Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 6 No. 5 (2022) 735 - 741 ISSN Media Electronic: 2580-0760

QSAR Study on Diacylgycerol Acyltransferase-1 (DGAT-1) Inhibitor as Anti-diabetic using PSO-SVM Methods

I Kadek Andrean Pramana Putra[1], Reza Rendian Septiawan [2], Isman Kurniawan[3]
[1]School of Computing, Telkom University
[2]School of Electrical Engineering, Telkom University
[1,3]Research Center of Human Centric Engineering, Telkom University
[1]andrepramana@student.telkomuniversity.ac.id, [2]zaseptiawan@telkomuniversity.ac.id,[3]ismankrn@telkomuniversity.ac.id

Abstract

Diabetes mellitus is a chronic disease that can occurred to anyone. Up until now, there are no specific drugs that have been found which can completely cure diabetes. One of the possible steps to treat diabetes mellitus is by inhibiting the growth of Diacylglycerol Acyltransferase-1 (DGAT-1) enzyme. This study aims to build a QSAR model on DGAT-1 inhibitors as anti-diabetic using Particle Swarm Optimization (PSO) and Support Vector Machine (SVM). Acyl-CoA: DGAT1 is a mikrosomal enzyme in lipogenesis which is increased in metabolically active cells to meet nutrient requirements. Microsomal enzymes that have an important in the triglyceride synthesis process of 1,2-diacylglycerol by catalyzing-acyl coa dependent acylations as anti-diabetics. The dataset used in this study consists of 228 samples containing molecular structures and their inhibitor activities. We reduce the number of features by removing features with a standard deviation less than the threshold value, followed by the PSO algorithm. The best-predicted result is obtained through the implementation of SVM with RBF kernel, with the score of R^2_{train} and R^2_{test} are 0.75 and 0.67, respectively.

Keywords: Diabetes, Diacylglycerol Acyltransferase-1, Particle Swarm Optimization, QSAR, Support Vector Machine

1. Introduction

Diabetes mellitus is a chronic disease caused by the lack of insulin hormone production in the body and the inability to use insulin properly [1]. Insulin is a hormone produced by the pancreas to digest glucose in blood [2]. The disease is usually characterized by high blood sugar level. Diabetes is divided into three types; i.e. type 1, type 2, and gestational diabetes [1],[2]. Diabetes mellitus type 1 is caused by an increase in blood sugar levels. This type of diabetes is caused by damages in the pancreas so that there is no insulin production at all in the body. Diabetes mellitus type 2 is caused by an increase in blood sugar level due to a decrease in insulin secretion by the pancreas gland. Gestational diabetes is caused by an increase in blood sugar level during the pregnancy period [1], [2].

In 2020, The Indonesian Ministry of Health reported that there are 463 million people worldwide in the age range of 20-79 years old suffered from diabetes in 2019. The prevalence rate is predicted to increase to 578 million people in 2030 and 700 million people in 2045 [2]. Indonesia is in the 7th position out of 10 countries in the world with the most diabetes, with a total of 10.7 million people. This number is predicted to increase to 13.7 million in 2030 and 16.6 million people in 2045 [3].

Up to now, there is no drugs that has been found which can cure diabetes completely [1]. One of the possible ways to treat diabetes is by inhibiting the growth of Diacylglycerol Acyltransferase-1 (DGAT-1) enzyme. Several diabetes drugs, such as Metformin, Glipalamide, and Glimepiride, have been developed to treat this disease, but those drugs can cause quite serious side effects, such as nausea, vomiting, hypoglycemia, dizziness, and tremors [1]. The conventional drug models are ineffective because new compounds with certain biological activities must be synthesized to determine their activity [4].

Therefore, in silico method, such as QSAR model, can be used to accelerate drug discovery. The application of the QSAR method can produce a model that can show the relationship between the individual compounds and their biological activities. QSAR method works by applying mathematical calculations based on the biological activity of a compound and also considering

Accepted: 14-07-2022 | Received in revised: 13-09-2022 | Published: 01-10-2022

the structural and chemical characteristics of the compound [5].

Several studies of QSAR method with and without metaheuristic algorithm on drug development have been done. In 2009, Saqib and Mohammad conducted a QSAR 3D study on triazolopiperazine-amides inhibitors of dipeptidyl peptidase-IV as an anti-diabetic and obtained results of r^2 and r^2_{pred} were 0.816 and 0.863 [6]. In 2019 Al-Fakih and Algamal conducted research on QSAR for the prediction of anti-diabetic activity using the binary gravitational search algorithm method and obtained TVBGSA results Q^2_{int} of 0.957 and Q^2_{LGO} of 0.951 [7].

In 2019, Faghihi et al. conducted molecular docking and QSAR study of quinazoline, 2-benzoxazolione, and Diazocoumarin derivatives as Anti-HIV-1 Agents, the QSAR model produce a statistically significant correlation coefficient R^2 and Q^2_{loo} are 0.84 and 0.73, respectively. This model is validated by a molecular test set which gives a satisfactory predictive value R^2_{test} is 0.79 [8]. In 2019 Kumar and Shindu conducted a QSAR study for DGAT-1 inhibitors based on the SMILES descriptor using the Monte-Carlo method and obtained statistical results r^2 of 0.8129 and Q^2 of 0.7962 [4].

In 2021 Kurniawan et al. conducting CoMFA research, molecular docking and molecular dynamics study on cycloguanil analogues as potent antimalarial agents, which obtain R^2 of training data and test data of 0.85 and 0.70, respectively, and Q^2 of Q^2_{loo} score of 0.77 for both of them [9]. One of the challenges in QSAR study is to select the optimal features. Metaheuristic algorithm can be used to solve this issue. However, to the best of our knowledge there is no report of implementation of metaheuristic algorithm to select the optimal features for the case of QSAR study on DGAT-1 inhibitor. PSO algorithm is one of the metaheuristics that has the advantage of the speed in solving an optimization problem.

This study aims to build a model that can predict DGAT-1 inhibitors as anti-diabetics using Particle Swarm Optimization (PSO)-Support Vector Machine (SVM) method. The several study on health sector, has been done previously in 2021, Dharmawan conducted a heart disease prediction using PSO and SVM methods which obtained results of accuracy and AUC were 0.848 and 0.898 [10]. The PSO algorithm is used to perform feature selection by removing irrelevant features in each model descriptor to produce a dominant descriptor in each model. In this study, the SVR algorithm is used which is a derivative of the SVM algorithm. The SVR algorithm was used to produce the best predictive value of pIC_{50} from DGAT-1 inhibitors. The predictive value can be used to provide recommendations for compounds that have the potential to be used in the drug discovery of anti-diabetic [11].

2. Research Methods

2.1 Dataset

The anti-diabetic compounds used in this study are of DGAT-1 inhibitors with 228 samples of compound derivatives retrieved from a ChEMBL database (https://www.ebi.ac.uk/chembl) with the target ChEMBL ID CHEMBL1075284. The molecular descriptors of each compound were calculated using PaDEL application which resulted in 1444 molecular descriptors. The experimental IC_{50} value of the compound in nano-molar (nM) units was converted to molar units (M). To produce the target value in a smaller range, we convert IC_{50} to pIC_{50}using the formula $pIC_{50} = -logIC_{50}$. Then, the pIC_{50}value was used as the target value for the development of the QSAR model [4].

2.2 Preprocessing

The variation of molecular descriptor data is reduced by using the standard deviation below 0.5 to obtain 508 descriptors followed by scaling the dataset using MinMax method. Then, the dataset is split into training data and test data with a ratio of 70:30 randomly.

2.3 Feature Selection

Feature selection is used to remove irrelevant features so that too complex models can be avoided. The feature selection technique used in this study is the Particle Swarm Optimization (PSO) algorithm. PSO is a stochastic population-based optimization technique inspired by bird groups and fish flocks, originally designed and introduced by Kennedy and Eberhart (1995) [12].

The flow of the PSO algorithm starts with a population of particles which position is a potential solution to the problem being studied and the velocity that is randomly initiated in the search space. In each iteration, we find the best position by updating the velocity and position of the particle. The illustration of the scheme from the PSO is shown in Figure 1.

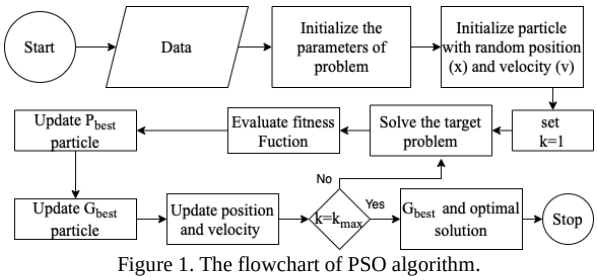


Figure 1. The flowchart of PSO algorithm.

Based on Figure 1, for each iteration of the PSO algorithm, the particle will be initialized with a random position(x) and velocity(v). The fitness value for each

DOI: https://doi.org/10.29207/resti.v6i5.4294

particle is calculated using the fitness function. The fitness function produced the two best position; best personal position (P_{best}) and best global position (G_{best}). P_{best} is the position where the particle is best visited. While G_{best} is the best position of the swarm has been visited since the first step. P_{best} and G_{best} are used to update the velocity of the particles [13]. PSO algorithm will end if the iteration value reaches the maximum value. The PSO equations of fitness function are shown in (1)-(3):

$$V_k(k+1) = w_k V_k(k) + c_1 r_1 \left(P_{best}(k)\right) + c_2 r_2 \left(G_{best}(k)\right) \quad (1)$$

$$k = 1,2,\dots,N \quad (2)$$

$$X_k(k+1) = X_k(k) + V_k(k+1) \quad (3)$$

where w represent a coefficient to control particle momentum by weighting the contribution from the previous velocity. The variable c_1 defined as a first acceleration constant that the particle has towards itself. c_2 is a second acceleration constant that a particle has towards its neighbors. The value of r_1 and r_2 set to random values is intended to give a stochastic process [13]. The parameters used in PSO algorithm is shown as Table 1.

Table 1. Parameters used in PSO algorithm.

Parameter	Value
Maximum Iteration (k)	200
Population	20
Velocity (w)	0.9
First acceleration constant (C_1)	2
Second acceleration constant (C_2)	2
r_1	[0,1]
r_2	[0,1]

2.4. Support Vector Machine

SVM algorithm is a supervised machine learning method which is widely used for classification and regression analysis [14]. Support Vector Regression (SVR) algorithm is an adaptation of SVM algorithm in its application used to solve classification problems that produce round or discrete values, while the SVR algorithm in its application is used to solve regression problems that produce real or continuous values [15].

SVM algorithm works by dividing between classes of hyperplanes in N-dimensional space [15]. The purpose of SVM is to get the location point of the hyperplane. Margin is the value of the distance between the hyperplane and the nearest pattern point in each class [13]. In general, data regression by SVM is done by creating a hyperplane, then the data used must be linear. Therefore, in non-linear data such as DGAT-1 inhibitors, it is necessary to include kernel functions to solve the problem. An illustration of the hyperplane is shown in Figure 2.

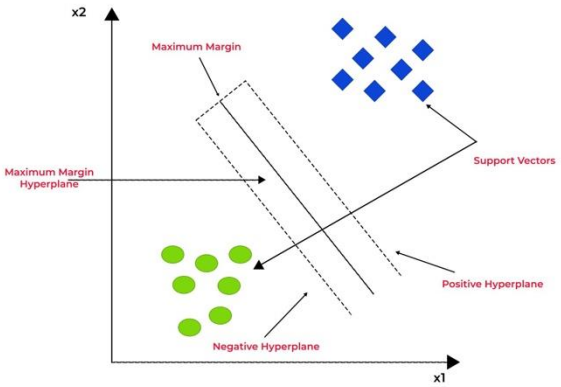


Figure 2. Hyperplane illustration of SVM algorithm

In Figure 2 the data is mapped to a vector space which has larger dimensions. In this vector space, a hyperplane is constructed to separate the two classes. The process of finding hyperplanes on SVM is as follows, there is a data set $\vec{x}_i \in (\vec{x}_1, \vec{x}_2, .. \overrightarrow{x_n})$ with $\vec{x}_i$ data consist of n attributes and two classes $y_i \in [-1, +1]$. The assumption of linear data and intermediate classes $+1$ and -1, could be separated by a hyperplane and defined on an equation (4) for class +1 and equation (5) for class -1 [16].

$$\vec{w} \cdot \vec{x} + b \geq 1 \quad (4)$$

$$\vec{w} \cdot \vec{x} + b \leq -1 \quad (5)$$

To produce an optimal hyperplane, a separating hyperplane is needed that can maximize the distance between the hyperplane and its closest point, which is $1/\|\vec{w}\|$. This can be defined as a Quadratic Programming problem, which is to find the minimum point of the equation. The formula to find the minimum point is written as equation (6) and (7) [10] [16].

$$y_i(\vec{x}_i \cdot \vec{w} + b) - 1 \geq 0 \quad (6)$$

$$\min \tau(\vec{w}) = \frac{1}{2}||\vec{w}||^2 + C\sum_i^n \varepsilon_i \quad (7)$$

where $\vec{w}$ is defined as weight, $\vec{x}$ is train data, b is relative field position, $\vec{x}_i$ is iteration of train data, and y_i is an iteration of the target class. The parameters selection is very important, because the selected parameter values can overfit the data. Thus, to resolve the problem of nonlinear this study uses 3 kernel functions with linear, polynomial, and RBF. Each kernel has unique parameters. The equation for each kernel function are shown in Table 2 [17], [18].

Table 2. Kernel function used in SVM

Kernel	Equation
Linear	$K(\vec{x}_i, \vec{x}_j) = (\vec{x}_i, \vec{x}_j)^T$
Polynomial	$K(\vec{x}_i, \vec{x}_j) = (\vec{x}_i^T + 1)^p$
RBF	$K(\vec{x}_i, \vec{x}_j) = \exp\left(-\frac{1}{2\sigma^2}||\vec{x}_i, \vec{x}_j - \vec{x}_i^T||^2\right)$

DOI: https://doi.org/10.29207/resti.v6i5.4294

2.5. Hyperparameter Tuning

The hyperparameter tuning is used to improve model performance. The tuning is carried out by applying all possible combinations of parameters and comparing parameter scores to produce the best combination of parameters. The hyperparameter tuning procedure is perfomed using the Grid Search Cross-Validation methods. The parameters of the SVM involved in this scheme and their range is shown in Table 3.

Table 3. The variation of parameters in SVM methods

Parameter	Value
C	$[10^{-3}, 10^{-2}, \dots, 10^{3}]$
Gamma	['Auto','Scale']
Degree	[2,3,4,5,6]

2.6 Model Validation

To validate the prediction models, internal and external validation methods are used by calculating certain parameters. In the internal validation process, the correlation coefficient used are R^2_{train} cross-validation and Q^2_{loo} leave-one-out by using training data. Meanwhile, in external validation correlation coefficient for test data (R^2_{test}) is used.

The model valid if the corresponding parameters values are within the respective thresholds. In addition, several validation methods were also carried out to confirm that the model was acceptable. The parameters used is in validation processed can be written as follows the equations (8)-(15) [19],[20] :

$$R^2_{train} = 1 - \frac{\sum(y_{train} - \hat{y}_{train})^2}{\sum(y_{train} - \bar{y}_{train})^2} \quad (8)$$

$$Q^2_{loo} = 1 - \frac{\sum(y_{train} - \hat{y}_{loo})^2}{\sum(y_{train} - \bar{y}_{train})^2} \quad (9)$$

$$R^2_{test} = 1 - \frac{\sum(y_{test} - \hat{y}_{test})^2}{\sum(y_{test} - \bar{y}_{train})^2} \quad (10)$$

$$k = \frac{\sum(y \times \hat{y})}{\sum(\hat{y})^2} \quad (11)$$

$$k' = \frac{\sum(y \times \hat{y})}{\sum(y)^2} \quad (12)$$

$$r^2 = \frac{[\sum(y - \bar{y})(\hat{y} - \bar{\hat{y}})]^2}{\sum(y - \bar{y})^2 \times \sum(\hat{y} - \bar{\hat{y}})^2} \quad (13)$$

$$r_0^2 = 1 - \frac{\sum(y - k \times \hat{y})^2}{\sum(y - \bar{y})^2} \quad (14)$$

$$r_0'^2 = 1 - \frac{\sum(\hat{y} - k' \times \hat{y})^2}{\sum(\hat{y} - \bar{\hat{y}})^2} \quad (15)$$

where y and $\hat{y}$ are representation of the experiment result and the predicted value of pIC_{50}, respectively, while $\bar{y}$, and $\bar{\hat{y}}$ defined as the average of experimental and predicted values, respectively. The acceptance of the model was considering according to the following criteria [19],[20].

$R^2 > 0.6$

$Q^2 > 0.5$

$0.85 \leq k \leq 1.15$ or $0.85 \leq k' \leq 1.15$

$\frac{(r^2 - r_0^2)}{r^2} < 0.1$ or $\frac{(r^2 - r_0'^2)}{r^2} < 0.1$

each model is validated by using Applicability Domain (AD), to ensure data points remain within the model domain by using the leverage method. The formula for AD validation is written as equation (16) [19],[20].

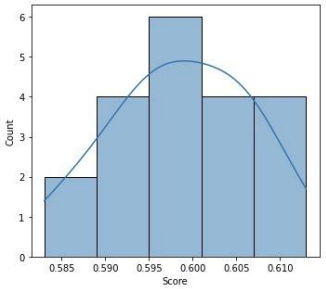
$$H = X(X^T X)^{-1} X^T \quad (16)$$

where X is defined as the values of a matrix obtained from the PLSR procedure, the critical leverage (h*) is represent as 3p/n, p and n, respectively, determined by the number of attributes and dataset involved in the training process. The model was acceptable if the calculated leverage of predicted value was smaller than critical leverage Then the predicted value on AD model is shown using William's plot [9].

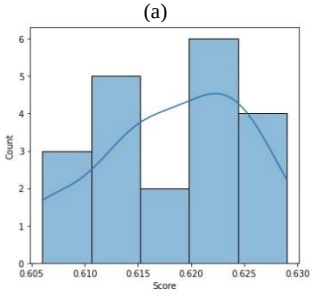
3. Results and Discussions

3.1 Feature Selection

The feature selection uses the PSO algorithm to produce an optimal descriptor that is selected based on the best fitness value. The fitness value is based on the regression of the SVR cross-validation model which is applied to individu in a population, where individu is a representation of several combinations of descriptor compounds. However, the implementation of feature selection on the PSO algorithm produced different objective score and the number of descriptors for each iteration. To acquire correct values, 20 iterations of the PSO algorithm were carried out, and comparisons were made based on the objective score for each iteration. The plot of the distribution of each kernel is shown in Figure 3.

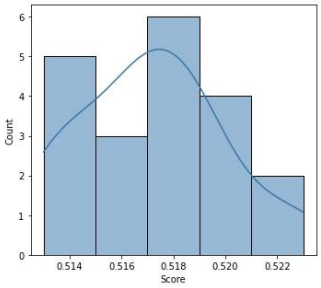


(a)



(b)

DOI: https://doi.org/10.29207/resti.v6i5.4294



(c)

Figure 3. The distribution of R^2_{score} to the number of iterations in linear (a), polynomial (b), and RBF kernels (c)

From Figure 3 we found that each kernel has a shifted distribution in different ranges. In particular, fitness values of the linear kernel are distributed between 0.585 to 0.610, while polynomial kernel scores are distributed between 0.605 to 0.630. The RBF kernel scores have a shifted distribution between 0.51 to 0.55. The number of features and the highest value of R^2_{score} in each kernel for iteration is shown in Table 4.

Table 4 The Summary of PSO result

Kernel	R^2_{score}	Features	Average	STD
Linear	0.613	251	0.589	0.030
Polynomial	0.629	240	0.595	0.033
RBF	0.523	218	0.512	0.009

The summary of best solution from PSO is shown in Table 4. We found the best optimal features selected on linear, polynomial and RBF kernel with R^2_{score} value are 251, 240, and 218, respectively. Each kernel also has different numbers of the features. We also discovered that the RBF kernel produced the lowest standard deviation (±0.009) compared to the other kernel. This indicates that the PSO solution in the RBF kernel performed almost similarly in every multiple-run scheme.

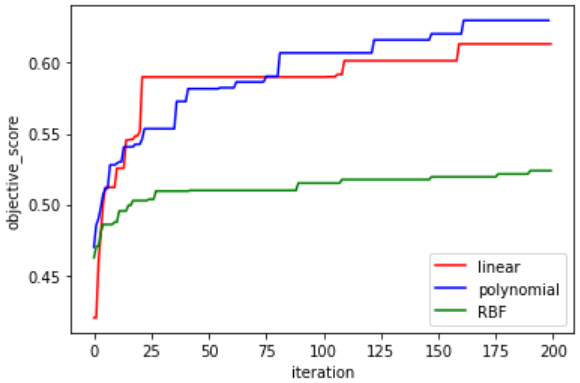


Figure 4. The convergent of PSO solution

Figure 4 shows the convergent plot for each kernel starts to produce stable fitness scores and found the global optimal around iteration number 165 to 190. This indicates that the PSO solutions are not premature to convergent value because on each kernel start to converge at relatively same time in last iteration.

3.2 Hyperparameter Tuning

We developed three prediction models using the SVM algorithm with difference combination of features according to the PSO results for predict the pIC50 target values. In addition, hyperparameter tuning is performed to find the best SVM parameters. The hyperparameter tuning result is shown in Table 5.

Table 5. The summary of hyperparameter tuning result

Parameter	Kernel		
	Linear	Polynomial	RBF
C	1(1)	1000 (1)	1 (1)
Gamma	-	Auto (scale)	Scale (scale)
Degree	-	3 (3)	-

*The value contained in the bracket is the model's default parameter value of SVM.

From Table 5 we can see that the value of parameter C for linear and RBF kernels is equal to 1, but the polynomial kernel's C parameter is equal to 1000. Gamma parameter for polynomial and RBF are set to auto and scale, respectively. The comparison of performance before and after tuning based on the R^2_{score} is shown in Figure 5.

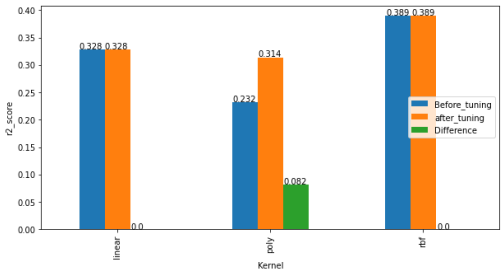
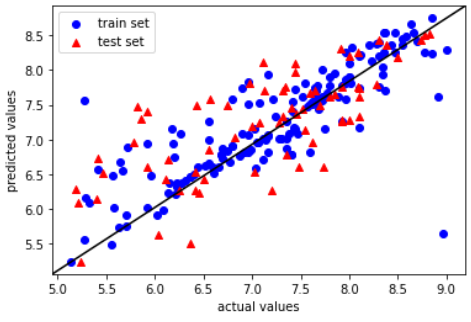


Figure 5. The comparison of performance before hyperparameter tuning and after hyperparameter tuning

From Figure 5 we can see that there is an increment in performance base on R^2_{score} for polynomial kernel by 0.082 while there is no performance increasing for both RBF and linear.

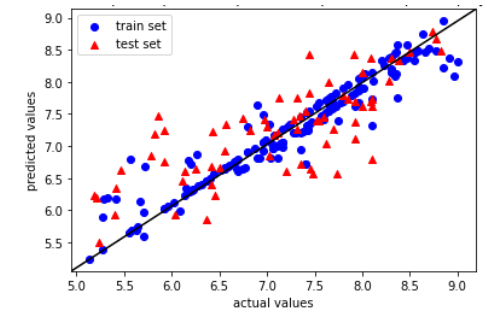
3.3 Validation Model

The prediction result from each kernel model is visualized using a scatter plot to compare the actual values and predicted values of pIC50, from training data and test data. The regression plot between predicted value and actual value for each kernel are shown in Figure 6.

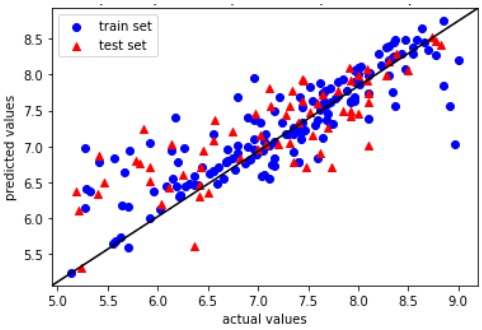


(a)

DOI: https://doi.org/10.29207/resti.v6i5.4294



(b)



(c)

Figure 6. The regression plot between of predicted value and actual value in (a) linear; polynomial; and (c) RBF in SVM kernels

From Figure 6 we found there the predicted values from training data and test data on polynomial and RBF kernel model are satisfactory. Meanwhile the linear kernel model prediction has several data points with significant deviation against actual values. This result indicates that the RBF and polynomial model performed better than the linear kernel model.

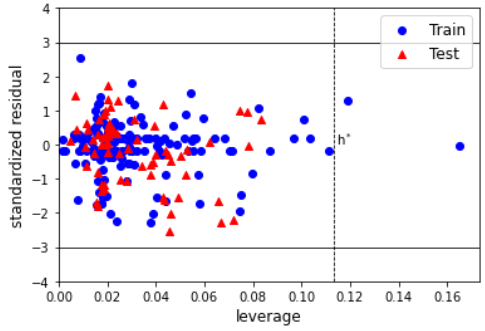
Table 6. Parameter validation of training data

Parameter	Training data			
	Linear	Polynomial	RBF	Threshold [9]
R^2	0.69	0.90	0.75	>0.6
Q^2_{loo}	0.48	0.51	0.51	>0.5
k'	1.00	0.97	1.00	0.85<k'<1.15
$\frac{(r^2 - r_0^2)}{r^2}$	0.00	0.00	0.00	<0.01
$\lvert r_0^2 - r_0^{`2}\rvert$	0.10	0.02	0.12	<0.3

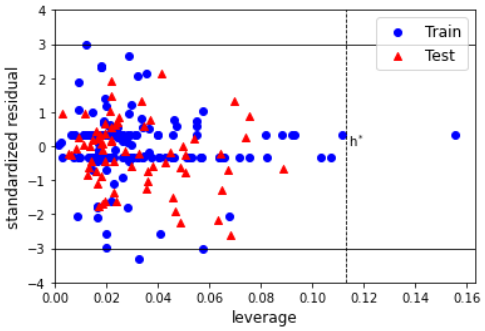
Table 7. Parameter validation of test data

Parameter	Test data			
	Linear	Polynomial	RBF	Threshold [9]
R^2	0.55	0.59	0.67	>0.6
k'	1.04	0.93	1.03	0.85<k'<1.15
$\frac{(r^2 - r_0^2)}{r^2}$	0.00	0.00	0.01	<0.01
$\lvert r_0^2 - r_0^{`2}\rvert$	0.25	0.12	0.25	<0.3

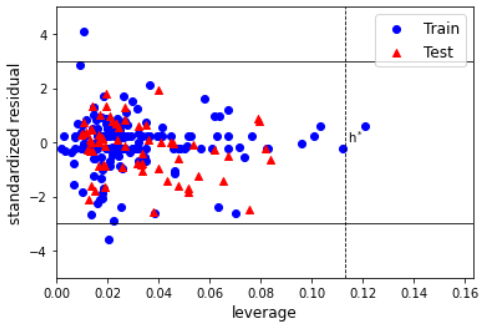
Table 6 and Table 7 shows, the model can be accepted if the result of the calculation on parameter validation is satisfied by the specified threshold. Internal validation of the SVM model with each kernel is acceptable, because the value of R^2_{train} is within the threshold and the highest value obtained from internal validation is the polynomial kernel with R^2_{train} and Q^2_{loo}values are 0.90 and 0.51 respectively. However, on external validation only the RBF kernel is acceptable, because the value of R^2_{test} from RBF models only satisfied by the threshold. Based on internal and external validation parameters, the RBF kernel model obtained the highest value of R^2_{train} and the R^2_{test} of 0.75 and 0.67, respectively. The standardized residual on each kernel is shown in Figure 7.



(a)



(b)



(c)

Figure 7. Applicability domain model in linear (a), polynomial (b), and RBF kernels (c).

Applicability Domain model in Figure 7 shows, that in the test data, the distribution of data points is still within the domain boundaries. However, in the training data there are several data points that are outside the domain boundaries, such as the linear kernel has two data points, the polynomial kernel has two data points, and the RBF kernel has three data points. It is certain that the model built has the potential to be applied in drug discovery because the dominant data point is still within the domain boundaries.

4. Conclusion

We developed model for QSAR Study on DGAT-1 inhibitor as anti-diabetic by utilizing PSO algorithm to

DOI: https://doi.org/10.29207/resti.v6i5.4294

reduce the features number and hyperparameter tuning is performed to improve the model and to produce the best parameter on the model. Based on the validation result we found that the best model satisfied with internal and external validation was developed by the RBF kernel model. As from the internal validation we obtained the score with R^2_{train} and Q^2_{loo} are 0.75 and 0.51, respectively. Then, in the external validation we obtained the R^2_{test} is 0.67.

Reference

[1] R. Joddy, S. Putra, A. Achmad, and H. Rachma, "Pharmaceutical Journal Of Indonesia Kejadian Efek Samping Potensial Terapi Obat Anti Diabetes Pasien Diabetes Melitus Berdasarkan Algoritma Naranjo," *PHARMACEUTICAL JOURNAL OF INDONESIA*, vol. 2017, no. 2, pp. 46–46, 2017, [Online]. Available: http://.pji.ub.ac.id

[2] S. Pangribowo, "Infodatin-2020-Diabetes-Melitus," *Pusat Data dan Informasi Kementrian Kesehatan RI*, pp. 1–10, 2020.

[3] "IDF Diabetes Atlas Ninth edition 2019," 2019. doi: 10.1016/S0140-6736(55)92135-8.

[4] P. Kumar, A. Kumar, and J. Sindhu, "In silico design of diacylglycerol acyltransferase-1 (DGAT1) inhibitors based on SMILES descriptors using Monte-Carlo method," *Taylor Francis Online*, vol. 30, no. 8, pp. 525–541, Jul. 2019, doi: 10.1080/1062936X.2019.1629998.

[5] M. R. Keyvanpour and Shirzad Mehrnoush Barani, "An Analysis of QSAR Research Based on Machine Learning Concepts," *PubMed*, vol. 18, no. 1, pp. 17–30, 2021, doi: 10.2174/1570163817666200316104404.

[6] U. Saqib and M. I. Siddiqi, "3D-QSAR studies on triazolopiperazine amide inhibitors of dipeptidyl peptidase-IV as anti-diabetic agents," *SAR and QSAR in Environmental Research*, vol. 20, no. 5–6, pp. 519–535, 2009, doi: 10.1080/10629360903278677.

[7] A. M. Al-Fakih, Z. Y. Algamal, M. H. Lee, M. Aziz, and H. T. M. Ali, "A QSAR model for predicting antidiabetic activity of dipeptidyl peptidase-IV inhibitors by enhanced binary gravitational search algorithm," *SAR and QSAR in Environmental Research*, vol. 30, no. 6, pp. 403–416, Jun. 2019, doi: 10.1080/1062936X.2019.1607899.

[8] K. Faghihi, M. Safakish, T. Zebardast, and A. Zarghi, "Molecular Docking and QSAR Study of 2-Benzoxazolinone, Quinazoline and Diazocoumarin Derivatives as Anti-HIV-1 Agents," *Iranian Journal of Pharmaceutical Research*, vol. 18, no. 3, pp. 1253–1263, 2019, doi: 10.22037/ijpr.2019.1100746.

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

[10] W. S. Dharmawan, "Dalam Prediksi Penyakit Jantung," *Jurnal Informatika, Manajemen dan Komputer*, vol. 13, no. 2, 2021.

[11] R. Amanullah Hakim, I. Kurniawan, and A. Aditsania, "Studi QSAR pada Senyawa Disulfida Aromatik sebagai Inhibitor SARS-CoV Mpro dengan Menggunakan Metode Genetic Algorithm-Support Vector Machine," *Jurnal Tugas Akhir Fakultas Informatika*, 2022.

[12] J. Kennedy and R. Eberhart, "Particle Swarm Optimization," *International Conference on Neural Networks*, vol. 4, no. 0, pp. 1942–1948, Nov. 1995, doi: 10.1109/icnn.1995.488968.

[13] R. Fahryandi, Y. Sibaroni, and A. Romadhony, "Klasifikasi Cyberbullying Terhadap Tokoh Publik Pada Komentar Sentimen Instagram Dengan Menggunakan Metode Support Vector Machine Dan Optimasi Fitur Berbasis Particle Swarm Optimization," Telkom University, Bandung, 2021.

[14] A. Nurlaily and M. T. S. R. Pradina Kusumawardani S.T., "Prediksi Diabetes Berdasarkan Faktor Risiko Behavioral Menggunakan Algoritma Support Vector Machine," Institut Teknologi Sepuluh November, Surabaya, 2018.

[15] N. Dini Maulana, B. Darma Setiawan, and C. Dewi, "Implementasi Metode Support Vector Regression (SVR) Dalam Peramalan Penjualan Roti (Studi Kasus: Harum Bakery)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 3, pp. 2986–2995, Mar. 2019, [Online]. Available: http://j-ptiik.ub.ac.id

[16] S. Aprilla, M. T. Furqon, and M. A. Fauzi, "Klasifikasi Penyakit Skizofrenia dan Episode Depresi Pada Gangguan Kejiwaan Dengan Menggunakan Metode Support Vector Machine (SVM)," 2018. [Online]. Available: http://j-ptiik.ub.ac.id

[17] N. P. Nanik Hendayanti, I. K. Putu Suniantara, and M. Nurhidayati, "Penerapan Support Vector Regression (Svr) dalam Memprediksi Jumlah Kunjungan Wisatawan Domestik ke Bali," *Jurnal Varian*, vol. 3, no. 1, Oct. 2019, doi: https://doi.org/10.30812/varian.v3i1.506.

[18] V. N. Vapnik, *The Nature of Statistical Learning Theory*, 2nd Edition. New York Berlin Heidelberg: Springer-Verlag, 1995.

[19] 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.

[20] 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 Komputer : Jurnal Ilmiah Teknologi Informasi*, vol. 12, no. 2, p. 62, Jul. 2021, doi: 10.24843/lkjiti.2021.v12.i02.p01.

DOI: https://doi.org/10.29207/resti.v6i5.4294