Malaysian Journal of Analytical Sciences (MJAS)
Published by Malaysian Analytical Sciences Society

KAFFIR LIME OIL QUALITY GRADING USING NON-LINEAR SUPPORT VECTOR MACHINE WITH DIFFERENT KERNELS

(Penentuan Kualiti Minyak Limau Purut Dengan Menggunakan Mesin Vektor Sokongan Bukan Linear Dengan Kernel Berbeza)

Nor Syahira Jak Jailani[1]*, Zuraida Muhammad[2], Nor Salwa Damanhuri[2], Muhd Hezri Fazalul Rahiman[1], Mohd Nasir Taib[1]

[1]College of Engineering,
Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[2]College of Engineering,
Universiti Teknologi MARA,
13500 Permatang Pauh, Pulau Pinang, Malaysia

**Corresponding author: norsyahira_jakjailani@yahoo.com*

Received: 23 September 2021; Accepted: 5 February 2022; Published: 28 April 2022

Abstract

Nowadays, Kaffir lime oil is one of the highly demand in the industries and sold with various market prices. But sometime the most expensive price of kaffir lime oil does not guarantee the best quality of this oil. Currently, kaffir lime oil quality grading is based only on human sensory evaluation such as smell and visual which accompanying with confusion and inconsistent result. This is due to human sensory evaluations have limitations such as easily fatigue and unable to manage mass products of the oil in time. This paper presents the classification of significant chemical compound of kaffir lime oil in the oil quality grading by using Non-Linear Support Vector Machine (NSVM). The objective of this study is to classify the quality of kaffir lime oil whether it is high or low in quality only by tuning the two different kernels into NSVM. This project has used up to 90 samples of kaffir lime oil data from various high to low quality to prepare NSVM with two different kernels. The implementation of this project was performed by using MATLAB version R2020A. The result appeared NSVM model with RBF kernel is better than NSVM model with Polynomial kernel. It was discovered that RBF kernel was able to generate 100% accuracy, specificity, precision, and sensitivity compared to Polynomial kernel. Since the finding and outcome were effective and significant, thus this study contributes a lot of benefits for future study especially in kaffir lime field.

Keywords: non-linear support vector machine, radial basis function, polynomial, kaffir lime oil, high/low quality

Abstrak

Saat ini, minyak limau purut mendapat permintaan tinggi di industri, dan dijual dengan harga yang berbeda di pasar. Tetapi harga minyak limau purut yang mahal tidak menjamin kualiti tinggi minyak ini. Pada masa ini, penilaian kualiti minyak limau purut berdasarkan penilaian deria seperti hidung dan mata menimbulkan kekeliruan dan hasil yang tidak konsisten. Ini kerana penilaian deria mempunyai batasan dan mudah letih dan tidak dapat menangani banyak sampel sekaligus. Oleh itu, kajian inimembentangkan klasifikasi sebatian kimia yang penting dalam minyak limau purut untuk gred kualiti minyak dengan

menggunakan Mesin Vektor Sokongan Bukan Linear (NSVM). Objektif kajian ini adalah untuk mengklasifikasikan kualiti minyak limau purut sama ada berkualiti tinggi atau rendah dengan menyesuaikan dua kernel yang berbeza ke dalam NSVM. Projek ini menggunakan 90 sampel data minyak limau purut dari berkualiti tinggi hingga rendah untuk melatih NSVM dengan kernel yang berbeza. Pelaksanaan projek ini dilakukan dengan menggunakan MATLAB versi R2020A. Hasilnya menunjukkan model NSVM dengan kernel RBF lebih baik daripada model NSVM dengan kernel Polinomial. Didapati kernel RBF mampu menghasilkan ketepatan, kekhususan, ketepatan, dan kepekaan 100% berbanding dengan kernel Polinomial. Penemuan dan hasil ini berkesan dan signifikan sekali gus akan mendorong dan memberi banyak faedah untuk kajian masa depan terutama dalam bidag limau purut

Kata kunci: mesin vektor sokongan bukan linear, fungsi asas radial, polinomial, minyak limau purut, kualiti tinggi/rendah

Introduction

Kaffir lime oil or scientifically known as Citrus Hystrix which is originate from Rutacae family can be found in most Asian Countries. It was widely consumed in Asian dishes and other few industries such as aromatherapy, medicines, and fragrances. Despite all, Kaffir lime oil can also be obtained from its peels and leaves. Currently, the oil quality grading was performed manually based on human sensory evaluation. It was observed based on the physical properties of the samples. Human experienced and perception on the oil long-lasting aroma, color and odor is the only method to classify the quality of the oil. Natural sensory quality evaluation method like this produces less accuracy and subjective as it may vary from everyone. Usually, the result performed by natural sensory train grader were not consistent since it may vary to each other and the process itself take a high time consuming in the procedure [1-3]. Many researchers encountered that essential oil can be classified based on their chemical compound. This method of classification is more accurate compared to the human sensory evaluation [4]. The chemical compounds of essential oil need to be analyzed in order to classify the grading of their quality [5].

The quality of kaffir lime oil can be determined based on their abundances of major chemical compound. Many researchers found the major chemical compound of kaffir lime oil was monoterpene components. The monoterpene components discovered were Citronellal, Limonene, β-pinene, α-pinene and Sabinene [6-9] Besides that, caryophyllene was also identified by Ngan et al., researcher from Vietnam [10] and terpinene-4-ol was found by Hongatanaworkit & Buchbauer as another major compounds of kaffir lime oil [11].

Furthermore, to overcome the inconsistent result of classifying oil grading and the amount time consuming, this project proposed a latest technique of classifying major compounds in Kaffir lime essential oil using Non-Linear Support Vector Machine (NSVM) to perform the automated grading. A SVM is a method in supervised learning that usually used for classification. SVM is used to overcome the problem of classification and regression with linear or non-linear data [12, 13]. However, NSVM will be applied when the data consumed can't be easily separated with linear line. Therefore, the kernels ought to be used to make non-separable data into separable data. Kernels is used to map out the data from the vector input into the high dimension vector space [12]. In this project, the RBF kernel and polynomial were choosing to be tuning into NSVM. The abundance percentages (%) of major compounds act as an input and will be supplied to NSVM model develops by kernel. The output will determine the quality of kaffir lime oil whether it is low or high in quality. Hence, the oil quality and grading of kaffir lime will be classified by using an automated intelligent nonlinear support vector machine (NSVM) which regards to their chemical compounds and reduce the time-consuming during grading processes.

Materials and Methods

Sample preparations

The 90 samples data of kaffir lime oil that contains high and low in quality were used in this study that contains high and low quality. These data were

obtained from 15 samples kaffir lime oil before applying synthetic data. 12 samples of kaffir lime oil obtained in the market with different various brands. All 12 brands kaffir lime oil was sent to Forest Research Institute Malaysia (FRIM) to extract the chemical compound by applying gas chromatography-mass spectrometry (GC-MS). It was found kaffir lime oil has almost 96 chemical compounds. Another three samples kaffir lime oil data get from previous researcher[7, 9, 14].

Z-score technique

The Z-score technique was utilised to group the data and assign an explicit grouping to the data population sample in this study. Previous studies have used Z-score approaches to establish a specific set, with Z-score being a widely used application for standardising data and calculating a factor score. Thus, the Z-score technique was applied to identify the significant chemical compound of kaffir lime oil [4]. In this technique, the two significant chemical compounds of kaffir lime oil were found, which were β-pinene, limonene, linalool, citronellal, and terpinene-4-ol. The five chemical compounds were used as a marker compound for the NSVM classification method. Table 1 shows samples kaffir lime oil was used in this project and its significant chemical abundance (%). There are 11 HIGH samples and 4 LOW samples of kaffir lime oil. Samples AA, AB and AC got from previous researcher that were extract directly from kaffir lime, so they were labelled as a high quality. Another. 8 samples HIGH kaffir lime oils are traded as high quality by certified trader and some of them were published as high-quality kaffir lime oil

The quality of kaffir lime oil was determined based on its chemical abundance (%). The higher the abundance(%) of this chemical, the higher quality of kaffir lime oil [6, 8, 15–17].

Table 1. Significant Kaffir lime chemical constituents used for classification (Group' 1' for HIGH quality and '2' for LOW quality

	Abundance (%)					**Group**
	β-pinene	**Limonene**	**Linalool**	**Citronellal**	**Terpinen-4-ol**	
AA	16.8	19.8	1	7.8	2.8	1
AB	8.974	28.649	0.541	8.29	0.47	1
AC	9.321	26.446	0.687	6.63	1.0541	1
WS	1.42	0.52	7.19	60.06	0.99	1
FF	0.6	0.46	5.68	54.73	0.62	1
T	18.13	22.7	0.44	0	11.15	1
SS	0.56	0.37	7.4	79.39	0	1
SE	10.47	15.06	1.85	28.32	5.98	1
S	16.37	24.39	0.38	0	10.84	1
NL	1.16	0.42	7.35	60.65	0.85	1
DE	16.19	23.01	0.8	4.45	9.93	1
SP	0	16.8	2.48	0	0	2
C	0.19	18.49	2.14	0	0	2
B	0	0	12.38	51.05	0	2
LL	2.19	50.61	7.14	0	0	2

Synthetic data

All 15 data was added with synthetic data to become 90 samples data to increase the performances classification method[18]. Synthetic data was created by adding the SNR signal to the original input-output data. The synthetic data equations can be represented as follows, as recommended by Sevgi [19]:

$$Ysynthetic = Y_{\mathrm{NSVM}} + Y_{SNR} \quad (1)$$

The original NSVM input-output data is Y_{NSVM}, and the SNR signal or random noise is Y_{SNR}. The trend of the original data and original data with synthetic data should be the same. The calculation minimum, maximum and median of synthetic data are done to support the trend similarity.

Data pre-processing

Data pre-processing done for this experiment consists of a holdout cross-validation strategy for splitting data. The data is divided into ratios 80%:20% for training data and for testing data that was recommended by the authors [5, 20-22], which were for training data set it has 72 samples while the testing data has 18 samples. The ratio is applied in that manner due to the higher number of the training data collected, hence the higher accuracy results appeared on the model [22]. On the other hand, the difference between these two data sets is that the training data is used in model creation while the test data is performed as the model output prediction.

Build the NSVM classifier model

In the real world, data is rarely linearly separable. Most are non-linear. NSVM is modified by entering the kernel functions into SVM. There are a few types of kernel functions of NSVM: polynomial, quadratic, RBF, and sigmoid [23]. When using NSVMs in practice, it's especially vital to pick the correct kernel parameters. Thus, this experiment will test two different kernels to find the best kernel for modelling the quality kaffir lime oil grading. The two different kernels are the Polynomial kernel and RBF kernel. In general, the RBF kernel is a good starting point. This kernel nonlinearly maps samples into a higher-dimensional space to handle the case where the relationship between class labels and attributes is non-linear, unlike the linear kernel[24]. The polynomial kernel is selected because this kernel gives 100% accuracy for agarwood oil grading in a previous study [20] compared to the RBF kernel that is given only 89.5% [25].

The compounds gathered from the 90 samples of kaffir lime oil were used to train NSVM with Polynomial kernel and RBF kernel. The abundances (%) of the chemical compounds act as input, and this input were fed into NSVM model development, and hence the output is set to be either '1' or low '2' quality. All the analytical work was performed automatically by using MATLAB software version R2020A. Figure 1 shows the flowchart of NSVM modelling.

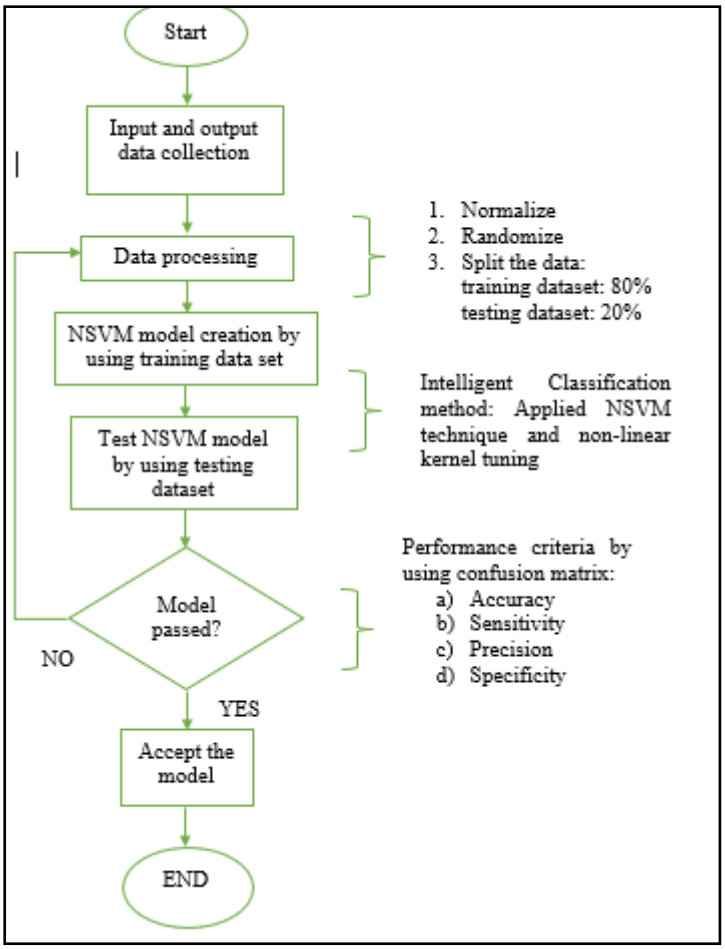

Figure 1. Flowchart NSVM modelling

Performance's measure

The performances for NSVM modelling were measured by using the confusion matrix. Other than that, measurements such as accuracy, sensitivity, specificity, precision, and Mean Absolute Error (MAE) would also be used to obtain more granular and precise model-building results. The accuracy (ACC) is defined as the effectiveness of the classifier. The calculation of the accuracy is as follows [29]:

$$ACC=(TP+TN)/(TP+FN+FP+TN) \quad (2)$$

Sensitivity (SENS) is defined as the rate of a true positive. The effectiveness of a classifier to identify the high quality of kaffir lime is as follow [29]:

$$SENS=TP/(TP+FN) \quad (3)$$

The specificity (SPEC) is interpreted as the rate of the negative. The effectiveness of a classifier to identify the low quality of kaffir lime is as follows [29]:

$$SPEC=TN/(TN+FP) \quad (4)$$

Precision (PREC) is interpreted how often it is correct when it predicts high quality of kaffir lime oil as follow:

$$PREC=TP/(TP+FP) \quad (5)$$

Results and Discussion

The confusion matrix result is shown in Table 2 for the polynomial kernel and Table 3 for the RBF kernel. Where TP (True positive): Classified as HIGH quality and the actual was HIGH quality, FP (False positive): Classifies as HIGH quality but the real is LOW quality, FN (False negative): classified as HIGH quality, but the actual was LOW quality and TN (True negative): classified as LOW negative and the actual is LOW quality. From the result in table 2 and table 3, the value of TP for Polynomial and RBF were 14 and the value of FN for both kernels is also the same, which is 0. FP value for table 1 was 4, while FP value in table 2 was 0. TN value in table 1 was 0 while TN value in table 2 was 4. This result shows that RBF kernel is better than the polynomial kernel because the best model does not have value of FP and FN.

Table 4 shows the comparison performances between the polynomial kernel with RBF kernel. It is clearly showing that RBF kernel is way better than polynomial. RBF kernel can reach up to 100% of its accuracy, sensitivity, specificity, precision, and 0% mean absolute error (MAE) compared to the polynomial kernel which only manage to achieve 100% in its sensitivity, 66.78% accuracy precision, and 0% specificity, and 0.22 for MAE. In general, the lower the MAE, the better a model's ability to match a dataset. We may compare the MAE of two distinct models to see which one delivers a better fit to a dataset when comparing them.

Table 2. Confusion matrix for polynomial

Data Class N=18	**Classified as HIGH**	**Classified as LOW**
actual HIGH	TP=14	FP=4
actual LOW	FN=0	TN=0

Table 3. Confusion matrix for RBF

Data Class N=18	**Classified as HIGH**	**Classified as LOW**
actual HIGH	TP=14	FP=0
actual LOW	FN=0	TN=4

Table 4. Confusion performances between polynomial kernel and RBF kernel

Performances	**Polynomial (%)**	**RBF (%)**
Accuracy	66.78	100
Sensitivity	100	100
Specificity	0	100
Precision	66.78	100
MAE	0.22	0

Conclusion

This paper has successfully evaluated the Kaffir lime oil quality either high or low by tuning in polynomial kernel and radial basis function (RBF) kernel into NSVM. The result shows RBF kernel was able to produce 100% in its accuracy, specificity, precision, and sensitivity compared to Polynomial kernel which was able only to produce 77.78% in its accuracy, 100% sensitivity, 0% specificity and 77.78% precision. Hence it is also can be concluded that RBF kernel parameter better than polynomial for kaffir lime oil grading classification. Since the finding and outcome were effective and significant, thus this study contributes a lot of benefits for future study especially in kaffir lime field.

Acknowledgements

A highly appreciate this research funder, the Malaysia Ministry of Higher Education's Fundamental Research Grant Scheme (FRGS-RACER). This research was supported by FRGS Grant no 600-IRMI/FRGS-RACER 513(115/2019) for my financial support. The authors wish to thank Universiti Teknologi MARA, Cawangan Pulau Pinang and Faculty of Electrical Engineering, Universiti Teknologi MARA, Cawangan Pulau Pinang for their endless technical assistance.

References

1. Abas, M. A., Zubir, N. S. A., Ismail, N., Yassin, I. M., Ali, N. A. M., Rahiman, M. H. F. and Taib, M. N. (2017). Agarwood oil quality classifier using machine learning. *Journal of Fundamental and Applied Sciences*, 9(4): 62-76.
2. Haron, M. H., Taib, M. N., Ismail, N., Mohd Ali, N. A. and Tajuddin, S. N. (2019). Statistical analysis of agarwood oil compounds based on GC-MS data. *9th IEEE Control System Grad. Research Colloquium:* pp. 27-30.
3. Azah, M. N., Husni, S. S., Mailina, J., Sahrim, L., Majid, J. A. and Faridz, Z. M. (2013). Classification of agarwood (gaharu) by resin content. *Journal of Tropical Forest Science*, 2013: 213-219.
4. Ismail, N., Azah, M. A. N., Jamil, M., Rahiman, M. H. F., Tajuddin, S. N. and Taib, M. N. (2013). Analysis of high quality agarwood oil chemical compounds by means of SPME/GC-MS and Z-score technique. *Malaysian Journal of Analytical Sciences*, 17(3): 403-413.
5. Jantan, H., Yassin, I. M., Zabidi, A., Ismail, N. and Ali, M. M. (2017). Differentiation of agarwood oil quality using support vector machine (SVM). *Journal Engineering Applied Sciences*, 12(15): 3810-3812.
6. Mohd Sauid, S. and Aswandi, F. A. (2018). Extraction methods of essential oil from kaffir lime (*Citrus hystrix*): A review. *Malaysian Journal of Chemical Engineering and Technology*, 1: 56-64.
7. Kasuan, N., Muhammad, Z., Yusoff, Z., Fazalul Rahiman, M. H., Taib, M. N. and Haiyee, Z. A. (2013). Extraction of *Citrus hystrix* DC (kaffir lime) essential oil using automated steam distillation process: analysis of volatile compounds. *Malaysian Journal of Analytical Sciences*, 17: 359-369.
8. Haiyee Z. A. and Winitkitcharoen, C. (2012). Extraction of volatile oil from kaffir lime leaves (*Citrus hystrix*) using pressurize liquid extraction. *International Journal Food, Nutrition Public Health*, 5: 201-210.
9. Muhammad, Z., Yusoff, Z. M. and Nordin, M. N. N. (2013). Steam Distillation with induction heating system: analysis of kaffir lime oil compound and production yield at various temperatures. *Malaysian Journal of Analytical Sciences*, 17(2): 340-347.
10. Ngan, T. T. K., Hien, T. T., Tien, L. X., Anh, T. T., Quan, P. M., Cang, M. H. and Toan, T. Q. (2019). Physico-chemical profile of essential oil of kaffir lime (*Citrus hystrix* DC) grown in An Giang Province, Vietnam. *Asian Journal Chemistry*, 31(12): 2855-2858.
11. Hongatanaworkit, T. and Buchbauer, G. (2007). Chemical composition and stimuating effect of *Citrus hystrix* oil on human. *Flavour Fragrance Journal*, 22(4): 443-449.
12. Dewi, C. and Basuki, A. (2018). Identifying citronella plants from UAV imagery using support vector machine. *Telkomnika*, 16(4): 1877–1885.

13. Windha Mega, P. D. and Haryoko (2019). Optimization of parameter support vector machine (SVM) using genetic algorithm to review go-jek's services. *4th International Conference Information Technology, Information System, and Electrical Engineering*, 6: 301-304.
14. Zakiah, M. Y., Zuraida, M., Nurhani, K., Mohd, H. F. R. and Mohd, N. T. (2013). Effect of temperature on kaffir lime oil by using hydro-diffusion steam distillation system. *Malaysian Journal of Analytical Sciences,* 17(2), 326-339.
15. Lubinska-Szczygieł, M., Różańska, A., Namieśnik, J., Dymerski, T., Shafreen, R. B., Weisz, M., ... & Gorinstein, S. (2018). Quality of limes juices based on the aroma and antioxidant properties. *Food Control*, 89: 270-279.
16. An, T. N. T., Ngan, T. T. K., Van, C. K., Anh, H. L. T., Minh, L. V. and Ay, N. V. (2021, March). The major and minor components of Kaffir Lime (*Citrus hystrix* DC) essential oil in the steam distillation process. In *IOP Conference Series: Materials Science and Engineering*, 1092(1): 012082.
17. Jailani, N. S. J., Muhammad, Z., Rahiman, M. H. F. and Taib, M. N. (2021). Analysis of kaffir lime oil chemical compounds by gas chromatography-mass spectrometry (GC-MS) and Z-Score technique. In *2021 IEEE International Conference on Automatic Control & Intelligent Systems:* pp. 110-113.
18. Ismail, N., Rahiman, M. H. F., Taib, M. N., Ali, N. A. M., Jamil, M. and Tajuddin, S. N. (2014). Application of ANN in agarwood oil grade classification. In *2014 IEEE 10th International Colloquium on Signal Processing and its Applications*: pp. 216-220.
19. Sevgi, L. (2007). Synthetic radar-Signal environment: Computer generation of signal, noise, and clutter. *IEEE Antennas and Propagation Magazine*, 49(5): 192-198.
20. Ismail, N. S., Ismail, N., Rahiman, M. H. F., Taib, M. N., Ali, N. A. M. and Tajuddin, S. N. (2018, October). Polynomial tuned kernel parameter in SVM of agarwood oil for quality classification. In *2018 IEEE International Conference on Automatic Control and Intelligent* Systems: pp. 77-82.
21. Ariffin, B. M., Ismail, N., Ali, N. A. M., Rahiman, M. H. F., Tajuddin, S. N. and Taib, M. N. (2018). Multilayer perceptron (MLP) tuned kernel parameter in support vector machine (SVM) for agarwood oil compounds quality classification. *J. Tomography System and Sensors Application*, 1(1): 1-15.
22. Raif, M. A. A., Ismail, N., Ali, N. A. M., Rahiman, M. H. F., Tajuddin, S. N. and Taib, M. N. (2019). Quadratic tuned kernel parameter in Non-linear support vector machine (SVM) for agarwood oil compounds quality classification. *Indonesia Journal Electronic Engineering Computer Science*, 17(3): 1371-1376.
23. Patle, A. and Chouhan, D. S. (2013). SVM kernel functions for classification. In *2013 International Conference on Advances in Technology and Engineering*, pp. 1-9.
24. Apostolidis-Afentoulis, V. and Lioufi, K. I. (2015). SVM classification with linear and RBF kernels: pp. 1-7.
25. Kamarulzaini, K. A. A., Ismail, N., Rahiman, M. H. F., Taib, M. N., Ali, N. A. M. and Tajuddin, S. N. (2018). Evaluation of RBF and MLP in SVM kernel tuned parameters for agarwood oil quality classification. In *2018 IEEE 14th International Colloquium on Signal Processing & Its Applications*, pp. 250-254.