Quadratic tuned kernel parameter in Non-linear support vector machine (SVM) for agarwood oil compounds quality classification

Muhamad Addin Akmal Bin Mohd Raif[1], Nurlaila Ismail[2], Nor Azah Mohd Ali[3], Mohd Hezri Fazalul Rahiman[4], Saiful Nizam Tajuddin[5], Mohd Nasir Taib[6]
[1,2,4,6]Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM), Malaysia
[3]Herbal of Product Development Programme, Natural Product Division,
Forest Research Institute Malaysia (FRIM), Malaysia
[5]Bio Aromatic Research Centre of Excellent, Universiti Malaysia Pahang, Malaysia

Article Info

ABSTRACT

This paper presents the analysis of agarwood oil compounds quality classification by tuning quadratic kernel parameter in Support Vector Machine (SVM). The experimental work involved of agarwood oil samples from low and high qualities. The input is abundances (%) of the agarwood oil compounds and the output is the quality of the oil either high or low. The input and output data were processed by following tasks; i) data processing which covers normalization, randomization and data splitting into two parts in which training and testing database (ratio of 80%:20%), and ii) data analysis which covers SVM development by tuning quadratic kernel parameter. The training dataset was used to be train the SVM model and the testing dataset was used to test the developed SVM model. All the analytical works are performed via MATLAB software version R2013a. The result showed that, quadratic tuned kernel parameter in SVM model was successful since it passed all the performance criteria's in which accuracy, precision, confusion matrix, sensitivity and specificity. The finding obtained in this paper is vital to the agarwood oil and its research area especially to the agarwood oil compounds classification system.

Corresponding Author:

Nurlaila Ismail,
Faculty of Electrical Engineering,
Universiti Teknologi MARA (UiTM), Malaysia.
Email: nurlaila@ieee.org

1. INTRODUCTION

The Agarwood or gaharu is the prominent name for resinous heartwood, where it is precious natural oil and from 'wounded/infected' Aquilaria trees. It also known as Oud oil, Aloeswood oil and Eaglewood, depends on its origin. The originated tree is located at North Eastern India, Bhutan and parts of South East Asia – especially Vietnam, Cambodia, Philippines and Indonesia [1]. The agarwood and its application has a wide history especially in medicinal manufacture, as an incense, and in perfumery elements [2]. Other than that, the agarwood oil also used as a symbol of wealth especially in the Middle East Countries, and being burnt during religious as well as wedding ceremony [2-5].

The enthusiasm for the utilization of the agarwood oil is expanding because of its special scent[3]. The discount cost for amazing agarwood oil is around US$30,000– US$50,000 per liter, by alluding to its aroma quality, lifespan, tar substance, and oil virtue [6-7]. The exchanged cost in view of the nature of the agarwood oil which is a high cost for high calibre and low cost for low quality. Expectedly, the evaluating techniques for agarwood oil is done in light of its physical properties; shading, scent, high fixative and

customer recognition. Normally, dim shading and dependable scent are delegated high calibre and sold at a chief cost [7]. Agarwood oil has been found tp be differentiated by two terms. A few countries utilize the term of high and low quality and another countries utilize term of grade; A, B, C and D. Each review is grouped in light of agarwood oil physical properties like shading, scent and purchaser recognition [2, 8].

There are several techniques performed to identify the quality of the agarwood oil [9-10]. Researchers from Japan have regarded Kanankoh as the highest and Jinkoh as the low quality [12-14]. The relationship between high and low quality of oil has been widely investigated by comparing their chemical compositions. One of the technique is used by distillation method where it is the earliest and widely used in extracting essential oil since it is economically usable and safety in use [15]. But this method has several disadvantages where the research came out with lack of documentation and research publication. Sometime, the process to get the result for standard quality control and characteristic for active compound in agarwood oil is disrupted because of adulteration occur during manufacturing standardize [15].

SVM is additionally one of the directed learning techniques utilized for regression and outliers detection [16]. In this technique, one maps the data into a higher dimensional info space and one builds an ideal isolating hyperplane in this space. SVM can be partitioned into two gatherings, linear SVM, and non-linear SVM. These two gatherings have an alternate strategy to isolate the information by utilizing hyper-plane and portion trap, separately. One of the fundamental elements of SVMs is the purported kernel trick where the calculation of dot products in high-dimensional component spaces utilizing basic capacities characterized by sets of information designs. This trick permits the definition of a nonlinear example of a calculation that can be thrown as far as dot products [17-22].

As mentioned before, SVM have two groups which is Linear SVM and Non-Linear SVM. The agarwood oil compound is a complex signal and the data cannot be divided by linear SVM classifier. Thus, there is a need to use a non-linear SVM technique for get the better classification result [18]. The way is projection the data on to a higher dimensional space with a kernel trick so the data can be isolated utilizing hyper-plane (nonlinear separator). In the event that we utilize a quadratic capacity to outline data into two measurements, at that point will less demanding for a nonlinear separator to characterize the data on the off chance that we utilize a quadratic capacity to delineate data into two measurements (a polar directions projection would be another probability). Kernel function is such a capacity, to the point that relates to a speck item in some extended component space [19]. The quadratic kernel is under family polynomial in two measurements. It similarity is for polynomial if d =1 then called as linear. In the event that d = 2, at that point called quadratic where is it normally utilized Quad-kernel for 2-dimensional vectors as given by [19, 24-25],

$$\vec{u} = (u\_(1) \ u\_2) \tag{1}$$

$$\vec{v} = (v\_(1) \ v\_2) \tag{2}$$

Consider $K(\vec{u}, \vec{v}) = (1 + \vec{u}^T \vec{v})^2$. To show that this is a kernel, i.e., that $K(\vec{u}, \vec{v}) = \phi \vec{u}^T \phi \vec{v}$ and $\phi(\vec{u}) = (1 \ u_1^2 \ \sqrt{2}u_1 u_2 \ u_2^2 \ \sqrt{2}u_1 \ \sqrt{2}u_2)$ [19]

The performance measure used in the study is the confusion matrix. It is including accuracy, sensitivity, precision and specificity. The confusion matrix divided into two part which the matrix row denotes the actual class and the matrix column shows the predicted class [2]. Four elements that involved in this measurement are true positives (tp), true negatives (tn), false positives (fp) and false negatives (fn). A confusion matrix is tabulated in Table 1.

Table 1. A Confusion Matrix [3]

DATA CLASS	Classified as Positive	Classified as Negative
Positive (+)	True positive (*tp*)	False negative (*fn*)
Negative (-)	False positive (*fp*)	True negative (*tn*)

in which; tp is the number of correctly classified to the class examples, tn is the number of correctly identified to the not class examples, fp is the example incorrectly identified to the class examples and fn is the examples incorrectly identified to the not class examples

2. RESEARCH METHOD

The 96 samples of agarwood oil for the project were obtained from the Forest Research Institute Malaysia (FRIM), Malaysia and Bio Aromatic Research Centre of Excellent (BARC), University Malaysia Pahang, Malaysia. Then, the identification of the agarwood oil chemical compound was done by using SVM. The sample preparation and compounds extraction were done by the previous researcher [6-7].

Figure 1 shows the flow chart for preparatory planned in this project. It start with the data collection from input and output data that have been done by previous researcher [6-7]. Then, follow by the data processing. At this stage, the process was normalizing the data, randomize the data and then split the data to training and testing dataset with the ratio of 80%: 20% respectively. Next is the SVM model development using training dataset was carried out in tuning quadratic kernel parameter and undergo the test for SVM model development by using testing dataset. The model should be able to pass all the performance criteria which are accuracy, confusion matrix, sensitivity, precision and specificity. If the model passed, the model will be accepted and if it is not, it will need to undergo data pre-processing precise as shown in the figure.
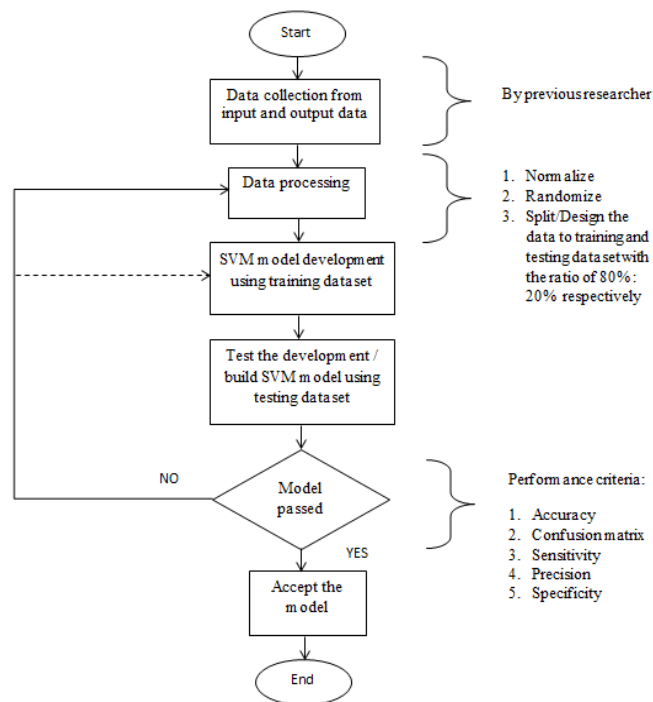

Figure 1. Flowchart of Experimental Set-up

Data collection used in this study consists of 96 sample of agarwood oil in term of abundances (%). The data are from high and low quality of three oil, and also belongs to seven compounds which are coded as C1 to C7. It is noted as C1 is β-agarofuran, C2 is α-agarofuran, C3 is 10-epi-Υ-eudesmol, C4 is Υeudesmol, C5 is longifolol, C6 is hexadecanol and C7 is eudesmol. Figure 2 shows the coding for pre-processing. After load the data of the sample of agarwood oil, the data is splitted into training and testing data set with the ratio of 80% and 20%, respectively. Before it is being splitted, the data has gone normalization and randomization process. After that, the data is fed to SVM for modelling them into two classes (high and low quality) of the oil.

```
xdata = dataAnn(:,1:7);
group = dataAnn(:,8);
P = cvpartition (group, 'Holdout', 0.20);
xtrain = xdata(P.training,:);
ytrain = group(P.training);
xtest = xdata(P.test,:);
ytest = group(P.test);
```

Figure 2. The Matlab Script for Load and Arrange the Data in Matlab

The training data set and testing data set were splitted from the data collection that consists of 96 samples with the ratio of 80% and 20%, respectively. It means that for training data set it has 77 samples while for testing data it has 19 samples. The different between these two data sets is training data was used to create a model while testing data set was used to predict the output based on the model. This ratio was used because the bigger data train is better to build a model.

Figure 3 shows the coding for SVM model using quadratic kernel function. After design into training and testing data set, SVM model was build using training data set. Because of the agarwood oil compound is a complex signal, the quadratic kernel trick was used in this method. Quadratic kernel function transforms the data into higher dimensional feature space to perform the linear separation.

```
SVMStruct = svmtrain(xtrain,ytrain,'kernel_function','quadratic');
```

Figure 3. The Matlab Script for SVM model using Quadratic Kernel

Figure 4 shows the coding for SVM classify that using the information in a SVM model that built before. This step is to get the predicted output. Each row of the testing data is test on the model and classified them either group 1(low quality) or group 2 (high quality) as shown in the Table 3.

```
yhattest = svmclassify(SVMStruct,xtest);
```

Figure 4. The Matlab Script for SVM Classify

3. RESULTS AND ANALYSIS

The finding from SVM revealed 14 support vectors from the training dataset and it is tabulated in Table 2. Basically, the support vector belongs to seven compounds coded as C1 to C7. It is noticed that the values are within -1.38 to 3.66 for C1 to C7. Specifically, for C1, the minimum vector is -0.74 and the maximum vector is 1.08. For C2, the minimum vector is -0.76 and the maximum vector is 1.57. For C3, the minimum vector is -1.16 and the maximum vector is 0.67. For C4, the minimum vector is -1.38 and the maximum vector is 1.02. For C5, the minimum vector is -0.31 and the maximum vector is 3.39. For C6, the minimum vector is -0.38 and the maximum vector is 3.66. For C7, the minimum vector is -0.37 and the maximum vector is 3.19. Overall, it can be seen that C6 has the highest vector that is 3.66.

Table 2. Support Vector for C1 to C7

Vector	C1	C2	C3	C4	C5	C6	C7
1	-0.74	-0.76	0.67	-1.38	-0.31	-0.38	-0.37
2	-0.55	-0.76	-0.99	-1.04	-0.31	-0.38	-0.37
3	1.08	1.57	0.48	1.02	-0.29	-0.33	-0.35
4	-0.55	-0.72	-0.99	-1.04	-0.30	-0.32	-0.33
5	-0.47	-0.62	-1.16	-0.73	3.39	3.66	3.19
6	-0.63	-0.16	-1.13	-0.78	-0.30	-0.24	0.76
7	-0.72	-0.75	0.67	-1.38	-0.30	-0.27	-0.33
8	-0.54	-0.75	-0.10	-1.04	-0.29	-0.38	-0.37
9	-0.70	-0.74	-0.04	-1.37	-0.27	-0.25	-0.37
10	-0.72	-0.73	0.67	-1.37	-0.28	-0.27	-0.36
11	-0.67	-0.60	0.66	-1.36	-0.26	-0.21	-0.36
12	-0.70	-0.75	-0.02	-1.37	-0.27	-0.35	-0.34
13	-0.74	-0.61	0.66	-1.35	-0.31	-0.21	-0.35

Table 4 (a) tabulates the result about the performance measure used in the research that is quadratic tuned kernel parameter. The value of true positive in quadratic is 4 and the value of false negative is 0. In other side, the value for false positive is 0 while true negative is 15. It means that from all the nineteen numbers of samples, there is four numbers of samples has strongly predicted to group 1 (low quality) and the other fifteen numbers of sample has strongly predicted to group 2 (high quality). By observation in Table 4 (b), it is clearly showed that all the classification behaviour which are the accuracy, sensitivity,

specificity and precision achieved 100%. Overall, the result shows high performance that related to SVM technique was proven as a good classifier for agarwood oil classification.

Table 3. Predicted Output Data

No. of samples	Group
1	2
2	1
3	2
4	2
5	1
6	2
7	2
8	2
9	2
10	1
11	2
12	2
13	2
14	2
15	1
16	2
17	2
18	2
19	2

Table 4. The Classification Performance Measures for Testing Data; (a) confusion matrix and (b) Accuracy, Sensitivity, Specificity and Precision

Data Class	Predicted as positive	Predicted as negative
Low quality	4 (tp)	0 (fn)
High quality	0 (fp)	15 (tn)

(a)

Parameter	Value
Accuracy (ACC)	100%
Sensitivity (SENS)	100%
Specificity (SPEC)	100%
Precision (PREC)	100%

(b)

Table 5 tabulates the test target, predicted test target, prediction errors and its m.s.e. In general, the prediction errors for test target are zero error for nineteen target data that leads zero error for m.s.e.

Table 5. Test Target, Predicted Test Target, Prediction Errors and its m.s.e

Test Target (ytest)	Predicted Test Target (yhattest)	Prediction Error (errTest)	Mean Square Error (m.s.e)
2	2	0	
1	1	0	
2	2	0	
2	2	0	
1	1	0	
2	2	0	
2	2	0	
2	2	0	
2	2	0	
1	1	0	0
2	2	0	
2	2	0	
2	2	0	
2	2	0	
1	1	0	
2	2	0	
2	2	0	
2	2	0	
2	2	0	

4. CONCLUSION

The study in this paper has successfully evaluated the agarwood oil quality to high and low using Quadratic Kernel parameter. The technique was chosen because it was one of the most ideal techniques for classification system of agarwood oil by implementing the abundances of significant chemical compound as

the input and agarwood oil quality as output. The data was extracted using Support Vector Machine (SVM). The result showed that this technique comes with 100% accuracy and passed other performance measures; sensitivity, specificity, precision and confusion matrix. The finding in this study was significance thus it will benefit other future work especially in agarwood oil classification system.

REFERENCES

[1] E.Botanicals. (2017). Agarwood Essential Oil. Available: https://www.edenbotanicals.com/agarwood-oil-oud-aloeswood.html

[2] R. Naef, "The volatile and semi-volatile constituents of agarwood, the infected heartwood of Aquilaria species: a review," *Flavour and Fragrance Journal*, vol. 26, pp. 73-87, 2011.

[3] Takemoto, Hiroaki, Ito, Michiho, Shiraki, Tomohiro, Yagura, Toru and Honda, Gisho, "Sedative effects of vapor inhalation of agarwood oil and spikenard extract and identification of their active components", *Journal of Natural Medicines,* Vol. 2, Issue 1, pg. 41-46, 2008.

[4] Chen, Huai-Qiong, Wei, Jian-He, Yang, Jun-Shan, Zhang, Zheng, Yang, Yun, Gao, Zhi-Hui, Sui, Chun, Gong, "Chemical constituents of agarwood originating from the endemic genus Aquilaria plants", *Chemistry & Biodiversity,* Vol. 9, Issue 2, pp. 236-250, 2012.

[5] N Sulaiman, M Ida Idayu, and AZ Ramlan, "Effects Of Extraction Methods on Yield and Chemical Compounds of Gaharu (Aquilaria Malaccensis), *Journal of Tropical Forest Science (JTFS)*, Vol. 27, No. 3, pp. 413-419, 2015.

[6] W. Hidayat, A. Y. M. Shakaff, M. N. Ahmad, and A. H. Adom, "Classification of agarwood oil using an electronic nose," *Sensors*, vol. 10, pp. 4675-4685, 2010.

[7] Barden, N. A. Anak, T. Mulliken, and M. Song, "Heart of the matter: agarwood use and trade and CITES implementation for Aquilaria malaccensis," TRAFFIC International, Cambridge, UK, 2000.

[8] G. A. Persoon and H. H. Beek, "Growing 'the wood of the gods': agarwood production in southeast Asia," Smallholder Tree Growing for Rural Development and Environmental Services, pp. 245-262, 2008.

[9] M. Nor Azah, Y. Chang, J. Mailina, S. Saidatul Husni, H. Nor Hasnida, and Y. Nik Yasmin, "Comparison of chemical profiles of selected gaharu oils from Peninsular Malaysia," *Malaysian Journal of Analytical Sciences*, vol. 12, pp. 338-340, 2008.

[10] N. A. B. Sidik, "Comparison of Gaharu (Aquilaria Malaccensis) Essential Oil Composition Between Each Country," *UMP*, 2008.

[11] M. Ishihara, T. Tsuneya, M. Shiga, and K. Uneyama, "Three sesquiterpenes from agarwood," *Phytochemistry*, vol. 30, pp. 563-566, 1991.

[12] B. Scholkopf and A. J. Smola, "Learning with kernels: support vector machines, regularization, optimization, and beyond: MIT press", 2001.

[13] M. Ishihara, T. Tsuneya, and K. Uneyama, "Fragrant sesquiterpenes from agarwood," *Phytochemistry*, vol. 33, pp. 1147-1155, 1993.

[14] M. Ishihara, T. Tsuneya, and K. Uneyama, "Components of the agarwood smoke on heating," *Journal of Essential Oil Research*, vol. 5, pp. 419-423, 1993.

[15] M. Ishihara, T. Tsuneya, and K. Uneyama, "Components of the volatile concentrate of agarwood," *Journal of Essential Oil Research*, vol. 5, pp. 283-289, 1993.

[16] S. F. Ab Rahman, "Analysis of agarwood oil composition via preparative thin layer chromatography," *UMP*, 2009.

[17] N. Cristianini and J. Shawe-Taylor, "An introduction to support vector machines and other kernel-based learning methods", Cambridge university press, 2000.

[18] B. Schèolkopf, C. J. Burges, and A. J. Smola, "Advances in kernel methods support vector learning," ed: Cambridge, Mass., MIT Press, 1999.

[19] Shin, Kyung-Shik, Lee, Taik Soo and Kim, Hyun-jung "An application of support vector machines in bankruptcy prediction model", *Expert systems with applications*, Vol 8, Issue 1, pp. 127-135, 2005.

[20] Suykens and Johan AK , "*Nonlinear modelling and support vector machines*", in Proceedings of the 18th IEEE instrumentation and measurement technology conference (IMTC 2001), Vol. 1, pp. 287-294, 2001.

[21] Qiang Wu and Ding-Xuan Zhou, "SVM Soft Margin Classifiers: Linear Programming versus Quadratic Programming", *Neural Computation*, Vol. 17, Issue 5, pp.1160-1187,2005.

[22] Taku Kudo, "*Fast methods for kernel-based text analysis*", in Proceedings of the 41st Annual Meeting on Association for Computational Linguistics, Vol. 1, pp. 24-31, 2003.

[23] Xiaofei Zhoua, Wenhan Jiang, Yingjie Tiana and Yong Shi, "Kernel subclass convex hull sample selection method for SVM on face recognition", *Neurocomputing*, Vol. 73, pp. 2234–2246, 2010.

[24] B. G. Lindsay, M. Markatou, and S. Ray, "Kernels, degrees of freedom, and power properties of quadratic distance goodness-of-fit tests," *Journal of the American Statistical Association*, vol. 109, pp. 395-410, 2014.

[25] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, pp. 293-300, 1999.