzzy index m on membership function of the model has been proposed. In the first phase of the study, an algorithm to find the optimal fuzzy index using fuzzy C-means (FCM) and to avoid testing all fuzzy index values on the FSVM model has been proposed. The operational complexity of the model can be reduced by this process. In the second phase of this study a FSVM algorithm to incorporate new membership function is proposed. The model is tested on six different datasets and kernel functions and the kernel with the highest classification accuracy is identified as an efficient kernel for the applied dataset. The experimental results on the chosen fuzzy index have proven that the proposed alternate methodology enhances classifier accuracy compared to other research findings. Hence, the model could be applied to diverse fields of FSVM applications.

Keywords: Fuzzy C-means (FCM), fuzzy index, fuzzy support vector machine (FSVM), kernel function, membership function, performance.

INTRODUCTION

With the emergence of machine learning concepts, the theory of support vector machine (SVM) (Cortes & Vapnik, 1995) laid the foundation for research in the areas of classification and prediction in the field of artificial intelligence. SVM, a prominent classifier, classifies a

dataset in higher dimensional space by the construction of optimal hyperplane. The algorithm is aimed at minimising the generalisation error and maximising the separation between margins of hyperplane. The SVM supports data of linearly separable and non-linearly separable cases. The datasets that are well defined into two different classes are linearly separable. For a dataset for which linear classification is not possible, a non-negative slack variable ξ_i , the error term, is introduced in the model to accommodate the violations that happen in the classification problem. As computational complexity of the data points expressed as dot products increases, the replacement of dot products by non-linear kernel function $k(x_i, x_j)$ reduces the complexity of higher dimensional mapping in feature space. The incorporation of kernel function into non-linear classification problem significantly influences the classifier model performance. The effect of kernel functions such as Gaussian RBF (Wang *et al.*, 2005; Tang, 2011; Wu *et al.*, 2014), polynomial (Wang *et al.*, 2005; Tang, 2011), linear kernel (Wang *et al.*, 2005), and quadratic kernel (Manning *et al.*, 2008) on FSVM classifier model performance have been studied in the past. SVM, a good classifier and a predictor due to its generalisation capability, has been used in a wide range of applications like financial time series forecasting (Tay & Cao, 2001), bioinformatics (Byvatov & Schneider, 2002), fault diagnosis (Li & Li, 2010) and forecasting stock index (Senthil Arasu *et al.*, 2014).

On the existence of outliers or noise, the data point may not show its exact belongingness to any of the classes defined. Lin and Wang (2002) reformulated the

* Corresponding author (psridevi@nitt.edu;  <https://orcid.org/0000-0003-2364-4266>)



SVM into a fuzzy support vector machine (FSVM) and incorporated fuzzy membership into the model such that different input points contribute differently to the learning of decision surface. The FSVM has been found to be very effective in many real time applications like credit risk evaluation (Wang *et al.*, 2005; Zhang *et al.*, 2014), text categorisation (Wang & Chiang, 2007), leukemia diagnosis (Perez *et al.*, 2008), ECG arrhythmia detection (Ozcan & Gorgen, 2010), etc.

The key issue in the FSVM model is the generation of appropriate membership for the data points to minimise the effect of outliers. Jiang *et al.* (2006), Shilton and Lai (2007), Tang (2011) and Wang *et al.* (2005) have reported new membership functions for FSVM models of linear and non-linear classification problems. As membership values have an influence on model performance, our focus was to propose an algorithm using fuzzy C-means (FCM) clustering to determine suitable membership values for classifier performance. Xie and Beni (1991) had proposed a new validity function by FCM algorithm to measure the overall average compactness and separation of fuzzy C-partition. Wu *et al.* (2014) studied a novel partition index maximisation (PIM) clustering method to define robust fuzzy membership for the FSVM model. Determination of an optimal fuzzy index value was focussed as an important objective of the present study, as different values of fuzzy index can affect membership values produced by the FCM algorithm, which in turn can effect the FSVM model performance. This is because of the fact that fuzzy index parameter (m), is defined as a variable in the objective function of the FCM algorithm. Instead of testing model performance on all values of m, this work focussed on defining an optimal fuzzy index value for suitable membership selection and the proposed model with new membership function was tested on six different benchmark datasets by deploying various kernel functions. The model parameter c and kernel parameter that produce the best model enhancement were chosen as the optimal parameters.

Fuzzy support vector machine (FSVM) model

Although SVM is an effective classifier tool for binary classification, it cannot enhance classifier generalisation performance in the presence of outliers or noises. Lin and Wang (2002) proposed an extended model of SVM known as the FSVM to reduce the effect of outliers or noise on model generalisation performance. In FSVM, each data point is assigned a fuzzy membership value to describe the attitude of the corresponding point to the learning of decision hyperplane. As the error term is influenced by the membership value of a data point,

the data point defined as an outlier is assigned a low membership value to minimise the effect of outliers on error term and is also treated as less important in the construction of the decision hyperplane. Lin and Wang (2004) automatically generated fuzzy memberships for training a data point from a heuristic strategy using confident factor and trashy factor and a mapping function. For instances belonging to multiple classes, Wang *et al.* (2005) proposed a bilateral weight based FSVM model and tested the performance of the classifier on different membership functions and kernel functions. Almasi *et al.* (2015) also proposed a new fuzzy membership function and an algorithm for both linearly separable and nonlinearly separable classification problems by adaptive particle swam optimisation (APSO). The FSVM model is described as follows:

A training set T is associated with membership values $(y_i, x_i, s_i) \dots (y_p, x_p, s_p)$ where each training data point $x_i \in \mathbb{R}^n$ is given a class label $y_i \in \{1, -1\}$ and fuzzy membership value s_i , where $0 < s_i \leq 1$ with $i=1 \dots p$.

Since the fuzzy membership s_i is the weight of the corresponding data point x_i toward one class and the parameter ε_i is a measure of error in FSVM, the term $s_i \varepsilon_i$ is a measure of the weighted error. The optimal hyperplane classification problem is then formulated as follows (Lin & Wang, 2002):

$$\text{Minimise } \frac{1}{2} w^T w + c \sum_{i=1}^p s_i \varepsilon_i$$

subject to,

$$y_i(w \cdot z_i + b) \geq 1 - \varepsilon_i, i = 1 \dots p$$

and $\varepsilon_i \geq 0, i = 1 \dots p$... (1)

Here c is a cost control parameter, which controls the trade-off between the maximum margin and minimisation of error. $s_i \in \{0, 1\}$ is a fuzzy membership degree of x_i belonging to y_i . ε_i is an error term of instance i. The assignment of a small value of s_i for data points considered as outliers reduces the effect of the error term on model performance, as the corresponding data point is given less importance in construction of the hyperplane. $z_i = \varphi(x_i)$ is a nonlinear function, which maps x_i to a high dimensional feature space.

Equation (1) could be transformed into a dual problem as

$$\text{Max}_\alpha \sum_{i,j} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j Y_i Y_j k(x_i, x_j)$$

subject to,

$$\sum_{i=1}^p \alpha_i Y_i = 0$$

$$0 \leq \alpha_i \leq s_i c, i = 1, \dots, p \quad \dots(2)$$

By solving equation (2), for the points x_i , where $\alpha_i > 0$ are identified as the support vectors. There are two possible existences of support vectors. The points x_i satisfying $0 < \alpha_i < s_i c$, falling on the hyperplane margins are identified as unbounded support vectors and the points x_i satisfying $\alpha_i = s_i c$ falling between the margin of the two classes are identified as bounded support vectors. The point x_i with $\alpha_i = 0$, is said to be properly classified and its error term value is taken as zero. Although x_i is correctly classified, it can take the error term value of $0 \leq \varepsilon_i < 1$ when it falls between the margin and the hyperplane of the concerned class. When x_i is misclassified, it takes the error term value $\varepsilon_i \geq 1$ (Soman *et al.*, 2009). The test sample class label can be predicted by solving the following classifier model.

$$Y(x) = \text{sgn} \left[\sum_{i=1}^{sv} \alpha_i y_i k(x, x_i) + b \right]$$

where,

$$b = \frac{1}{NS_v} \sum_{i \in S_v} y_{sv} - \sum_{i \in S_p} \alpha_i Y_i K(x, x_i) \quad \dots(3)$$

Based on the sign of the classifier model computation, test data get classified into one of the defined classes. In a study on FSVM, Zhang *et al.* (2014) proposed a multi-criteria optimisation classifier based on kernel, fuzzification and using penalty factors (KFP-MCOC) to address outliers, class imbalance and non-linearity problems. The present study has also incorporated fuzzy membership and kernel functions like radial basis function (RBF), polynomial, multi-layer perceptron (MLP), and quadratic to address non-linearity issues.

FCM based FSVM model

The key issues in FSVM research are, setting of appropriate membership to reduce effects of outliers and reduction of computational complexity in training the model. Earlier studies (Lin & Wang, 2002; Wang & Chiang, 2007; Tang, 2011) have proposed different membership functions for the FSVM model and substantiated experimental results on the applied datasets. FCM, a known clustering algorithm introduced by Bezdek *et al.* (1984) groups data of similar

characteristics by finding the centroid of clusters, and assigns membership values for data points belonging to the respective clusters. The algorithm is run iteratively until the optimality condition for the number of clusters and objective function are attained. FCM minimises the objective function with respect to fuzzy membership μ_{ij} and cluster centroid V_j ,

$$J_m = \sum_{i=1}^n \sum_{j=1}^c (\mu_{ij})^m d^2(X_i, V_j)$$

$$\text{subject to, } \sum_{j=1}^c \mu_{ij} = 1 \text{ and } 0 < \mu_{ij} \leq 1 \quad \dots(4)$$

where, μ_{ij} , the fuzzy membership is computed using

$$\mu_{ij} = \frac{\left(\frac{1}{d^2(X_i, V_j)} \right)^{\frac{1}{(m-1)}}}{\sum_{j=1}^c \left(\frac{1}{d^2(X_i, V_j)} \right)^{\frac{1}{(m-1)}}}$$

$$d^2(X_i, V_j) = (X_i - V_j)^T A (X_i - V_j)$$

A is a P X P positive definite matrix, P is the dimension of vectors. X_i ($i = 1, 2, \dots, n$), n is the number of vectors (or data points), c is the number of clusters, and $m > 1$ is the fuzziness index.

V_j , the centroid is computed using $V_j = \frac{\sum_{i=1}^n (\mu_{ij})^m X_i}{\sum_{i=1}^n (\mu_{ij})^m}$
 $S = \frac{J_m}{n * (d_{min})^2}$, is the cluster validity function which must be minimum for optimal cluster validity.

Where, J_m is the compactness function (equation 4), n is the number of data points, and $(d_{min})^2$ is the distance between two cluster centroids defined as separation.

To enhance the training efficiency of the classifier, researchers have also focussed on hybrid models. Xiao and Tong (2014) assessed various image segmentation methods in brain MR image segmentation and found that unsupervised classification algorithm of FCM and FSVM model have shown good prospects in image segmentation accuracy with high noise and high bias. In a similar study, cluster centres formed using FCM with a suitable fuzziness parameter is taken as a representation of samples closer to cluster centres and the new training dataset is taken to formulate FSVM (Xiong *et al.*, 2005). Also, the combination of multilayer FCM and FSVM algorithm was attempted for multiclass problems using classification strategies (Li & Yu, 2010). Yang *et al.* (2011) proposed a KFCM based FSVM algorithm by

determining the farthest pair of clusters to form a new training dataset with membership degrees using KFCM, and incorporated FSVM for the new dataset. Hussain *et al.* (2010) also adopted FCM based FSVM model in image steganography to exploit the strengths of soft computing techniques. In a related study by Wu *et al.* (2014), an algorithm has been proposed to set reasonable membership and to reduce computational complexity using the PIM fuzzy clustering based FSVM method.

The FSVM model has its objective function and constraints defined with membership s_i to nullify the effect of outliers in model performance (equation 2). Therefore, a membership function was proposed to derive membership value for the model generalisation performance. As the error term is influenced by the membership value of a data point, the data point defined as an outlier should be assigned a low membership value to minimise the effect of outliers on error term and is also treated as less important in the construction of the decision hyperplane.

$\frac{|\mu_1 - \mu_2|}{m} = s_i$, is proposed as new membership function for the FSVM model.

where, μ_1 is the probability of data belonging to class 1 and μ_2 is the probability of data belonging to class 2, and m is the fuzzy index parameter.

This study attempted to use FCM algorithm to derive membership values for data points and focused on incorporating new membership values for the FSVM model using fuzzy index parameter influence. As the value of m is considered 2 (based on literature) in the study, also as two membership values are derived for a data point from FCM and as one input on membership variable s_i is to be fed into the model, proposed membership function tested the influence of m by taking the difference of 2 cluster membership values of data points divided by m . Placing the highest value in denominator may give a low membership value for a data point acting as outlier, and thereby minimising its effect on error terms. Hence, the influence of fuzzy index parameter in deriving new membership is also analysed and incorporated in the research. The new FSVM model for the analysis is described as;

$$\text{Max}_{\alpha} \sum_{i,j} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j Y_i Y_j k(x_i, x_j)$$

subject to,

$$\sum_{i=1}^p \alpha_i Y_i = 0$$

$$0 \leq \alpha_i \leq \frac{|\mu_1 - \mu_2|}{m} * c, i = 1, \dots, p \quad \dots(5)$$

$\frac{|\mu_1 - \mu_2|}{m} = s_i$ is defined as the new membership for the FSVM model. As the membership value should be low for nullifying the effects of outliers on model performance, the new membership function was proposed by dividing the difference taken from membership values by fuzzy index parameter (m) and its significance on FSVM model performance was tested.

As the incorporation of kernel function $k(x_i, x_j)$ in the model reduces the complexity of higher dimensional mapping in feature space, and also significantly influences the classifier model performance, the following kernel functions were incorporated to test the FSVM classifier performance.

1. Polynomial kernel, $k(x_i, x_j) = (x_i^T \cdot x_j + 1)^d$; where d is a polynomial degree ... (6)

2. RBF kernel, $k(x_i, x_j) = \exp(\gamma \|x_i - x_j\|^2)$; where γ is the parameter controlling radius ... (7)

3. MLP kernel, $k(x_i, x_j) = \tanh(p_1 * x_i * x_j + p_2)$; where $p_1 > 0$ & $p_2 < 0$... (8)

4. Quadratic kernel, $k(x_i, x_j) = (x_i^T \cdot x_j + 1)^2$... (9)

The following work discusses the generation of an efficient algorithm for FCM based FSVM classifier model by adopting a different strategy.

Proposed model for optimal fuzzy index selection based on FCM algorithm

This study presents the selection of the optimal value of fuzzy index (m) through fuzzy C-means (FCM), the best known algorithm for fuzzy clustering (Xie & Beni, 1991). Wu (2012) proposed a method for selection of the fuzzy index parameter m and showed the influence of m on the membership function. It was proposed that for large m values, FCM is more robust to outliers and noise. The fuzzy index has also been shown to influence the selection of optimal cluster number (Rezaee *et al.*, 1998; Chen *et al.*, 2001) in clustering. As the assignment of appropriate membership value for data points influence the effect of outliers or noise on the FSVM classifier model performance, the generation of required membership value is addressed as an important issue in most of the research. FCM can be used to determine

the membership values for dataset and also as different values of m will generate different membership values, suitable membership values based on optimal choice of m could be established. Therefore, the choice of the parameter m is to be based on the optimality condition of m . Furthermore, the FCM objective function J is a minimisation function and cluster validity function S must also be minimum for optimal cluster validity. Hence, we propose a strategy for evolving optimal choice of fuzzy index m . The computational steps are described in the algorithm.

The proposed model computational steps for selection of the optimal fuzzy index (m) based on fuzzy C-means algorithm (Figure 1) are as follows:

Algorithm: 1

- Step 1: Initialise FCM parameters as number of clusters $c = 2$ and fuzzy index $m = 2$
- Step 2: Run the FCM algorithm and find the fuzzy membership for data (μ_1, μ_2) , objective function J and validity function S (equation 4), where μ_1 is the possibility of data belonging to class 1 and μ_2 is the possibility of data belonging to class 2.
- Step 3: Repeat the procedure for all possible values of m and terminate the process when there is only minimal change in J .
- Step 4: Find weighted average of all observed values of S and J as $SW_1 = S/\text{grand total}$ and $JW_2 = J/\text{grand total}$ and compute combined weighted average $(SW_1 + JW_2)$ of all values of m .
- Step 5: From the computed values of $(SW_1 + JW_2)$, look for minimum $(SW_1 + JW_2)$ value and choose the corresponding fuzzy index as optimal value of m .
- Step 6: Membership values of corresponding m is noted and computation is stopped.

Proposed model for FSVM based on new membership generation

The membership obtained through FCM algorithm for an optimal value of m is used to compute new membership values through the proposed membership function. Then, the FSVM model is used for final classification of data. As outliers has to be assigned with low membership to reduce its impact on classifier performance, a new membership function to assign the lowest value for outliers was defined. Data points other than outliers will not get affected with the new calculation of new membership. The computational steps involved in the proposed algorithm (Figure 2) are as follows:

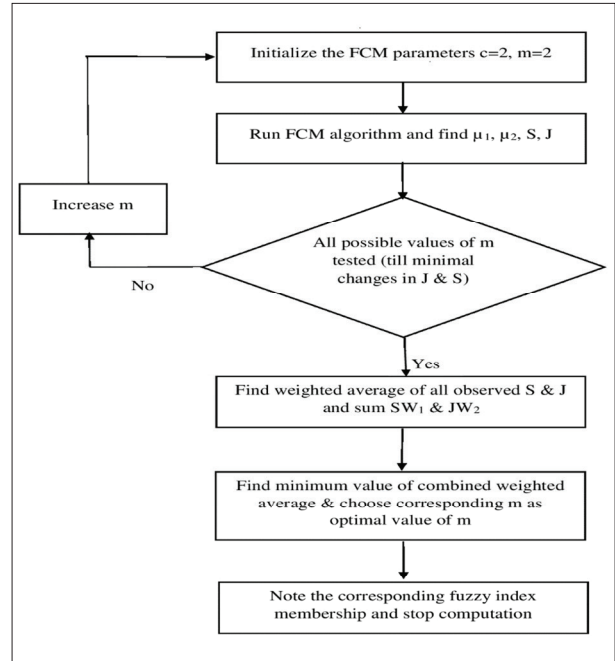


Figure 1: Flowchart for optimal fuzzy index (m) selection based on FCM algorithm

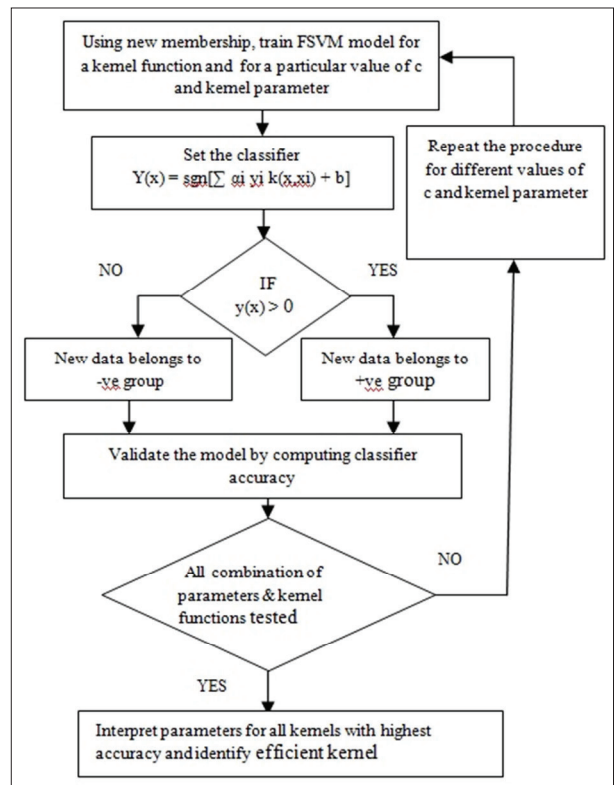


Figure 2: Flowchart for FSVM based on new membership generation

Algorithm 2:

Step 1: Calculate new membership s_i for data using optimal fuzzy index m value as $s_i = \frac{|\mu_1 - \mu_2|}{m}$;

where, m is the optimal fuzzy index (as derived from algorithm 1),

μ_1 is the possibility of data belonging to class 1 and μ_2 is the possibility of data belonging to class 2 (as derived from algorithm 1),

Step 2: Using new membership train the FSVM model for a kernel function (equation 6 for polynomial) and for a particular value of c and kernel parameter (say d for polynomial).

Step 3: Set the classifier and classify the test data into defined classes.

Step 4: Validate the model by computing classifier accuracy.

Step 5: Repeat the procedure for different values of c and kernel parameter until all combinations of parameters are tested.

Step 6: Repeat from step 2 to step 5 to test classifier performance for other kernel functions (equation 7; equation 8; equation 9) consecutively.

Step 7: Interpret the parameters for all kernel function with the highest accuracy and identify the kernel with the highest accuracy as the best kernel

METHODOLOGY

Empirical analysis

In this section, the performance of the FCM-based FSVM classifier is evaluated on six different benchmark datasets. The credit risk dataset was taken from Beynon and Peel (2001). Datasets such as PIMA, Heart, Australian credit, Thoracic surgery and MONK were downloaded from <http://archive.ics.uci.edu/ml>. The dataset descriptions are as follows:

Credit risk: The credit risk dataset is derived from the Financial Analysis Made Easy (FAME) CD-Rom database of corporations in the United Kingdom. It contains 30 failed and 30 non-failed firms described by 12 variables and 2 classes that classify them as good creditors and bad creditors.

PIMA: PIMA dataset consists of real world data of diabetes among Indians collected by the US National Institute of Diabetes and Kidney Diseases. It is defined by 8 variables and 768 instances with two class labels: 1 for positive result and 0 for negative result.

Heart: This dataset consists of 270 instances described by 14 features and classified into 2 classes: 1 indicating the presence of heart disease and 2 indicating its absence.

Australian credit: This dataset consists of credit card applications defined by a good mix of continuous and nominal attributes. It is described by 14 attributes and 690 instances with two class labels - 0 for not creditworthy applicants and 1 for creditworthy applicants.

Thoracic surgery: This dataset consists of post-operative life expectancy of lung cancer patients with the label 1 - indicating death within one year after surgery and 2 - indicating survival. It is defined with 17 attributes and 470 instances.

MONK: The Monk's problem was the basis of the first international comparison of learning algorithms and is a collection of 3 binary classification problems (MONK-1, MONK-2 and MONK-3) over a 6 attributes domain. Each problem involves learning a binary function defined over this domain from 432 instances that belong to class 0 or 1.

The performance of the FSVM classifier using kernels such as polynomial, RBF, MLP and quadratic kernel function was tested and compared. The parameter values taken for the different kernels are: polynomial kernel parameter, $d = \{2,3,4,5\}$; RBF kernel parameter $\Upsilon = \{2^{-3}, 2^{-1}, 2^1, 2^3, 2^5, 2^7, 2^9, 2^{11}\}$; MLP kernel parameter $p = \{-1,1\}, [-0.7,0.7], [-0.5,0.5]$ etc.} and quadratic kernel parameter of degree 2. For FSVM model constraint $0 \leq \alpha_i \leq s_i c$, the value of c was also tested on a range of values $c = \{2^{-3}, 2^{-1}, 2^1, 2^3, 2^5, 2^7, 2^9, 2^{11}, 2^{13}, \dots\}$. s_i is the fuzzy membership value computed from the proposed algorithm. Grid search method (Ögüt et al., 2009) was used for assessing the combination of c and kernel parameter into the model. A five-fold cross validation strategy was adopted for the input dataset. The FCM based FSVM model was run in Intel(R) core(TM) i7-2620M CPU, 8GB RAM, 64 bit operating system and MATLAB version 7.9.0.529 was used as analysis tool.

RESULTS AND DISCUSSION

The experiments were performed in two phases. In the first phase, FCM algorithm was used to find the membership values for data. The fuzzy index m is kept as a variable and the FCM algorithm was run for possible values of m and membership values, objective function J and validity function S were recorded on each run.

As described in Figure 3, when the changes in values of J reach a minimum and S to be small, the algorithm was terminated. The proposed procedure in the defined algorithm was followed to identify the optimal value of fuzzy index and the corresponding membership values of data were noted.

In the second phase of the experiment, new membership values (s_j) were generated based on the proposed membership function only for optimal fuzzy index (m) membership values, and the values were taken as one of the inputs into the FSVM model. As different membership values can very well influence the FSVM classifier performance, the choice of an appropriate membership value that can enhance classifier performance is addressed in the present study by finding optimal fuzzy index and its corresponding membership. The process defined in FSVM algorithm was done only for optimal m and not on all possible m values. We also attempted to prove that the classifier performance is not

affected by the methodology proposed by comparing with other FSVM models. Furthermore, the setting of low membership to minimise the effect of outliers and its influence on model enhancement was justified. Quadratic programming was used as a method for solving FSVM model. The FSVM model was run deploying different kernels and the influence of kernels on classifier accuracy was tested. Grid search method was used to record classifier accuracy on combinations of c and kernel parameter and the classifier accuracy rate of all kernels deployed was compared. The kernel with the highest classification accuracy on specific parameters was chosen as the best parameter value and the kernel function that could fit the model for the specified dataset. Table 1 and Figure 3 show the computational results of the selection of optimal parameter m for credit risk dataset.

The computational results of the proposed method shows that the optimal value of m is chosen at 5, because the combined weighted average is minimum at $m = 5$.

Table 1: Computational results of optimal parameter m selection for credit risk dataset

m	J	S	JW_1	SW_2	$JW_1 + SW_2$
2	192.1832	1056831.227	0.289388	3.3741E - 11	0.289388
2.5	142.4192	28890856.51	0.214454	9.22388E - 10	0.214454
3	102.9758	961763497.8	0.15506	3.07059E - 08	0.15506
3.5	73.62244	31603143196	0.11086	1.00898E - 06	0.110861
4	52.34698	1.17017E + 12	0.078824	3.73597E - 05	0.078861
4.5	37.11376	1.48992E + 12	0.055886	4.75681E - 05	0.055933
5	26.27489	1.68536E + 14	0.039565	0.00538079	0.044945
5.5	18.58333	7.56E + 15	0.027983	0.241430572	0.269413
6	18.58297	2.36E + 16	0.027982	0.753103523	0.781086

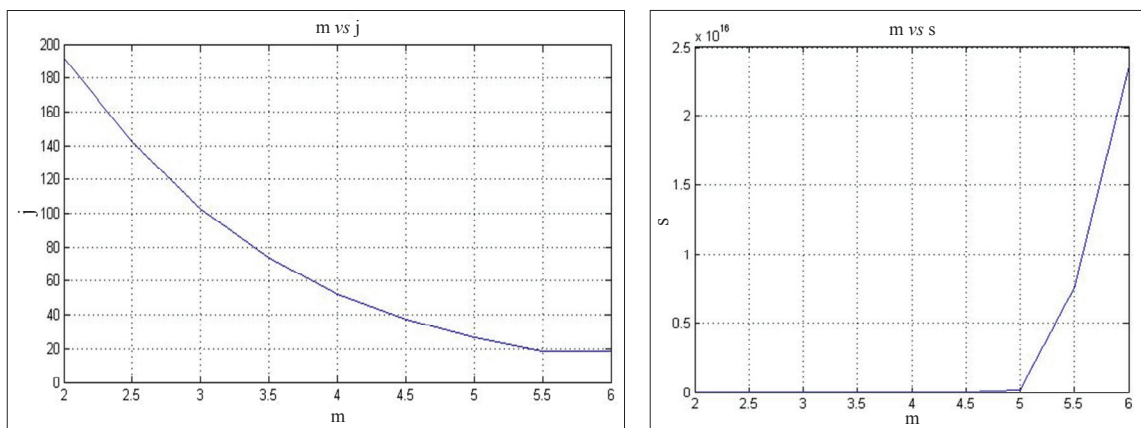


Figure 3: Choice of optimal fuzzy index parameter m for credit risk dataset

Table 2: FSVM classifier performance on different kernels for credit risk dataset

Kernel function	Parameters	Run time (s)	Accuracy (%)	Error rate (%)
RBF	$c = 2^{11}, r = 2^5$	0.103	83.57	16.43
Polynomial	$c = 2^3, d = 4$	0.139	81.67	18.33
MLP	$c = 2^5, p = [0.1, -0.1]$	0.146	78.33	21.67
Quadratic	$c = 2^7, n = 2$	0.173	80.07	19.93

Table 3: Computational results of optimal parameter m selection for PIMA dataset

m	S	J	JW_1	SW_2	$JW_1 + SW_2$
2	2.39919E + 11	3986825	6.6584E - 38	0.2675210	0.267521
2.5	7.65982E + 14	3075630	2.1258E - 34	0.2063787	0.206379
3	7.82E + 18	2292078	2.17E - 30	0.1538013	0.153801
3.5	2.14E + 22	1677363	5.9383E - 27	0.1125532	0.112553
4	2.16E + 26	1214804	5.9916E - 23	0.0815149	0.081515
4.5	5.75E + 29	874174.5	1.5959E - 19	0.0586582	0.058658
5	5.63E + 33	626418.1	1.5618E - 15	0.0420334	0.042033
5.5	1.50E + 37	447586.3	4.1671E - 12	0.0300336	0.030034
6	7.62E + 40	319149.9	2.1145E - 08	0.0214153	0.021415
6.5	7.19E + 44	227224	0.00019961	0.0152470	0.015447
7	3.60E + 48	161590.6	0.99980037	0.0108429	1.010643

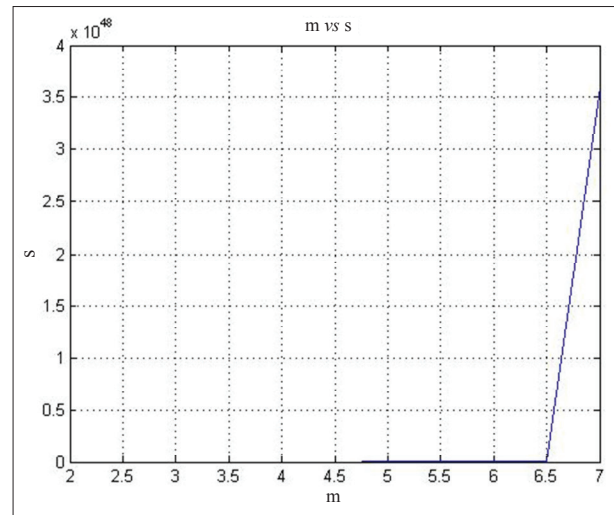
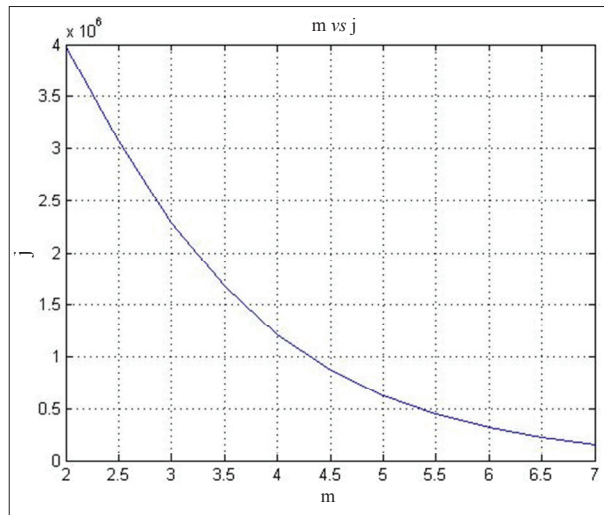


Figure 4: Choice of optimal parameter m for PIMA dataset

Table 2 describes the FSVM classifier performance using RBF, polynomial, MLP and quadratic kernels on credit risk dataset for $m = 5$. The table shows that the FSVM model attained the maximum classifier accuracy of 83.57 % with an error rate of 16.43 % for RBF kernel on parameter values $c = 2^{11}$ and $\gamma = 2^5$. The proposed

model has shown significant improvement over earlier models proposed for fuzzy index value of $m = 5$ with less computation time.

Table 3 and Figure 4 projects the computational results of optimal fuzzy index parameter m selection using FCM for PIMA dataset. The result shows that the

optimal fuzzy index m is chosen at 6.5 using FCM and the proposed algorithm.

Table 4 shows the computational results of the FSVM model for $m = 6.5$ on RBF, polynomial, MLP and quadratic kernels for the PIMA dataset. The

analysis result shows that the parameter values of $c = 2^9$, $\gamma = 2^3$ for RBF kernel fit best with the FSVM model for a classifier accuracy rate of 78.36 %. For the fuzzy index value of $m = 6.5$, the proposed model has shown significant improvement in classifier accuracy over other FSVM classification algorithms.

Table 4: FSVM classifier performance on different kernels for PIMA dataset

Kernel function	Parameters	Run time (s)	Accuracy (%)	Error rate (%)
RBF	$c = 2^9, r = 2^3$	46.94	78.36	21.64
Polynomial	$c = 2^{-3}, d = 2$	14.9	77.85	22.15
MLP	$c = 2^5, p = [0.05, -0.05]$	19.84	76.17	23.83
Quadratic	$c = 2^5, n = 2$	52.76	78.21	21.79

Table 5: Computational results of optimal parameter m selection for Heart dataset

m	J	S	JW_1	SW_2	$JW_1 + SW_2$
2	1117304	18294483934	0.294477	3.21201E - 30	0.294477
2.5	808208	3.54953E + 13	0.213011	6.232E - 27	0.213011
3	578632	6.80E + 16	0.152504	1.19E - 23	0.152504
3.5	412172.4	1.32E + 20	0.108632	2.31E - 20	0.108632
4	292804.6	2.53E + 23	0.077172	4.44E - 17	0.077172
4.5	207683.6	4.79E + 26	0.054737	8.41E - 14	0.054737
5	147169.3	9.45E + 29	0.038788	1.66E - 10	0.038788
5.5	104224.9	1.77E + 33	0.02747	3.10E - 07	0.02747
6	73782.38	3.28E + 36	0.019446	5.76E - 04	0.020022
6.5	52217.4	5.69E + 39	0.013762	9.99E - 01	1.013186

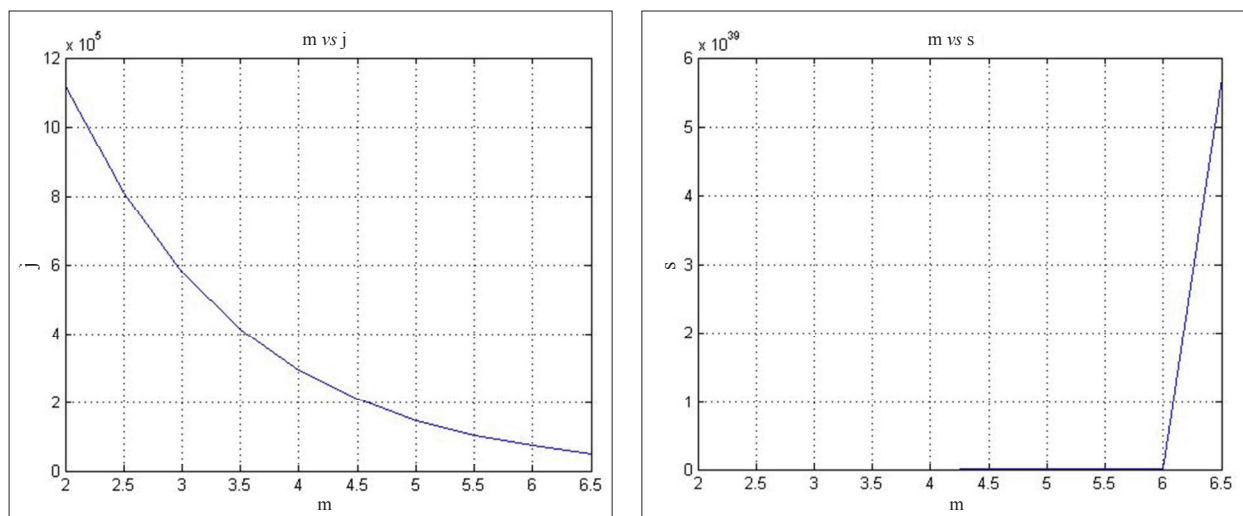


Figure 5: Choice of optimal parameter m for Heart dataset

Table 5 and Figure 5 projects the computational results of optimal fuzzy index parameter m selection using FCM clustering for the Heart dataset. The computational result shows that the optimal value of m is chosen at m = 6 from the proposed method.

The analysis shows that for the fuzzy index value of m = 6, the classifier achieved a maximum accuracy rate of 85.96 % with an error rate of 14.04 % for MLP kernel function with parameter values c = 2⁵ and p = ± 0.07.

Table 6 describes the FSVM model accuracy on the kernels like RBF, polynomial, MLP and quadratic kernel for m = 6 on the Heart dataset.

The results of different datasets from previous research findings, current research findings and other research paper findings are compared in Table 7.

Table 6: FSVM classifier performance on different kernels for Heart dataset

Kernel function	Parameters	Run time (s)	Accuracy (%)	Error rate (%)
RBF	c = 2 ⁷ , γ = 2 ⁵	0.18	84.44	15.56
Polynomial	c = 2 ⁻⁵ , d = 3	0.67	82.22	17.78
MLP	c = 2 ⁵ , p = [0.07, -0.07]	0.34	85.96	14.04
Quadratic	c = 2 ⁻³ , n = 2	0.36	80.74	19.26

Table 7: Comparative results of classifier performance by research findings

Dataset	FSVM model accuracy (%)		
	Our findings	Accuracy (%)	Other research findings Reference
Credit risk	83.57 (m = 5)	76.00 (U-FSVM) 79.00 (B-FSVM)	Wang <i>et al.</i> (2005)
PIMA	78.36 (m = 6.5)	77.56 (FCM-FSVM) 77.6 (PIM-FSVM) (m = 2)	Yang <i>et al.</i> (2011) Wu <i>et al.</i> (2014)
Heart	85.96 (m = 6)	84.8 (FSVM-KT) 84.5 (FSVM-kNN)	Lin and Wang (2004)

This study aims at identifying the optimal value of fuzzy index and runs the FSVM model only on optimal m based on the proposed methodology. This study shows that the complexity of the computational procedure is minimised by finding optimal m in the initial stage of computation than testing the impact of m in the FSVM model. Furthermore, with respect to classifier accuracy, the model outperforms models reported by Wang *et al.* (2005) that uses bilateral FSVM, Yang *et al.* (2011) that uses FCM-FSVM and Wu *et al.* (2014) that uses PIM-FSVM. As shown in Table 7, the classifier rate of the proposed model on optimal fuzzy index selection is observed to be 83.57, 78.36 and 85.96 % for credit risk,

PIMA and Heart dataset, respectively. It is interesting to note that in other models reported in literature, accuracies were found to be 79 % for credit risk and 77.56 % and 77.6 % for PIMA dataset. Wu *et al.* (2014) and Yang *et al.* (2011) tested the FCM-FSVM model only on the Gaussian kernel and for m = 2 as default. The researchers have not tested the influence of different values of m and the kernel functions on the model, which is taken as a research gap for the present study. The Heart dataset output also shows an improved accuracy rate of 85.96 % with the proposed method, over the method proposed by Lin and Wang (2004) with a classifier accuracy rate of 84.8 % and 84.5 %. The runtime analysis also

shows that the computation time taken by the model is significantly satisfactory. Inclusion of optimal value of m as a contributing variable in this study supported to generate appropriate membership values and contributed in reducing the effect of outliers present in the dataset. Hence, the influence of the proposed model on reducing the effect of outliers or noise is justified with the enhancement of FSVM classifier performance.

The classifier is also tested on datasets like Australian credit, Thoracic and MONK to substantiate the model performance using the same methodology (Tables 8 – 10).

It is observed that for the Australian credit dataset ($m = 18$), the RBF kernel suits well for parameters $c = 2^9$

and $r = 2^3$ showing the highest accuracy rate of 86.09 %. For the MONK dataset ($m = 4.5$) RBF and polynomial kernel shows a significant performance of 96.76 % for the parameters as shown in Table 9. But as the running time is found to be less for the polynomial kernel than for the RBF, the former could be taken as the best kernel to suit the dataset. In a study by Wu *et al.* (2014) using PIM fuzzy clustering based FSVM, a model accuracy of 94.44 % was reported for $m = 2$ and Gaussian kernel function. Similarly for the Thoracic dataset ($m = 26$), RBF and MLP kernel dominate with an accuracy rate of 85.11 %, but as RBF has taken just 0.364 seconds as computation time, it could be opted as the best kernel. It is obvious from the analysis that our proposed model performs significantly better for all types of datasets chosen for the study.

Table 8: FSVM classifier performance on different kernels for Australian credit dataset ($m = 18$)

Kernel function	Parameters	Run time (s)	Accuracy (%)	Error rate (%)
RBF	$c = 2^9, r = 2^3$	5.923	86.09	13.91
Polynomial	$c = 2^1, d = 3$	34.82	85.80	14.2
MLP	$c = 2^9, p = \pm 0.07$	9.692	83.19	16.81
Quadratic	$c = 2^5, n = 2$	50.14	85.09	14.91

Table 9: FSVM classifier performance on different kernels for MONK dataset ($m = 4.5$)

Kernel function	Parameters	Run time (s)	Accuracy (%)	Error rate (%)
RBF	$c = 2^{19}, r = 2^1$	21.723	96.76	3.24
Polynomial	$c = 2^1, d = 6$	12.167	96.76	3.24
MLP	$c = 2^{11}, p = [0.5, -0.5]$	10.499	72.92	27.08
Quadratic	$c = 2^{13}, n = 2$	13.711	82.87	17.13

Table 10: FSVM classifier performance on different kernels for Thoracic dataset ($m = 26$)

Kernel function	Parameters	Run time (s)	Accuracy (%)	Error rate (%)
RBF	$c = 2^9, r = 25$	0.364	85.11	14.89
Polynomial	$c = 2^7, d = 2$	32.24	81.28	18.72
MLP	$c = 2^{13}, p = [0.05, -0.05]$	0.61	85.11	14.89
Quadratic	$c = 2^1, n = 2$	5.22	84.26	15.74

CONCLUSION

In this paper, we have proposed an algorithm for the selection of optimal fuzziness index m using FCM, which has a significant contribution in the classification accuracy of the FSVM model. It has been proposed that the membership values must be set low to reduce the effect of outliers on the FSVM model performance. We have used a new membership function, where membership values are calculated only for the optimal fuzzy index values. The new membership values taken into the FSVM model are tested using kernels like RBF, polynomial, MLP and quadratic kernel function. The experimental results on the chosen fuzzy index have proven that the proposed alternate methodology significantly enhances the classifier accuracy compared to other research findings and also has reduced the complexity of the computational process by the simplified approach of the proposed algorithm.

In this study, we have been able to identify an efficient kernel function and parameters that can define the FSVM model effect in terms of classifier generalisation performance. The key issues of the FSVM such as suitable fuzzy membership setting and computational complexity are well addressed in the study and could help researchers working on diverse applications of FSVM.

REFERENCES

- Almasi O.N., Gooqeri H.S., Asl B.S. & Tang W.M. (2015). A new fuzzy membership assignment approach for fuzzy SVM based on adaptive PSO in classification problems. *Journal of Mathematics and Computer Science* **14**: 171 – 182.
- Beynon M.J. & Peel M.J. (2001). Variable precision rough set theory and data discretisation: an application to corporate failure prediction. *Omega* **29**(6): 561 – 576.
- Bezdek J.C., Ehrlich R. & Full W. (1984). FCM: the fuzzy c-means clustering algorithm. *Computers and Geosciences* **10**(2): 191 – 203.
- Byvatov E. & Schneider G. (2002). Support vector machine applications in bioinformatics. *Applied Bioinformatics* **2**: 67 – 77.
- Chen C.F., Lee J.M. & Li C. (2001). The validity measurement of fuzzy c-means classifier for remotely sensed images. *22nd Asian Conference on Remote Sensing* Singapore, 5 – 9 November.
- Cortes C. & Vapnik V. (1995). Support-vector networks. *Machine Learning* **20**(3): 273 – 297. DOI: <https://doi.org/10.1007/BF00994018>
- Hussain H.S., Aljunid S.A., Yahya S. & Ali F.H.M. (2010). A novel hybrid fuzzy- SVM image steganographic model. *Proceedings of IEEE International Symposium in Information Technology (ITSim)*, volume 1, Kuala Lumpur, Malaysia, 15 – 17 June, pp. 1 – 6. DOI: <https://doi.org/10.1109/ITSIM.2010.5561300>
- Jiang X., Yi Z. & Lv J.C. (2006). Fuzzy SVM with a new fuzzy membership function. *Neural Computing and Applications* **15**(3 – 4): 268 – 276.
- Li J. & Yu Z. (2010). An improved adaptive support vector machine algorithm with combinational fuzzy C-means clustering. *Proceedings of the IEEE 2nd International Conference on Advanced Computer Control (ICACC)*, volume 3, Shenyang, China, 27 – 29 March. pp. 269 – 272.
- Li Y.B. & Li Y. (2010). Notice of retraction survey on uncertainty support vector machine and its application in fault diagnosis. *Proceedings of the 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT)*, Chengdu, China, 9 – 11 July, pp. 561 – 565.
- Lin C.F. & Wang S.D. (2002). Fuzzy support vector machines. *IEEE Transactions on Neural Networks* **13**(2): 464 – 471. DOI: <https://doi.org/10.1109/72.991432>
- Lin C.F. & Wang S.D. (2004). Training algorithms for fuzzy support vector machines with noisy data. *Pattern Recognition Letters* **25**(14): 1647 – 1656. DOI: <https://doi.org/10.1016/j.patrec.2004.06.009>
- Manning C.D., Raghavan P. & Schütze H. (2008). Support vector machines and machine learning on documents. *Introduction to Information Retrieval*, pp. 319 – 348. Cambridge University Press, Cambridge, UK. DOI: <https://doi.org/10.1017/CBO9780511809071.016>
- Ögüt H., Doğanay M.M. & Aktaş R. (2009). Detecting stock-price manipulation in an emerging market: the case of Turkey. *Expert Systems with Applications* **36**(9): 11944 – 11949. DOI: <https://doi.org/10.1016/j.eswa.2009.03.065>
- Ozcan N.O. & Gurgun F. (2010). Fuzzy support vector machines for ECG arrhythmia detection. *Proceedings of the IEEE 20th International Conference on Pattern Recognition (ICPR)*, Istanbul, Turkey, 23 – 26 August, pp. 2973 – 2976. DOI: <https://doi.org/10.1109/ICPR.2010.728>
- Perez M., Rubin D.M., Scott L.E., Marwala T. & Stevens W. (2008). A hybrid fuzzy-SVM classifier, applied to gene expression profiling for automated leukaemia diagnosis. *Proceedings of the IEEE 25th Convention of Electrical and Electronics Engineers in Israel Eilat*, Israel, 3 – 5, December. DOI: <https://doi.org/10.1109/EEEI.2008.4736603>
- Rezaee M.R., Lelieveldt B.P. & Reiber J.H. (1998). A new cluster validity index for the fuzzy c-mean. *Pattern Recognition Letters* **19**(3): 237 – 246.
- Senthil Arasu B., Jeevananthan B., Thamaraiselvan N. & Janarthanan B. (2014). Performance of data mining techniques in forecasting stock index - evidence from India and US. *Journal of the National Science Foundation of Sri Lanka* **42**(2): 177 – 191.
- Shilton A. & Lai D.T.H. (2007). Iterative fuzzy support vector machine classification. *Proceedings of the IEEE*

- International Fuzzy Systems Conference*, London, UK, 23 – 26 July, pp. 1 – 6.
DOI: <https://doi.org/10.1109/FUZZY.2007.4295570>
20. Soman K., Loganathan R. & Ajay V. (2009). *Machine Learning with SVM and Other Kernel Methods*, pp. 123 – 147. PHI Learning Pvt. Ltd., New Delhi, India.
 21. Tang W.M. (2011). Fuzzy SVM with a new fuzzy membership function to solve the two-class problems. *Neural Processing Letters* **34**(3): 209 – 219.
DOI: <https://doi.org/10.1007/s11063-011-9192-y>
 22. Tay F.E. & Cao L. (2001). Application of support vector machines in financial time series forecasting. *Omega* **29**: 309 – 317.
 23. Wang T.Y. & Chiang H.M. (2007). Fuzzy support vector machine for multi-class text categorization. *Information Processing and Management* **43**(4): 914 – 929.
DOI: <https://doi.org/10.1016/j.ipm.2006.09.011>
 24. Wang Y., Wang S. & Lai K.K. (2005). A new fuzzy support vector machine to evaluate credit risk. *IEEE Transactions on Fuzzy Systems* **13**(6): 820 – 831.
DOI: <https://doi.org/10.1109/TFUZZ.2005.859320>
 25. Wu K.L. (2012). Analysis of parameter selections for fuzzy c-means. *Pattern Recognition* **45**(1): 407 – 415.
DOI: <https://doi.org/10.1016/j.patcog.2011.07.012>
 26. Wu Z., Zhang H. & Liu J. (2014). A fuzzy support vector machine algorithm for classification based on a novel PIM fuzzy clustering method. *Neurocomputing* **125**: 119 – 124.
DOI: <https://doi.org/10.1016/j.neucom.2012.07.049>
 27. Xiao J. & Tong Y. (2014). Research of brain MRI image segmentation algorithm based on FCM and SVM. *IEEE 26th Chinese Control and Decision Conference (CCDC)*, Changsha, China, 31 May – 2 June, pp. 1712 – 1716.
DOI: <https://doi.org/10.1109/CCDC.2014.6852445>
 28. Xie X.L. & Beni G. (1991). A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13**(8): 841 – 847.
DOI: <https://doi.org/10.1109/34.85677>
 29. Xiong S.W., Liu H.B. & Niu X.X. (2005). Fuzzy support vector machine based on FCM clustering. *Proceedings of the Fourth International Conference on Machine Learning and Cybernetics*, Guangzhou, China, 18 – 21 August, pp. 2608 – 2613.
 30. Yang X., Zhang G., Lu J. & Ma J. (2011). A kernel fuzzy c-means clustering-based fuzzy support vector machine algorithm for classification problems with outliers or noises. *IEEE Transactions on Fuzzy Systems* **19**(1): 105 – 115.
DOI: <https://doi.org/10.1109/TFUZZ.2010.2087382>
 31. Zhang Z., Gao G. & Shi Y. (2014). Credit risk evaluation using multi-criteria optimization classifier with kernel, fuzzification and penalty factors. *European Journal of Operational Research* **237**(1): 335 – 348.
DOI: <https://doi.org/10.1016/j.ejor.2014.01.044>