

GSA to Obtain SVM Kernel Parameter for Thyroid Nodule Classification

Dias Aziz Pramudita*¹, Aina Musdholifah²

¹Program Studi S2 Ilmu Komputer FMIPA UGM, Yogyakarta, Indonesia

²Departemen Ilmu Komputer dan Elektronika, FMIPA UGM, Yogyakarta, Indonesia

e-mail: *dias.aziz.pramudita@mail.ugm.ac.id, aina.m@ugm.ac.id

Abstrak

Support Vector Machine (SVM) merupakan salah satu metode populer dalam permasalahan klasifikasi dikarenakan solusinya yang global optima. Namun, pemilihan nilai parameter dan kernel yang sesuai masih menjadi kendala dalam prosesnya. Permasalahan tersebut dapat diselesaikan dengan menambahkan proses optimasi parameter pada SVM. Gravitational Search Algorithm (GSA) merupakan algoritma optimasi yang terinspirasi oleh interaksi massa dan hukum gravitasi Newton. Algoritma hybrid GSA-SVM ini akan digunakan untuk meningkatkan performa akurasi klasifikasi nodul tiroid. Data yang digunakan dalam penelitian ini berupa data citra nodul tiroid hasil ultrasonografi yang diperoleh dari RSUP Dr. Sardjito, Yogyakarta. Data tersebut digunakan sebagai objek penelitian untuk mengevaluasi penggunaan algoritma GSA-SVM. Evaluasi yang dilakukan dengan membandingkan akurasi klasifikasi menggunakan SVM dengan parameter default dengan parameter yang telah dioptimasi menggunakan kombinasi GSA-SVM. Hasil penelitian ini menunjukkan bahwa Gravitational Search Algorithm (GSA) dapat digunakan untuk mencari parameter optimal pada classifier SVM dan meningkatkan akurasi. Hal ini dibuktikan dengan meningkatnya akurasi pada kernel polynomial dari 58.5366 % menjadi 89.4309 %, dan 41.4634 % menjadi 98.374 % pada kernel RBF.

Kata kunci— Gravitational Search Algorithm, SVM, Nodul Tiroid, GSA-SVM

Abstract

Support Vector Machine (SVM) is one of the most popular methods of classification problems due to its global optima solution. However, the selection of appropriate parameters and kernel values remains an obstacle in the process. The problem can be solved by adding the best value of parameter during optimization process in SVM. Gravitational Search Algorithm (GSA) will be used to optimize parameters of SVM. GSA is an optimization algorithm that is inspired by mass interaction and Newton's law of gravity. This research hybridizes the GSA and SVM to increase system accuracy. The proposed approach had been implemented to improve the classification performance of Thyroid Nodule. The data used in this research are ultrasonography image of Thyroid Nodule obtained from RSUP Dr. Sardjito, Yogyakarta. This research had been evaluated by comparing the default SVM parameters with the proposed method in term of accuracy. The experiment results showed that the use of GSA on SVM is capable to increase system accuracy. In the polynomial kernel the accuracy rose up from 58.5366 % to 89.4309 %, and 41.4634 % to 98.374 % in Polynomial kernel.

Keywords— Gravitational Search Algorithm, SVM, Thyroid Nodule, GSA-SVM

1. INTRODUCTION

In machine learning and pattern recognition, classification refers to an algorithmic procedure for assigning a given piece of input data into one of a given number of categories based on quantitative information on one or more characteristics inherent in the items [1]. Classification and prediction play important roles in data mining and their problems arise in many data mining applications such as computer vision, speech recognition, natural language processing and so on [2]. A bundle of approaches was developed to build classification models including SVM which were developed by Vapnik and his colleagues [1]. They have recently attracted a lot of researchers from the machine learning (ML) and pattern classification community for their fascinating properties such as high generalization performance and globally optimal solutions [2].

Support vector machines (SVM) has recently been used in a range of problems including pattern recognition, bioinformatics, and text categorization. SVM classifies data with different class labels by determining a set of support vectors that are members of the set of training inputs that outline a hyperplane in the feature space. SVM provides a generic mechanism that fits the hyperplane surface to the training data using kernel function (e.g. linear, polynomial, or RBF) for the SVM during the training process that selects support vectors along the surface of this function [2].

One problem that faces the user of SVM is how to choose a kernel and the specific parameters for that kernel, it is the crucial step in handling a learning task with SVM since it has a heavy impact on the classification accuracy. Parameters that should be optimized include the penalty parameter C and the kernel function parameters such as the gamma (γ) for the radial basis functions (RBF) kernel. In other words, the largest problems encountered in setting up the SVM model are how to select the kernel function and its parameter values [3].

The proposed algorithm will be used to classify thyroid nodules into malignant or benign classes. In medical circumstances, marking off where the malignant and benign thyroid nodules is a crucial point, although the prevalence of malignant nodules that occur is relatively small only about 7-15% [4], however knowing a lesion is benign or malignant will affect the following treatment, for example if a malignant lesions is detected, biopsy needle is necessary, and vice versa. If there is a malignant lesion that is considered as a benign condition, that is the problem because it will not get appropriate therapy.

Ultrasonography (USG) is an imaging modality that has sensitivity in detecting the existence of thyroid nodule. Eventhough the sensitivity of an USG in detecting malignant nodule is about 63-94%, specificity 61,95%, and accuracy amount to 78-94%, USG has a major disadvantage due to its operator dependent. This causes subjectivity of the operator in the diagnosis. Imaging interpretation result in the examination depends on operator's expertise and skills, beside that operator fatigue can also cause diagnostic errors. Moreover, noise and speckle found on the image increasingly influences the accuracy of the diagnosis [5]. Therefore, a computer-based system is necessary to assist radiologist to classify the malignancy of a thyroid nodule.

2. METHODS

2.1 Literature Review

2.1.1 Image Processing

In general, digital image processing refers to the processing of two-dimensional images using a computer. Although an image has a lot of information, but often the quality is decreased (degradation), for example it contains noise, the color is too contrasting, less sharp, blurring, etc. This makes more difficult to interpret information of the image. Therefore, digital image processing is necessary [6].

Digital image processing is a technique to manipulate, modify, or increase the quality of an image in various ways [6]. Fig. 1 shows the complete stages of digital image processing.

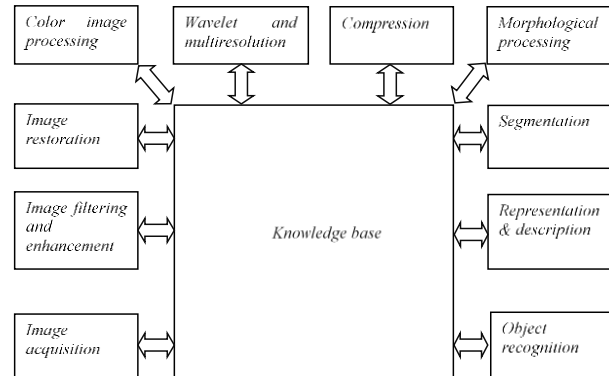


Figure 1. Digital image processing complete stages [6]

Fig. 1 shows that there are 10 processes in digital image processing. In fact, not all stages are always done, it depends on the images or the objective of the system. In this research, the steps of pre-processing stages are:

1. *Image filtering and enhancement*, in this stage noise filtering had been applied two times, adaptive median for smoother, and SRBF to remove speckle.
2. *Segmentation*, after enhancement process, the main object of the image has been separated from background. The method used in this stage is *Fast Global Minimization for Active Contour (FGMAC)*.
3. *Representation & description*, the main object has been represented according to the characteristic of the edge. Then, it has been extracted to get the value of the features. This research provides geometrical and statistical aspects to extract the features of object, 8 features used to describe the characteristic of the nodule, they are: *convexity, solidity, aspect ratio, compactness, circularity, dispersy, tortuosity, rectangularity*.

2.1.2 Support Vector Machine

SVM is a set of related supervised learning methods used for classification and regression. SVM can be defined as systems which use the hypothesis space of a linear function in a high dimensional feature space, trained with a learning algorithm from optimization theory that implements a learning bias derived from statistical learning theory. SVM has properties that are not possessed by other learning machines, specifically in the process of finding the best hyperplane that maximize the margin between non-linear input spaces and feature spaces using the kernel rules [7].

The best hyperplane between the two classes can be obtained by measuring the margin of the hyperplane and looking for the maximum point. Illustration of margin can be seen in Fig. 2 which is shown by the red line. Margin shows the distance between the hyperplane and the closest data from each class. The closest data from each class called support vector [8].

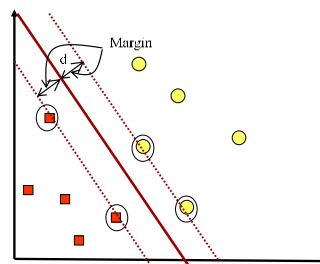


Figure 2. Illustration of the optimized hyperplane on SVM

In the real problem, generally data cannot be classified 100% correctly (linearly separable), it needs kernel function. Kernel function used in this research are shown in the Table 1.

Table 1. SVM kernel

Name of Kernel	Equation
Linear	$K(x, y) = (x \cdot y)$ (1)
Polynomial	$K(x, y) = [(x \cdot y) + 1]^d$ (2)
RBF	$K(x, y) = \exp(-\gamma \ x - y\ ^2)$ (3)

As it was mentioned, SVM is a classifier, given a set of training examples, each marked as belonging to one of two categories, and the SVM training algorithm builds a model that predicts whether a new example falls into one category or the other.

Given a training set of D , which

$$D = \{(O_i, y_i) \mid O_i \in R^p, y_i \in \{-1, 1\}\}_{i=1}^{PT}, \quad (4)$$

where PT is the number of samples, pair (O_i, y_i) represent the i^{th} training input sample and its label respectively, and $O_i = (O_{i1}, O_{i2}, \dots, O_{ip})$ is a p -dimensional vector in the feature space.

The generalized linear SVM finds an optimal separating hyperplane which can be written as the set of points O satisfying: $(w \cdot O - b = 0)$, by solving the following optimization problem [7]:

$$\min_{w, b, \zeta} \frac{1}{2} \langle w \cdot w \rangle + C \sum_{i=1}^{PT} \zeta_i \quad \text{subject to: } y_i (\langle w \cdot O_i \rangle + b)_{\zeta \geq 0} + \zeta - 1 \geq 0, \quad (5)$$

where C is the penalty parameter, which controls the trade-off between the complexity of the decision function and the number of training examples that have been misclassified, and ζ_i is the non-negative slack variable. A good separation is achieved by the hyperplane that has the largest distance to the nearest training data points of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier. This optimization model can be solved by introducing the Lagrange multipliers $0 \leq \alpha_i \leq C$ for its dual optimization model. After the optimal solution α_i^* is obtained, the optimal hyperplane parameters w^* and b^* can be determined, and the indicator function (classifier) can be written as:

$$\text{sign}(\langle w^* \cdot O \rangle + b^*) \quad \text{or} \quad \text{sign}(\sum_{i=1}^{PT} y_i \alpha_i^* \langle O_i \cdot O \rangle + b^*) \quad \text{for linear kernel,} \quad (6)$$

In the nonlinearly separable cases, the SVM maps the training points, nonlinearly, to a high-dimensional feature space using kernel function $K(O_i, O_j)$ where linear separation may be possible. The crucial idea is to use kernels to reduce a complex classification task to one that can be solved with separating hyperplanes. After selecting the kernel function, the nonlinear SVM classifier becomes:

$$\text{sign}(\sum_{i=1}^{PT} y_i \alpha_i^* \exp(-\gamma \|O_i - O_j\|^2) + b^*) \quad \text{for RBF kernel,} \quad (7)$$

$$\text{sign}(\sum_{i=1}^{PT} y_i \alpha_i^* ((O_i \cdot O_j) + 1)^d + b^*) \quad \text{for Polynomial kernel} \quad (8)$$

The performance of an SVM can be controlled through the term C and γ in RBF kernel, and C and d in Polynomial kernel, which are called hyperparameters. These parameters influence the number of support vectors and the maximization margin of the SVM.

2.1.3 Gravitational Search Algorithm

The GSA was first introduced by Rashedi et al. as a new stochastic population-based heuristic optimization tool [9]. This approach provides an iterative method that simulates mass interactions, and moves through a multi-dimensional search space in the influence of gravitation. This heuristic algorithm has been inspired by the Newtonian laws of gravity and motion [9]. The effectiveness of GSA in solving a set of nonlinear benchmark functions has been proven. Moreover, the results confirm that GSA is a suitable tool for the optimization of engineering problems.

In GSA, agents are considered as objects and their performance is measured by their masses. All objects attract each other by the gravity force, and this force causes a global movement of all objects with heavier masses. The heavy masses – which correspond to good solutions – move more slowly than the lighter ones, this guarantees the exploitation step of the

algorithm.

In GSA, each mass (agent) has four specifications: position, inertial mass, active gravitational mass, and passive gravitational mass. The position of the mass corresponds to a solution of the problem, and its gravitational and inertial masses are determined using a fitness function. In other words, each mass presents a solution, and the algorithm is navigated by properly adjusting the gravitational and inertia masses. By lapse of time, it expected that masses be attracted by the heaviest mass. That mass will present the optimum solution in the search space.

Now, consider a system with N agents (masses). The position of i^{th} agent is defined by:

$$X_i = (X_i^1, \dots, X_i^d, \dots, X_i^N) \quad \text{for } i = 1, 2, 3, \dots, N. \quad (9)$$

Where X_i^d presents the position of i^{th} agent in the d^{th} dimension.

The gravitational constant, G , is initialized in the beginning and will be reduced with time to control the search accuracy. In other words, G is a function of the initial value (G_0) and time(t):

$$G(t) = G_0 \times \left(1 - \frac{t}{T}\right), \quad (10)$$

Gravitational and inertia masses are simply calculated by the fitness evaluation. A heavier mass means a more efficient agent. This means that better agents have higher attractions and walk more slowly. Assuming the equality of the gravitational and inertia mass, the values of masses are calculated using the map of fitness. Gravitational and inertia masses are updated using following equations:

$$M_{ai} = M_{pi} = M_{ii} = M_i, \quad i = 1, 2, \dots, N. \quad (11)$$

$$m_i(t) = \frac{fit_i(t) - worst(t)}{best(t) - worst(t)}, \quad (12)$$

$$M_i(t) = \frac{m_i(t)}{\sum_{j=1}^N m_j(t)}, \quad (13)$$

Where $fit_i(t)$ represent the fitness value of agent i at time t , and $best(t)$ and $worst(t)$ are defined as follows (maximization problem):

$$best(t) = \max_{j \in \{1, \dots, N\}} fit_j(t), \quad (14)$$

$$worst(t) = \min_{j \in \{1, \dots, N\}} fit_j(t), \quad (15)$$

At a specific time ' t ', the force acting on mass ' i ' from mass ' j ' defined as follow:

$$F_{ij}^d(t) = G(t) \frac{M_{pi}(t) \times M_{aj}(t)}{R_{ij}(t) + \varepsilon} (X_j^d(t) - X_i^d(t)), \quad (16)$$

Where M_{aj} is the active gravitational mass related to agent j , M_{pi} is the passive gravitational mass related to agent i , $G(t)$ is gravitational constant at time t , ε is a small constant, and $R_{ij}(t)$ is the Euclidean distance between agent ' i ' and ' j ':

$$R_{ij}(t) = \|X_i(t), X_j(t)\|_2, \quad (17)$$

So, the total force that acts on agent i in a dimension d be randomly weighted sum of d^{th} components of the forces exerted from other agents:

$$F_i^d(t) = \sum_{j=1, j \neq i}^N rand_j F_{ij}^d(t), \quad (18)$$

Where $rand_j$ is a random number in the interval $[0, 1]$.

Hence, by the law of motion, the acceleration of the agent i at time t , and in direction d^{th} is given as follow:

$$a_i^d(t) = \frac{F_i^d(t)}{M_{ii}(t)}, \quad (19)$$

Where M_{ii} is the inertial mass of i^{th} agent.

Furthermore, the next velocity of an agent is considered as a fraction of its current velocity added to its acceleration. Therefore, its position and velocity could be calculated as follows:

$$v_i^d(t+1) = rand_i \times v_i^d(t) + a_i^d(t), \quad (20)$$

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1), \quad (21)$$

where $rand_i$ is a uniform random variable in the interval $[0, 1]$ to give randomized

characteristic to the search. Fig. 3 illustrates the flow of the GSA.

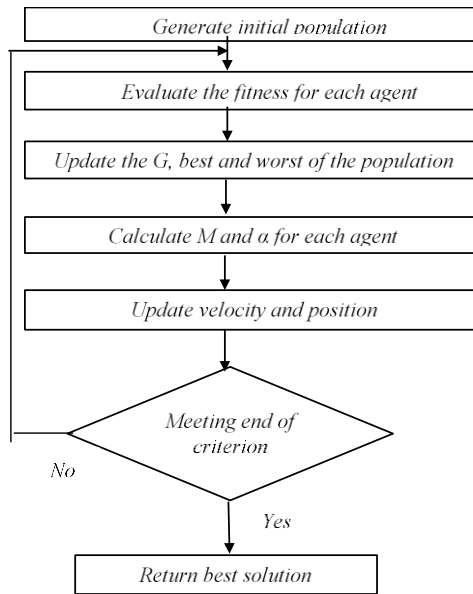


Figure 3. Flowchart of general principle of GSA [9]

2.2 The Detail Method of GSA-SVM System

There are two key factors when using the GSA as an optimization algorithm: one is what to choose as the masses of the GSA, another is how to define the fitness function which evaluates the goodness of a particle. This experiment tried to find which kernel that perform better in this problem. The research comparing the results of each kernel, *linear*, *polynomial*, and *RBF*.

2.2.1 Mass Representation

In this problem, the agent represented by the combination of parameter of each kernel, for example: in *Polynomial* kernel, the agent represented by the combination of parameter C and parameter d , while parameter C and parameter γ used in *RBF* kernel. The mass of each agent calculated by its accuracy. The accuracy is obtained during the training process using Eq. (7) and Eq. (8). Then, the value is normalized using Eq. (12) and Eq. (13) for further processing.

2.2.2 Fitness Function Definition

Classification accuracy is the criteria used to design a fitness function. Thus, the fitness function should be designed such that the supermass with high classification accuracy. For the fitness definition, the classification accuracy (acc_i) or hit rate denoting the percentage of correctly classified examples is evaluated by Eq. (22). The numbers of correctly classified examples are indicated by TP and TN while FP and FN represent incorrectly classified.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%, \quad (22)$$

2.2.3 The Proposed GSA-SVM

The detailed steps of the algorithm can be seen in the Fig. 4. The experiment designed by splitting the data set into two parts, 60 % as a training set, and the rest 40 % as the testing set. The training set is used to compare among others kernel performance, while testing set is untouchable.

In the Fig. 4, the step begin with search space identification and initialization. Secondly, training accuracy of each agent is evaluated as a fitness value. Then, after updating agent's

position, if the criteria is reached, the system will be stopped and the optimized value will be used to test the testing set.

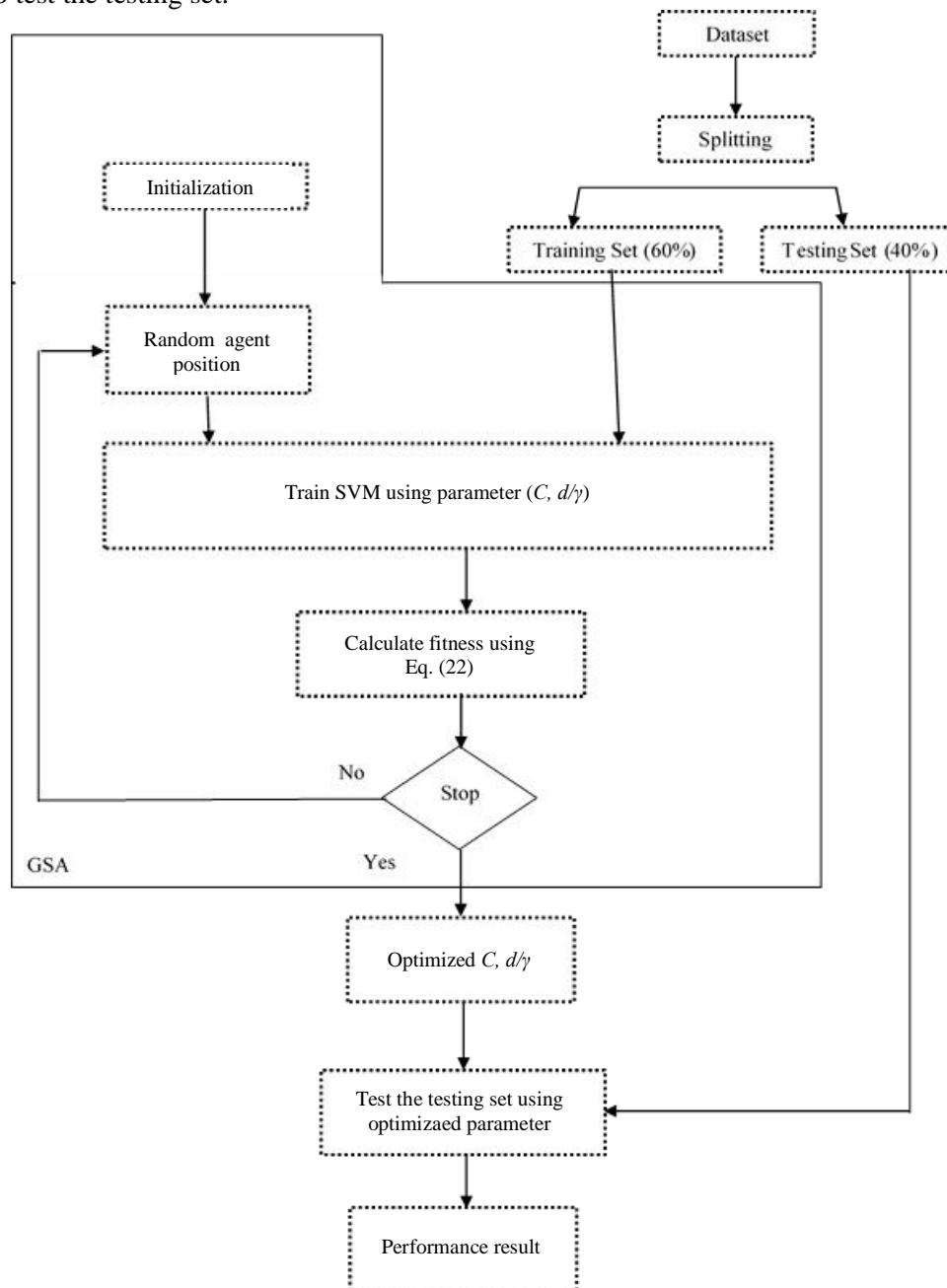


Figure 4. Flowchart of GSA-SVM hybrid system

3. RESULTS AND DISCUSSION

3.1 Data Description

The data sets used in the experimentation are all obtained from Radiology Installation of Rumah Sakit Umum Pusat (RSUP) Dr. Sardjito, Yogyakarta with range of periode from 2011 to 2014. 160 data as total is obtained from previous research [10]. The entire data belong to *.bmp RGB format.

3.2 Pre-processing

The data obtained from RSUP Dr. Sardjito is processed before being classified. The presence of noises or other obstacles can interfere the classification stages, so that it should be removed first. Complete digital image processing stages in this research can be seen in Fig. 5.

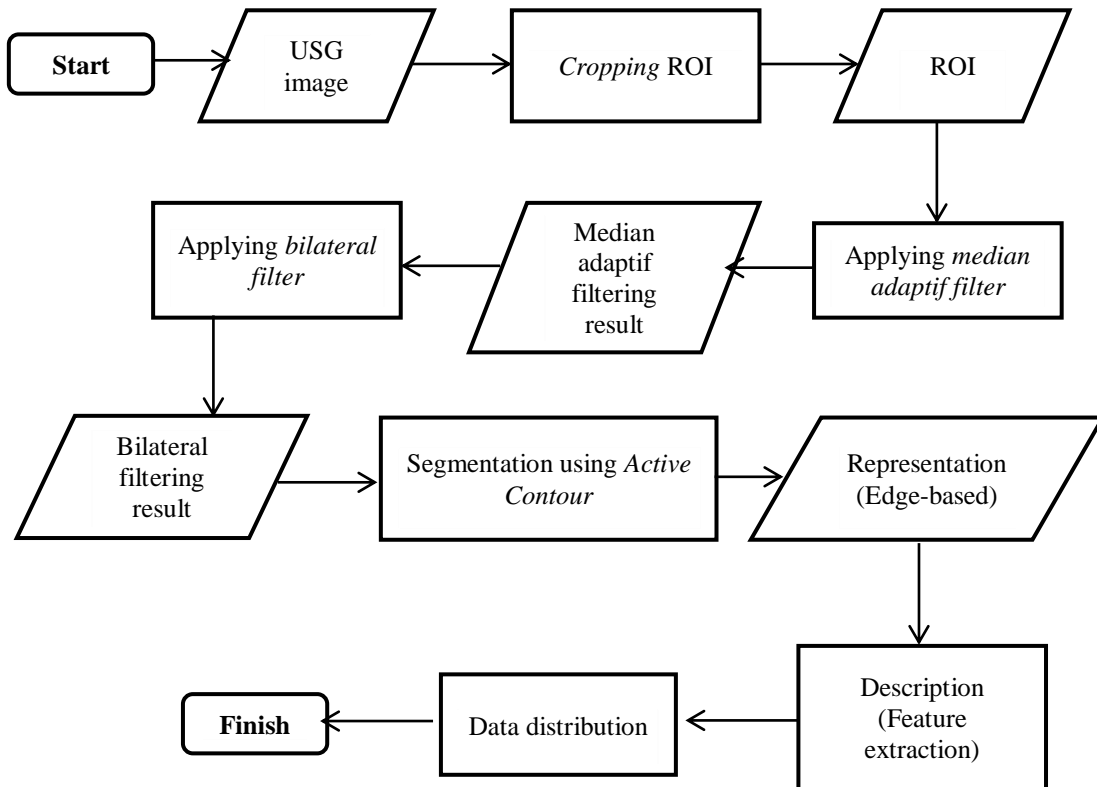


Figure 5. Flowchart of pre-process image data

According to the Fig. 5, the first step of this pre-processing is to get ROI of the image. The cropping process has been done manually by the experts, Dr. Endang Sri Wulandari as a radiologist and 2 assistants that accompanied by the researcher as the director of the use of the system. The result of cropping process can be seen in the Fig. 6.



Figure 6. (a) Real image from USG, (b) Cropping result as ROI of the image

Noise reduction process had done in two steps. First, median adaptif filter has been applied on the ROI, then bilateral filter at a later stage. Fig. 7 (b) shows the result of median adaptif filtering, while bilateral filtering result is provided by Fig. 7 (c). It can be seen that the last image presented smoothest image among others.

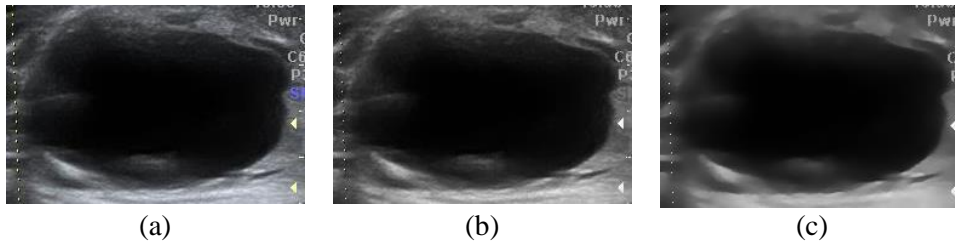


Figure 7. Image (a) ROI, (b) result of median adaptif filtering (c) result of bilateral filtering

The segmentation process is carried out based on the edges of the nodule. The method used is Fast Global Minimization For Active Contour (FGMAC). Segmented image illustrated as white while the background is black

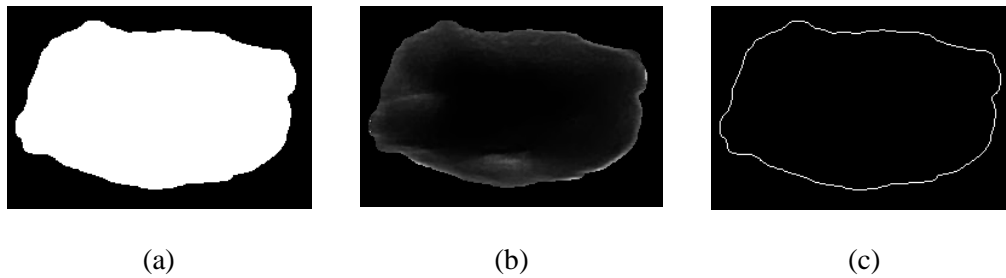


Figure 8. Image (a) segmentation result, (b) masking area, (c) edge detection

After segmenting the image, the next step is representation and description stage. The process is done by extracting the image features. Geometric and statistical characteristics were chosen considering that the radiologist recognized the edge of the ultrasound image also based on the shape of the edge and firmness of the edges. Data should be described as a value to simplify the calculation in the *.xlsx format, before it is splitted out into training and testing set.

	A	B	C	D	E	F	G	H	I	J	K	L	M
	Convexity	Solidity	Aspek Ratio	Compactness	Circularity	Dispersy	Turtolisy	Rectangularity	Class				
2	0.090358145	244.5841321	1	0.276719349	0.205497176	0.006465835	0.620398249	0.754037817	Irregular				
3	0.402961917	65535	1	0	0.493083507	0.024505742	0.856422418	0.767406409	Irregular				
4	0.106757835	161.009901	1	0.291067703	0.375151189	0.007462531	0.620427789	0.571393474	Irregular				
5	0.476450778	80.55518395	1	0.055833328	1.916639577	0.011585204	0.724782378	0.872643174	Well Define				
6	0.333262017	165.3770492	1	0.057652197	1.348245	0.013474438	0.78572329	0.747873146	Well Define				
7	0.499269288	3.674832962	1	1.138539955	0.782135692	0.038659827	0.90480445	0.803763146	Well Define				
8	0.627172971	65535	1	0	2.649841176	0.039804322	0.812008176	0.80685491	Well Define				
9	0.731275243	74.91743119	1	0.026702775	2.175993908	0.019912296	0.742483164	0.866526556	Well Define				
10	0.255987568	104.5238095	1	0.114378907	0.194758808	0.022488145	0.911288833	0.77440057	Well Define				
11	0.159596472	-74.725	1	-0.49807699	0.278174314	0.009744151	0.827716327	0.769323875	Well Define				
12	0.257729644	18.25759768	1	0.799092588	1.608000933	0.016493859	0.727575246	0.803641704	Well Define				
13	0.585196901	2.42095882	1	1.458253424	9.45461093	0.033829414	0.761595535	0.824170708	Well Define				
14	0.178109503	-27910.66667	1	-0.00956918	0.137428254	0.007367848	0.782899278	0.834106579	Well Define				
15	0.187781627	58.5868396	1	0.473942757	0.84681856	0.007756232	0.753427239	0.762967669	Well Define				
16	0.508996923	-5.395348837	1	-0.722188912	1.11298283	0.032729167	0.799261187	0.816807027	Well Define				
17	0.370467478	79.90804598	1	0.088619077	0.465235592	0.027959196	0.848001738	0.821091591	Well Define				
18	0.473542501	-634.5	1	-0.007342021	2.109901828	0.021436518	0.762699281	0.824198005	Well Define				
19	0.30163093	138.9641694	1	0.07810763	0.557178673	0.009037657	0.740047066	0.80743084	Well Define				
20	0.445818681	18.59876543	1	0.292907832	1.151147823	0.025849495	0.832989625	0.79789728	Well Define				
21	0.475398169	-38.34375	1	-0.137284996	1.523172632	0.027853313	0.84909354	0.791468727	Well Define				
22	0.475472482	-5.020217729	1	-0.861817051	1.03382696	0.035753502	0.801474336	0.833873157	Well Define				

Figure 9. Description of features value of the segmented image

3.2 Experiment Results

The experiment scenario is comparing the proposed system GSA-SVM with SVM using default parameter. Then, to get the result fairly, the data used to compare the system is testing set, while the training set used to train the GSA-SVM version.

Firstly, we test the default SVM parameter in all three different kernel using testing set, the result showed by Table 2.

Table 2. Experiment result using SVM default parameter

No	Kernel Name	Parameter			Accuracy
		C	d	γ	
1	Linear	1	-	-	41.4634 %
2	Polynomial	1	3	-	58.5366 %
3	RBF	1	-	0.5	41.4634 %

As can be seen from the Table 2, that the performance of the system is very low. As explained in the introduction section that the problem using SVM is how to choose the appropriate kernel and its parameters based on the data and the problem.

In SVM, different data, different problem required different parameters as well. This means SVM needs to change the parameters in every case to get the optimal solution. In this study, the search process for the best parameters for each kernel is done using GSA.

The scenario started with the polynomial kernel which has parameter C and d . In the training process, to get reliable result k -fold cross validation is applied where the k is 5. The training process is done by running it in different range of iteration and number of agents. Table 3 presents the result of GSA-SVM in polynomial kernel.

Table 3. GSA-SVM Polynomial kernel result

Kernel	Number of agent	Iteration	Fold	C	d	Accuracy	
						Training data (%)	Validation data (%)
GSA-SVM Polynomial	5	25	1	738.6248	6	62.2449	64
			2	124.8736	1	88.7755	76
			3	922.6634	9	58.1633	68
			4	904.3556	1	91.9192	62.5
			5	917.3361	2	100	58.3333
	10	25	1	99.3063	1	87.7551	72
			2	246.3117	10	55.102	68
			3	419.3443	2	100	68
			4	482.7784	2	100	66.6667
			5	834.3613	1	89.899	70.8333
	15	25	1	632.3596	1	89.7959	72
			2	557.0479	1	89.7959	84
			3	594.3899	2	100	68
			4	144.1742	2	100	66.6667
			5	895.4578	1	89.899	70.8333

It can be seen from the Table 3 that the highlighted value is the highest parameter obtained from the training set. The value of the parameter C is 557.0479, and the value of parameter d is 1 get the accuracy of the training set up to 89.79 % while the validation set reach 84 %.

Table 4. GSA-SVM RBF kernel result

Kernel	Number of agent	Iteration	Fold	C	γ	Accuracy	
						Training data (%)	Validation data (%)
GSA-SVM RBF	5	25	1	711.87	4	62.67	65
			2	747.7228	6	65	40
			3	360.7225	3	60	40
			4	294.3697	4	57.8	37.5
			5	619.5888	0.3	40.23	45.833
	10	25	1	422.5763	0.6	70.98	48
			2	623.3334	7	45.43	44
			3	843.9802	4	67.74	40
			4	670.3629	5	56.98	58.333
			5	309.7921	7	50.12	45.833
	15	25	1	655.741	2	78.67	52
			2	517.724	1	56.45	44
			3	463.0432	4	40.76	44
			4	330.8484	9	55.19	58.333
			5	236.9869	5	51.32	50

In RBF kernel, the optimum parameters obtained in the *fold-1* of the first experiment where the number of agent is 5 and Iteration is 25. The maximum accuracy of the training process is 62.67 % while the validation data reach 65 %. Eventhough it only slightly increased from the default one, it is a big improvement. For the final result, it presented in Table 5.

Table 5. Parameters obtained and system performance using GSA-SVM

No	Kernel Name	Parameters			Accuracy
		C	d	γ	
1	<i>Linear</i>	1	-	-	41.4634 %
2	<i>Polynomial</i>	945.71	1	-	89.4309 %
3	<i>RBF</i>	711.87	-	4	98.374 %

As can be seen from the Table 5, the best parameters obtained from the training processes and resulted high accuracy fro the system compare with the default parameter result in Table 2. The default polynomial parameter C and d (1, 3) resulted 58.54 % while after the training process best parameters obtained are (945.71, 1) resulted 89.43 %. Moreover, default RBF parameter C and γ (1, 0.5) resulted only 41.46%, while using parameters obtained from training process that are (711.87, 4) resulted 98.37 %.

4. CONCLUSIONS

Based on the experiment result, it can be concluded that Gravitational Search Algorithm (GSA) can be used to find optimal parameters in the SVM classifier. This is evidenced by the increased accuracy in the polynomial kernel from 58.5366% to 89.4309%, and 41.4634% to 98.374% in the RBF kernel.

REFERENCES

- [1] Z. Tao, L. Huiling, W. Wenwen, and Y. Xia, "GA-SVM Based Feature Selection and Parameter Optimization in Hospitalization Expense Modeling," *Appl. Soft Comput. J.*, vol. 73, pp. 456–466, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1568494618306264>. [Accessed: 20-Nov-2018].
- [2] M. Aladeemy, S. Tutun, and M. T. Khasawneh, "A New Hybrid Approach for Feature Selection and Support Vector Machine Model Selection Based on Self-adaptive Cohort Intelligence," vol. 88, pp. 118–131, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417417304475>. [Accessed: 20-Sept-2017].
- [3] S. Sarafrazi and H. Nezamabadi-Pour, "Facing the classification of binary problems with a GSA-SVM hybrid system," *Math. Comput. Model.*, vol. 57, no. 1–2, pp. 270–278, 2013. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S089571771100389X>. [Accessed: 29-Aug-2017].
- [4] B. R. Haugen, E. K. Alexander, K. C. Bible, G. M. Doherty, S. J. Mandel, Y. E. Nikiforov, F. Pacini, G. W. Randolph, A. M. Sawka, M. Schlumberger, K. G. Schuff, S. I. Sherman, J. A. Sosa, D. L. Steward, R. M. Tuttle, and L. Wartofsky, "2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer," *Thyroid*, vol. 26, no. 1, pp. 1–133, 2016. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/26462967>. [Accessed: 22-Sept-2017].

-
- [5] Y. J. Choi, J. H. Baek, H. S. Park, W. H. Shim, T. Y. Kim, Y. K. Shong, and J. H. Lee, "A Computer-Aided Diagnosis System Using Artificial Intelligence for the Diagnosis and Characterization of Thyroid Nodules on Ultrasound: Initial Clinical Assessment," *Thyroid*, vol. 27, no. 4, pp. 546–552, 2017. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/28071987>. [Accessed: 22-Sept-2017].
- [6] R. Gonzalez and R. Woods, *Digital image processing*, 4th Editio. Upper Saddle River, NJ, USA: Pearson, 2018.
- [7] C. Cortes and V. Vapnik, "Support-Vector Networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995. [Online]. Available: <https://link.springer.com/article/10.1007/BF00994018>. [Accessed: 20-Sept-2017].
- [8] E. Prasetyo, *Data Mining: Konsep dan Aplikasi menggunakan Matlab*, 1st ed. Yogyakarta: ANDI, 2013.
- [9] E. Rashedi, H. Nezamabadi-pour, and S. Saryazdi, "GSA: A Gravitational Search Algorithm," *Inf. Sci. (Ny)*, vol. 179, no. 13, pp. 2232–2248, 2009. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0020025509001200>. [Accessed: 20-Sept-2017].
- [10] E. S. Wulandari, "Kesesuaian Penentuan Fitur Bentuk, Tepi dan Orientasi Nodul Tiroid Jinak dan Ganas Antara Pembacaan Ultrasonografi oleh Dokter Spesialis Radiologi dan Computer Aided Diagnostic (CAD)," Tesis, Fakultas Teknik, Universitas Gadjah Mada, 2017. [Online]. Available: http://etd.repository.ugm.ac.id/index.php?mod=penelitian_detail&sub=PenelitianDetail&act=view&typ=html&buku_id=128402&obyek_id=4. [Accessed: 21-Sept-2017].