QSAR Study of Larvicidal Phytocompounds as Anti-Aedes Aegypti by using GA-SVM Method

Komang Triolascarya[1], Reza Rendian Septiawan[2], Isman Kurniawan[3]
[1,3]School of Computing, Telkom University
[2]School of Electrical Engineering, Telkom University
[3]Research Center of Human Centric Engineering, Telkom University
[1]triolascarya@student.telkomuniversity.ac.id, [2]zaseptiawan@telkomuniversity.ac.id, [3]ismankrn@telkomuniversity.ac.id

Abstract

Aedes aegypti is one of the most dangerous mosquitoes that can cause several deadly diseases, such as dengue fever, Chikungunya, Zika, and jaundice with high mortality rate. For now, no specific drug has been found that can cure the disease caused by Aedes Aegypti. One possible solution for handling this problem is to inhibit the growth and development of Aedes aegypti larvae. This study aims to implement Genetic Algorithm-Support Vector Machine to develop Quantitative Structure-Activity Relationship model for identification larvicidal phytocompounds as anti-aedes-aegypti. Hyperparameter tuning was performed to improve the performance of the models. Based on the result, we found that the best model was developed by the RBF kernel with the value of R^2 and Q^2_{loo} score are 0.64 and 0.64, respectively.

Keywords: Aedes aegypti, Genetic Algorithm, QSAR, Support Vector Machine

1. Introduction

Mosquitoes are the main vector of several diseases those attack humans and animals that cause thousands of deaths every year. Aedes aegypti is one of the most dangerous mosquitoes that can cause several diseases, such as dengue fever, Chikungunya, Zika, and jaundice [1],[2]. Dengue fever is considered one of the dangerous diseases caused by Aedes aegypti because the mortality rate is high and continues to increase every year [1]. The mortality rate of dengue fever has grown significantly worldwide at this time [1]. It is estimated that nearly thirty-nine million people worldwide are infected annually [1]. Symptoms of dengue fever are characterized by high fever accompanied by severe headache, muscle and joint pain, nausea or vomiting, and swollen glands [3].

Recently, no specific drug has been found that can cure dengue fever [3]. One possible solution for the treatment of dengue fever is to inhibit the growth and development of Aedes aegypti larvae. Several chemical products have been tested against larvae of Aedes aegypti such as phenolic acids, spinosyns, coumarins, et. Al [2]. Those compounds can help to inhibit the growth and development of Aedes aegypti larvae [2],[4],[5]. However, some chemical products are toxic

and harmful to the environment [2]. Therefore, it is expected that the larvicides of plant products will be a source of raw materials and a safer alternative, which results in little waste, and is non-toxic to non-target organisms [6],[7], one of them is a group of larvicidal phytocompound [8]. It is known that the design of conventional drugs is not effective because the new compounds with certain biological activities need to be synthesized to determine their activity [9]. Hence, we need a model that can predict drug candidate activities, such as Quantitative Structure-Activity Relationship (QSAR).

QSAR is an alternative method developed for linking chemical molecules with activity biologically based on their chemical structure [10]. QSAR uses chemometric methods to describe the biological activity or nature of varying physicochemical properties as a function of a molecular descriptor that describes the structure of the chemical molecule [11]. Therefore, computed descriptors can be used to predict new compounds [11]. One of the challenges in QSAR study is to obtain optimal feature. One of the solutions to solve is use meta heuristic algorithm, such as Genetic Algorithm (GA) to select the optimal feature.

Several QSAR studies with and without meta heuristic have been done by several researchers. In 2012, A. Bahesthi performed QSAR modeling to analyze the activity of 68 urea derivatives as antimalarials using the GA-Multiple linear regression method [12]. The model validation was validated using external validation, namely leave-one-out (LOO) cross validation and y-randomization test, and obtaining the squares of the correlation coefficients R^2 0.801 and 0.803, respectively [12]. In 2017, Doucet, et. al. predicts the toxicity of a derivative of piperidine Aedes aegypti using QSAR models [13]. They predict the toxicity of 33 piperidine derivatives against Aedes aegypti [13]. The predicted toxicity was calculated using Ordinary Least Squares-Multi Linear Regression from QSARINS and Support Vector Machine (SVM), with the coefficient of determination (r^2) 0.85 and 0.80, respectively [13].

In 2020, Javidfar performed modeling larvicidal phytocompounds against Aedes aegypti using the index of ideality correlation [2]. They developed three QSAR models to predict pLC5062 plant-derived compounds to fight Aedes aegypti by method-based Monte Carlo on the IIC criteria, with the excellent predictive of the models ($r^2_{Val} = 0.856$ to 0.977) [2]. In 2020, Farisi Rahman, et. Al. carried out a QSAR model derived from Fusidic Acid as an Antimalarial Agent using the Simulated Annealing (SA) – SVM method [14]. The results showed that SA as a feature selection resulted in a satisfactory combination of features. Then, for the best validation results are generated by the RBF kernel [14].

In 2021, Fajar, et. al. predicts the activity of indenopyrazole derivatives as anti-cancer drugs using the QSAR model with the SA-SVM method, with three kernel models for SVM, namely the RBF kernel, linear kernel, and polynomial kernel [15]. Based on the three kernels, the RBF kernel produces an R^2 score train and the best test is 0.79 and 0.60, respectively [15]. Also, QSAR Model has been implemented to identify other disease [16],[17],[18],[19],[20]. However, to the best of our knowledge there is no report of the implementation of meta heuristic, such as GA, to select the features for the case of larvicidal phytocompounds.

In this study, we aim to build QSAR Model to predict larvicidal phytocompound activity as anti-Aedes aegypti with the Genetic Algorithm-Support Vector Machine methods. GA is generally a search-based algorithm built on the concept of natural selection and descendants [21]. GA is a subdivision of a much larger area of computing known as Evolutionary computing [21]. Meanwhile, SVM is a supervised learning technique that determines to classify different categories of data from different disciplines for classification problem solving and regression analysis [22].

2. Research Methods

In this research, we aim to build QSAR Model to predict larvicidal phytocompound activity as anti-Aedes aegypti with the Genetic Algorithm-Support Vector Machine (SVM) methods. Genetic Algorithm is used as a feature selection technique, while the SVM is used as a prediction model. The flowchart design of the research procedures is depicted in Figure 1.
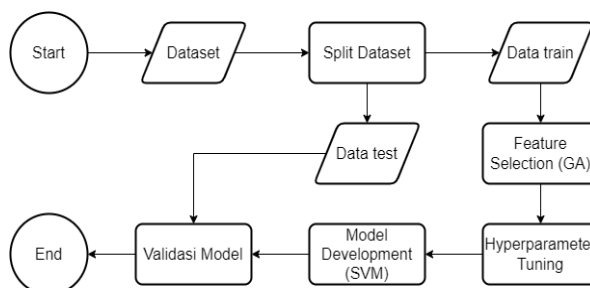

Figure 1. Flowchart of the research procedures

2.1 Datasets

A collection of 62 samples used in this study was obtained from Ref [2]. The observed data is regarding the larvicidal activity against Aedes aegypti is LC_{50} which is converted into molar units and is expressed on a negative logarithmic scale ($-\log LC_{50}$) or called pLC_{50} [2]. The molecule descriptor of the phytocompound was calculated from SMILES structure by using the PaDEL application. Meanwhile, the observed value of pLC_{50} is used as the target value for developing the QSAR model. We performed data reduction on the dataset by calculating the variance value of each feature and the feature with variance value less than 0.5 were removed. Then, the dataset is split into train and test set with the ratio of 70:30.

2.2 Feature Selection

Feature selection is used to reduce dimensions by reducing the number of irrelevant features. The feature selection techniques used in this study is Genetic Algorithm. In computer science, GA is a metaheuristic algorithm inspired by natural processes that belongs to the larger class of evolutionary algorithms [21]. GA are usually used to generate quality solutions for optimization and search problems using selection, crossover, and mutation operators [23]. Flowchart of GA is shown in Figure 2.

GA usually starts by initializing the population and runs in several iterations. At the end of each iteration, a new generation will be obtained and put into the next iteration, the algorithm will end when it reaches the maximum number of iterations or finds the best solution. To evaluate the optimal solution that generated from GA, we performed the fitness function. The equation of fitness function are formulated in Equation (1) – (3)
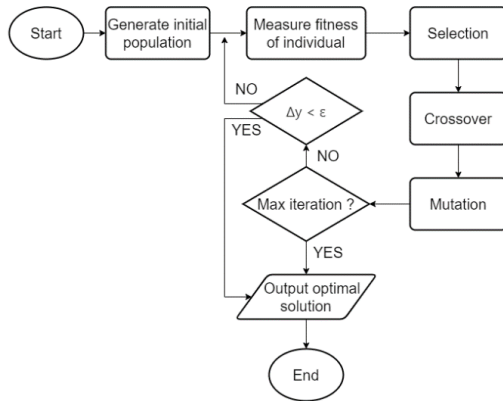
Figure 2. Flowchart of GA.

$$FF = (w_1 \times R^2) + (w_2 \times s_f) \qquad (1)$$

$$R^2 = 1 - \frac{RSS}{TSS} \qquad (2)$$

$$S_f = 1 - \frac{number\ of\ selected\ feature}{total\ feature} \qquad (3)$$

where w_1 represent the coefficient of determination (R^2) of selected feature weight and w_2 is the number of feature weight. Meanwhile, the variable RSS and TSS represent sum of squares of residuals and total sum of squares value, respectively. The parameters used in genetic algorithm are presented in Table 1.

Table 1. Parameter used in GA as feature selection.

Parameter	Values
n_iteration	100
population_size	10
selective_pressure	2
elitism	2
mutation_rate	0.05
weight_score	0.8

2.3 Prediction Model

To predict the model in this study we developed SVM model. Support Vector Regression (SVR) is one of the popular options for predicting and determining curve fitting in both linear and non-linear regression types [22]. This SVR model is the basic elements used in Support Vector Machine (SVM). SVM works by dividing between classes based on hyperplane division in N-dimensional space [22].
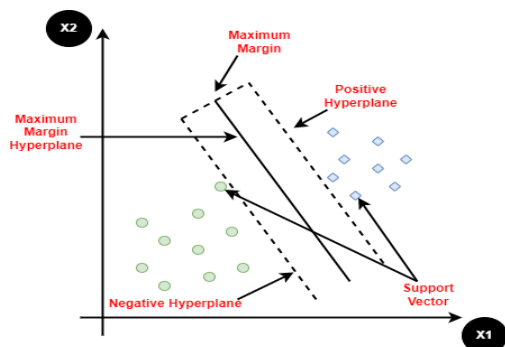

Figure 3. Illustration of hyperplane in SVM.

The essence of SVM is to get the optimal hyperplane location, then measure the margin and find the maximum point of the hyperplane [24]. The generalized equation for hyperplane represented in Equation (4)

$$y = wX + b \qquad (4)$$

where w is weights and b are the intercept at X = 0. In this case we use the SVR model with its respective kernels i.e., linear kernel, polynomial, and RBF. Then, to improve the performance of the model, we perform a hyperparameter tuning procedure. The ranges of parameters values in the hyperparameter tuning of the model selected are presented in Table 2.

Table 2. Ranges of parameters values in hyperparameter tuning step.

Parameter	Ranges
C	$[10^{-3}, 10^{-2}, ..., 10^3]$
gamma	['auto', 'scale', $10^{-4}, 10^{-3}, ..., 10$]
degree	[2, 3, 4, 5, 6]

2.4 Model Validation

To validate the QSAR will be carried out test twice, namely the internal validation and the external validation test which later the value will be compared to the threshold value that determines model acceptance. The internal validation test was carried out by calculating the coefficient of determination of (R^2_{train}) and Leave-One-Out (LOO) cross-validation (Q^2_{loo}) using training data. Meanwhile, in the test external validation is done by calculating the coefficient of determination (R^2_{test}) using test data [25]. Those parameters are formulated in Equation (5) – (16)

$$R^2_{train} = 1 - \frac{\sum(y_{train} - \hat{y}_{train})^2}{\sum(y_{train} - \bar{y}_{train})^2} \qquad (5)$$

$$Q^2_{loo} = 1 - \frac{\sum(y_{train} - \hat{y}_{loo})^2}{\sum(y_{train} - \bar{y}_{train})^2} \qquad (6)$$

$$R^2_{test} = 1 - \frac{\sum(y_{test} - \hat{y}_{test})^2}{\sum(y_{test} - \bar{y}_{train})^2} \qquad (7)$$

$$k = \frac{\sum(y \times \hat{y})}{\sum(\hat{y})^2} \qquad (8)$$

$$k' = \frac{\sum(y \times \hat{y})}{\sum(y)^2} \qquad (9)$$

$$r^2 = \frac{[\sum(y - \bar{y})(\hat{y} - \bar{\hat{y}})]^2}{\sum(y - \bar{y})^2 \times \sum(\hat{y} - \bar{\hat{y}})^2} \qquad (10)$$

$$r_0^2 = 1 - \frac{\sum(y - k \times \hat{y})^2}{\sum(y - \bar{y})^2} \qquad (11)$$

$$r_0'^2 = 1 - \frac{\sum(\hat{y} - k' \times \hat{y})^2}{\sum(\hat{y} - \bar{\hat{y}})^2} \qquad (12)$$

$$r_m^2 = r^2 \times \left(1 - \sqrt{r^2 - r_0^2}\right) \qquad (13)$$

$$r_m'^2 = r^2 \times \left(1 - \sqrt{r^2 - r_0'^2}\right) \qquad (14)$$

$$\overline{r_m^2} = \frac{(r_m^2 - r_m'^2)}{2} \qquad (15)$$

$$\Delta r_m^2 = |r_m^2 - r_m'^2| \qquad (16)$$

the variable y and $\hat{y}$ represent the actual value and the predicted value of the pLC_{50} value, respectively. While $\bar{y}$ and $\bar{\hat{y}}$ represents the average actual value and predicted value, respectively. Models that can be considered as a valid model if they meet the criteria shown in Table 3.

Table 3. Threshold of parameter for Model Validation

Parameter	Threshold [25]
R^2	> 0.6
Q_{loo}^2	> 0.5
k'	$0.85 \leq k' \leq 1.15$
$\dfrac{(r^2 - r_0^2)}{r^2}$	< 0.1
$|r_0^2 - r_0'^2|$	< 0.3
$\overline{r_m^2}$	> 0.5
Δr_m^2	< 0.2

The Applicability Domain (AD) of the model is determined to ensure the data set lies in the model domain. Determination of AD is calculated using the leverage method which is formulated in Equation (17)

$$H = X(X^TX)^{-1}X^T \qquad (17)$$

where X represents the score matrix obtained from the PLSR procedure and the critical leverage value (h*), and X^T represent transpose of X. The equation for critical leverage (h*) represented in Equation (18)

$$h^* = 3p/n \qquad (18)$$

where p defined the number of attributes and n is the data involved in the training process. The predicted value of the data can be accepted if the calculated leverage value is less than the critical leverage [25].

3. Results and Discussions

3.1 Feature Selection

Genetic algorithms are mostly criteria-based probability in nature. On the contrary, the algorithm works well against local random search, which uses random solutions, cannot identify the best solution. Therefore, feature selection using the genetic algorithm is carried out with multiple runs of 20 times to ensure that the objective results of the scores obtained are consistent and this can be confirmed by looking at the results of the standard deviation of 20 multiple runs of each model used. The distribution of the objective scores for each kernel is shown in Figure 4.

Based on Figure 4, as for linear kernel, the distribution of the highest score objective value most often appears in the range 0.81 to 0.82 and the lowest score objective appears in the range 0.79 to 0.80. As for the polynomial kernel, the highest score objective value distribution most often appears in the range 0.76 to 0.80 and the lowest objective score appears in the range 0.72 to 0.74. As for RBF kernel, the distribution of the highest score objective value most often appears in the range 0.79 to

0.80 and the lowest score objective occurs in the range 0.77 to 0.79.
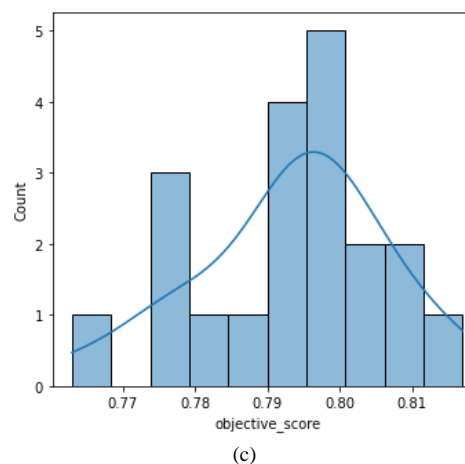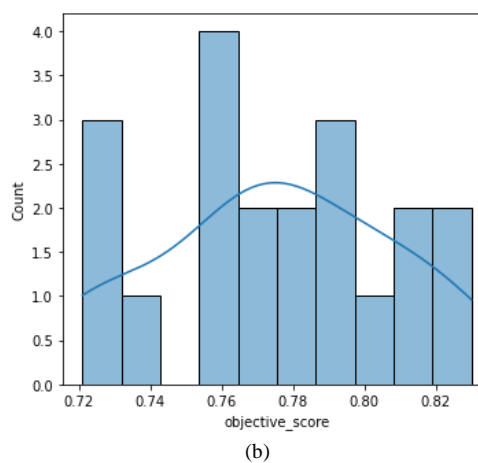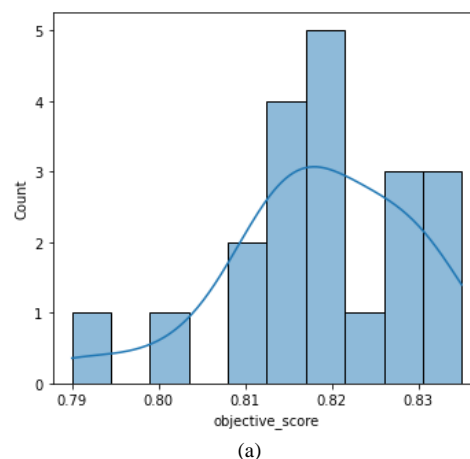
(a)

(b)

(c)

Figure 4. The distribution chart of objective score with (a) linear kernel; (b) polynomial kernel; (c) RBF kernel.

Table 4. Total feature and objective score selected on each kernel.

Kernel	Total Feature	Objective Score	Avg ± St.dev
Linear	234	0.832	±0.010
Polynomial	212	0.830	±0.032
RBF	207	0.817	±0.013

Summary of selected feature is shown in Table 4. Based on the result, we found that the optimal number of features for linear, polynomial, and RBF kernel are 234, 212, and 207, respectively. We also found that the linear and RBF kernel produced the lowest standard deviation compared to the other model. This indicates that the GA solution in the linear and RBF kernel performed almost similarly in every multiple-run scheme.

The convergence plot of GA shows in Figure 5, the highest objective score is obtained by the linear kernel, and in linear kernel gets the best score objective with the fastest iteration. While the lowest score objective is found in the RBF kernel, in the RBF kernel it is faster to get the best score objective (based on iterations) than the polynomial kernel.
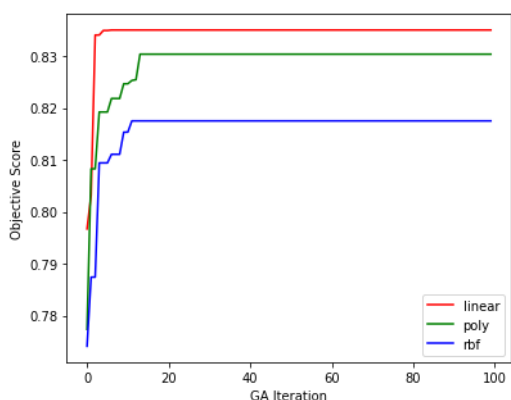

Figure 5. Convergence plot of GA for each kernel

3.2 Hyperparameter Tuning

The summary of hyperparameter tuning result is presented in Table 5, hyperparameter tuning is used to obtain the best parameters for all Kernel in SVR model. For parameters 'C' and 'gamma' we find that each kernel has a different value, while the 'degree' parameter in each kernel has the same value.

Table 5. Summary of hyperparameter tuning result

Parameter	Kernel		
	Linear	Polynomial	RBF
C	0.1 (1)	1 (1)	10 (1)
gamma	-	'scale' ('scale')	0.01('scale')
degree	-	2 (3)	-

*The value inside the parenthesis is the default value of the svm parameter

We present a comparison of the R^2 score between non-tuned and tuned kernel in Figure 5. The result shows that a significant difference between the tuned and non-tuned kernel. We found the improvement increase of R^2 score of linear, polynomial, RBF kernel are 0.331, 0.389, and 0.309, respectively. The highest increase occurs in the polynomial kernel because, the polynomial kernel uses 3 parameters compared to other kernels. The comparison of hyperparameter tuning result presented in Figure 6.
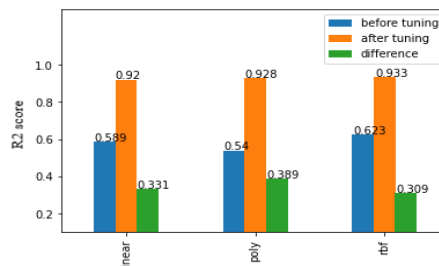

Figure 6. Comparison of hyperparameter tuning result for each kernel

3.3 Model Validation

The comparison of the predicted value and the actual value of pLC50 for each kernel is shown in Figure 7. The x-axis and y-axis represent the actual value and the predicted value, respectively. Each model shows a strong relationship between the predicted model and the true value. We also find that each data point for all kernels is located close to the diagonal line with a not very significant difference.
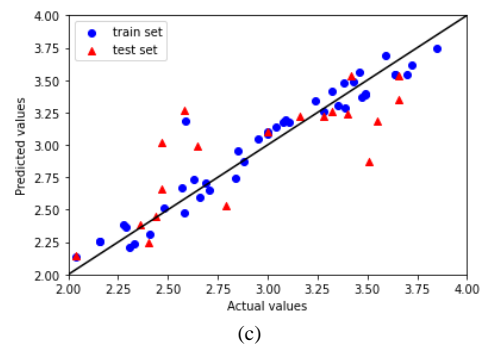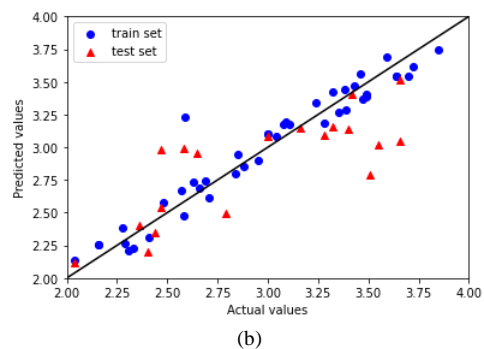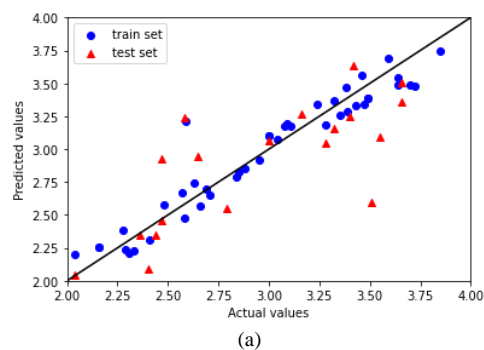

(a)

(b)

(c)

Figure 7. Scatter plot of predicted pLC_{50} vs. Actual pLC_{50} for each Kernel; (a) Linear; (b)Polynomial; (c) RBF

The summary of validation result is present in Table 6 and Table 7, to validate the QSAR model, several statistical parameters were calculated and compared with the threshold value [25]. In the training set, we found for each model met the criteria on the threshold value. However, in the testing set, the linear model is the only invalid model, because there are values that do not meet the threshold criteria. Then, the best R^2 score is obtained by the RBF kernel which is caused by the low feature number of the RBF kernel is 207, and the worst R^2 score is obtained by the linear kernel because the feature number of the linear kernel is 234.

Table 6. Summary of validation result for each kernel in training set

Parameter	Kernel			Threshold
	Linear	Polynomial	RBF	
R^2	0.92	0.92	0.93	> 0.6
Q^2_{loo}	0.61	0.61	0.64	> 0.5
k'	0.98	0.98	1.03	$0.85 \le k' \le 1.15$
$\frac{(r^2 - r_0^2)}{r^2}$	0.02	0.02	0.00	< 0.1
$|r_0^2 - r_0'^2|$	0.01	0.01	0.00	< 0.3
$\overline{r_m^2}$	0.78	0.78	0.87	> 0.5
Δr_m^2	0.07	0.07	0.05	< 0.2

Table 7. Summary of validation result for each kernel in testing set

Parameter	Kernel			Threshold
	Linear	Polynomial	RBF	
R^2	0.58	0.61	0.64	> 0.6
Q^2_{loo}	-	-	-	> 0.5
k'	0.88	0.88	0.97	$0.85 \le k' \le 1.15$
$\frac{(r^2 - r_0^2)}{r^2}$	0.06	0.03	0.02	< 0.1
$|r_0^2 - r_0'^2|$	0.00	0.03	0.04	< 0.3
$\overline{r_m^2}$	0.46	0.50	0.56	> 0.5
Δr_m^2	0.00	0.05	0.07	< 0.2

The William's plot that represents the applicability domain (AD) of the model is shown in Figure 7, based on Figure 7, for each data set and data train on each kernel there is no value higher than the critical leverage (h*), meaning that the data is reliable. The AD William's plot shows that our model results are proportional to the existing leverage approach as well as predictive models were mostly acceptable for all responses. The results of the analysis in Figure 8 show that the predictions of each model are most likely correct.
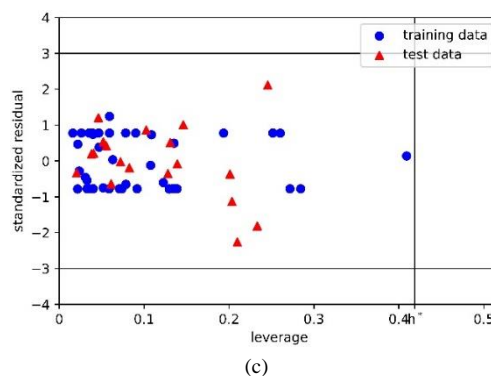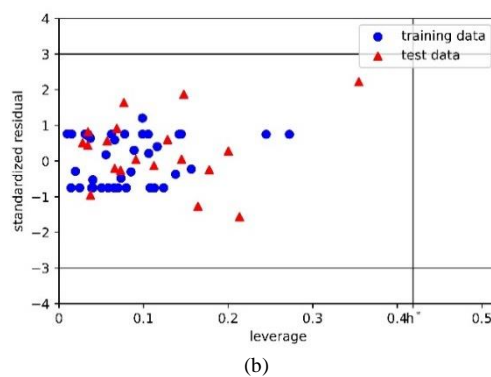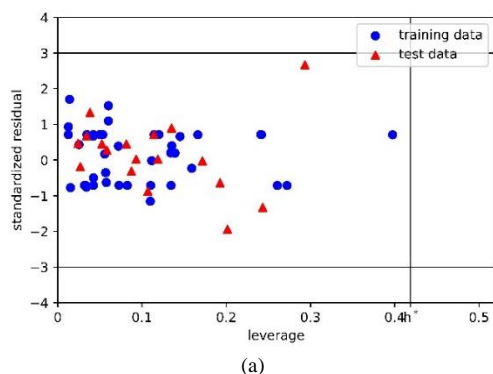
(a)

(b)

(c)

Figure 8. Williams plots of applicability domain for (a) linear kernel; (b) polynomial kernel; (c) RBF kernel.

4. Conclusion

We have developed a QSAR model by using GA-SVM method, SVM method to identify larvicidal compounds as anti-Aedes aegypti. The number of features is reduced by the variance threshold. Then, feature selection is continued by calculating the statistical parameters of the genetic algorithm. We performed model performance improvement with hyperparameter tuning procedure. Based on the validation results, we found that the best model was developed by the RBF kernel that satisfies all criteria with the value of R^2 and Q^2_{loo} score is 0.64 and 0.64, respectively.

References

[1] K. V. Lakshmi, A. V. Sudhikumar, and E. M. Aneesh, "Larvicidal activity of phytoextracts against dengue fever vector, Aedes aegypti - A review," *Plant Sci. Today*, vol.5, no. 4, pp. 167–174, 2018, doi: 10.14719/pst.2018.5.4.407.

[2] M. Javidfar and S. Ahmadi, "QSAR modelling of larvicidal phytocompounds against Aedes aegypti using index of ideality of correlation," *SAR QSAR Environ. Res.*, vol. 31, no. 10, pp. 717–739, 2020, doi: 10.1080/1062936X.2020.1806922.

[3] J. E Cogan, "Dengue and severe dengue," *Who*, 2018, [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue.

[4] S. Chalom *et al.*, "Utilization of electrocoagulation for the isolation of alkaloids from the aerial parts of Stemona aphylla and their mosquitocidal activities against Aedes aegypti," *Ecotoxicol. Environ. Saf.*, vol. 182, no. July, p. 109448, 2019, doi: 10.1016/j.ecoenv.2019.109448.

[5] P. Yogarajalakshmi *et al.*, "Toxicological screening of marine red algae Champia parvula (C. Agardh) against the dengue

mosquito vector Aedes aegypti (Linn.) and its non-toxicity against three beneficial aquatic predators," *Aquat. Toxicol.*, vol. 222, no. March, p. 105474, 2020, doi: 10.1016/j.aquatox.2020.105474.

[6] R. Pavela, "Essential oils for the development of eco-friendly mosquito larvicides: A review," *Ind. Crops Prod.*, vol. 76, pp. 174–187, 2015, doi: 10.1016/j.indcrop.2015.06.050.

[7] R. Pavela, "History, presence and perspective of using plant extracts as commercial botanical insecticides and farm products for protection against insects - A review," *Plant Prot. Sci.*, vol. 52, no. 4, pp. 229–241, 2016, doi: 10.17221/31/2016-PPS.

[8] M. R. S. Silvério, L. S. Espindola, N. P. Lopes, and P. C. Vieira, "Plant natural products for the control of Aedes aegypti: The main vector of important arboviruses," *Molecules*, vol. 25, no. 15, 2020, doi: 10.3390/molecules25153484.

[9] S. Vilar, G. Cozza, and S. Moro, "Medicinal Chemistry and the Molecular Operating Environment (MOE): Application of QSAR and Molecular Docking to Drug Discovery," *Curr. Top. Med. Chem.*, vol. 8, no. 18, pp. 1555–1572, 2008, doi: 10.2174/156802608786786624.

[10] M. R. Keyvanpour and M. B. Shirzad, "An Analysis of QSAR Research Based on Machine Learning Concepts," *Curr. Drug Discov. Technol.*, vol. 18, no. 1, pp. 17–30, 2021.

[11] C. B. R. Santos *et al.*, "A SAR and QSAR study of new artemisinin compounds with antimalarial activity," *Molecules*, vol. 19, no. 1, pp. 367–399, 2014, doi: 10.3390/molecules19010367.

[12] A. Beheshti, E. Pourbasheer, M. Nekoei, and S. Vahdani, "QSAR modeling of antimalarial activity of urea derivatives using genetic algorithm-multiple linear regressions," *J. Saudi Chem. Soc.*, vol. 20, no. 3, pp. 282–290, 2016, doi: 10.1016/j.jscs.2012.07.019.

[13] J. P. Doucet, E. Papa, A. Doucet-Panaye, and J. Devillers, "QSAR models for predicting the toxicity of piperidine derivatives against Aedes aegypti," *SAR QSAR Environ. Res.*, vol. 28, no. 6, pp. 451–470, 2017, doi: 10.1080/1062936X.2017.1328855.

[14] M. Fajar Rizqi, R. Rendian Septiawan, and I. Kurniawan, "Implementation of Simulated Annealing-Support Vector Machine on QSAR Study of Indenopyrazole Derivative as Anti-Cancer Agent," *2021 9th Int. Conf. Inf. Commun. Technol. ICoICT 2021*, no. Icidm, pp. 662–668, 2021, doi: 10.1109/ICoICT52021.2021.9527416.

[15] M. Fajar Rizqi, R. Rendian Septiawan, and I. Kurniawan, "Implementation of Simulated Annealing-Support Vector Machine on QSAR Study of Indenopyrazole Derivative as Anti-Cancer Agent," *2021 9th Int. Conf. Inf. Commun. Technol. ICoICT 2021*, pp. 662–668, 2021, doi: 10.1109/ICoICT52021.2021.9527416.

[16] H. F. Azmi, K. M. Lhaksmana, and I. Kurniawan, "QSAR Study of Fusidic Acid Derivative as Anti-Malaria Agents by using Artificial Neural Network-Genetic Algorithm," *2020 8th Int. Conf. Inf. Commun. Technol.*, 2020.

[17] F. Rahman, K. M. Lhaksmana, and I. Kurniawan, "Implementation of Simulated Annealing-Support Vector Machine on QSAR Study of Fusidic Acid Derivatives as Anti-Malarial Agent," *2020 6th Int. Conf. Interact. Digit. Media*, 2020.

[18] A. Arwansyah, A. R. Arif, G. Syahputra, S. Sukarti, and I. Kurniawan, "Theoretical studies of Thiazolyl-Pyrazoline derivatives as promising drugs against malaria by QSAR modelling combined with molecular docking and molecular dynamics simulation," *SAR QSAR Environ. Res.*, vol. 47, no. 12, pp. 988–1001, 2021.

[19] I. Kurniawan, R. Wardhani, M. Rosalinda, and N. Ikhsan, "QSAR Study for Prediction of HIV-1 Protease Inhibitor Using the Gravitational Search Algorithm–Neural Network (GSA-NN) Methods," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 12, no. 2, pp. 62–77, 2021.

[20] A. Arwansyah *et al.*, "Molecular modelling on SARS-CoV-2 papain-like protease: an integrated study with homology modelling, molecular docking, and molecular dynamics simulations," *SAR QSAR Environ. Res.*, vol. 32, no. 9, pp. 699–718, 2021.

[21] A. Lambora, K. Gupta, and K. Chopra, "Genetic Algorithm- A Literature Review," *Proc. Int. Conf. Mach. Learn. Big Data, Cloud Parallel Comput. Trends, Prespectives Prospect. Com. 2019*, no. 1998, pp. 380–384, 2019, doi: 10.1109/COMITCon.2019.8862255.

[22] D. Parbat and M. Chakraborty, "A python based support vector regression model for prediction of COVID19 cases in India," *Chaos, Solitons and Fractals*, vol. 138, pp. 3–7, 2020, doi: 10.1016/j.chaos.2020.109942.

[23] C. Audet and W. Hare, "Genetic Algorithms," *Springer Ser. Oper. Res. Financ. Eng.*, pp. 57–73, 2017, doi: 10.1007/978-3-319-68913-5_4.

[24] I. Ahmad, M. Basheri, M. J. Iqbal, and A. Rahim, "Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection," *IEEE Access*, vol. 6, no. c, pp. 33789–33795, 2018, doi: 10.1109/ACCESS.2018.2841987.

[25] I. Kurniawan, M. S. Fareza, and P. Iswanto, "Comfa, molecular docking and molecular dynamics studies on cycloguanil analogues as potent antimalarial agents," *Indones. J. Chem.*, vol. 21, no. 1, pp. 66–76, 2021, doi: 10.22146/ijc.52388.