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. 6 (2022) 974 - 980 ISSN Media Electronic: 2580-0760

DPP IV Inhibitors Activities Prediction as An Anti-Diabetic Agent using Particle Swarm Optimization-Support Vector Machine Method

Reza Rendian Septiawan[1], Bambang Hadi Prakoso[2], Isman Kurniawan[2]
[1]School of Electrical Engineering, Telkom University
[2]School of Informatics, Telkom University
zaseptiawan@telkomuniversity.ac.id, bambangprakoso@students.telkomuniversity.ac.id,
ismankrn@telkomuniversity.ac.id*

Abstract

Diabetes mellitus is a chronic illness that can affect anyone, while the medicine that can entirely cure diabetes has not been discovered yet. Dipeptidyl Peptidase IV (DPP IV) inhibitor is one of the agents with potency as an anti-diabetic treatment. In this work, we utilized the machine learning method to predict the activity of DPP IV as an anti-diabetic agent. We combined Particle Swarm Optimization (PSO) method for features selection and the Support Vector Machine (SVM) for the prediction model. Three SVM kernels, i.e., radial basis function (RBF), polynomial, and linear, were utilized, and their performance was compared. A Hyperparameter tuning procedure was conducted to improve the performance of models. According to the results, we found that the best model obtained from SVM with RBF kernel with the value R2 of train and test set are 0.79 and 0.85, respectively.

Keywords: dipeptidyl peptidase IV inhibitor, particle swarm optimization, quantitative structure-activity relationship, support vector machine

1. Introduction

Diabetes mellitus (usually known as just diabetes) is a metabolic disorder caused by a loss of β-cells in the pancreas that affects insulin production [1]. Diabetes can be easily detected by a prolonged high blood sugar level. In general, diabetes can be divided into three types: type 1 diabetes, type 2 diabetes, and gestational diabetes [1], [2]. Type 1 diabetes is due to the loss of β-cells in the pancreas, causing a deficiency in insulin produced by the body. Type 2 diabetes is caused by cells' failure to properly respond to insulin. Gestational diabetes occurs in pregnant women and is caused by a sudden weight gain during a gestational period [3].

Diabetes can be treated by oral anti-diabetic drugs that are widely available, such as metformin [4]. Unfortunately, such drugs can have some side effects, for example, gas (flatulence) and diarrhea on metformin [5]. Therefore, research for new anti-diabetic agents is needed to overcome any problems with diabetic treatments. One agent that has potency in controlling blood sugar levels is the dipeptidyl peptidase IV (DPP IV) inhibitor. DPP IV inhibitor is a class of oral anti-diabetic drugs that inhibit DPP IV enzyme [6], [7]. Some researches that have been done on DPP IV inhibitor show their potency as a treatment for diabetes [8], [9]. To increase the effectivity of DPP IV inhibitor as an anti-diabetic agent, a structural optimization process is needed [10]. The drug design process can be accelerated by using a Quantitative Structure-Activity Relationship (QSAR) approach [11]. QSAR method is already proven to be effective in the drug design process by building a relationship between the activities of tested compounds with their molecular structures. Some models, such as regression models and classification models, can be used on building an efficient QSAR model in drug design [12].

In [13], Sharma et al. already developed a QSAR model for some derivatives of trifluorophenyl as DPP IV inhibitors by using 3D-QSAR Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Indice Analysis (CoMSIA). In their study, the model that they developed based on its structural alignment shows a good prediction with r^2 values are 0.963 and 0.934 for CoMFA and CoMSIA, respectively. Their model is useful for designing new DPP IV inhibitors. In [14], Jiang et al. developed a QSAR model for a set of arymethylamines as a DPP IV inhibitor by using a CoMFA approach with r^2 0.953. In [15], Patel and Ghate did a 3D-QSAR analysis by using CoMFA and CoMSIA on 36 derivatives of quinoline and

Accepted: 02-09-2022 | Received in revised: 11-11-2022 | Published: 29-12-2022

isoquinoline as DPP IV inhibitors that show values of conventional coefficient (r^2) 0.991 and 0.983 and values of correlation coefficient (r^2_{pred}) 0.874 and 0.847 for CoMFA and CoMSIA, respectively, for their best model. Saqib and Siddiqi [16] analyzed 45 derivatives of triazolopiperazime amida as DPP IV inhibitors by using a 3D-QSAR that shows values of r^2 is 0.868 for both CoMFA and CoMSIA approaches, and r^2_{pred} are 0.816 and 0.863 for CoMFA and CoMSIA, respectively, for their best model. All [12]–[16] show that CoMFA and CoMSIA approaches can be used in designing a new anti-diabetic agent effectively. One of the challenges in a QSAR study is to obtain an optimal number of features. This issue can be solved by implementing a meta-heuristic method to select features. However, to the best of our knowledge, there is no report of the implementation of a meta-heuristic method in the QSAR study on DPP IV inhibitors as anti-diabetic agents.

In this work, we aim to develop a QSAR model to predict the activities of DPP IV inhibitors by using Particle Swarm Optimization (PSO) – Support Vector Machine (SVM). In the first step, the PSO method is used to do a feature selection process to produce the best combination of features [17]. After that, the SVM method is used in the second step to getting the most accurate model. The SVM method itself is already used in many QSAR studies as a trusted method for building accurate models [18].

2. Research Method

In this work a dataset of DPP IV inhibitor compounds is used to build a model via a two-step process: (i) feature selection by a PSO method, and (ii) model building by an SVR method. Later the model is optimized by using a hyperparameter tuning process. After the model is optimized, finally the model is validated by using a Leave-One-Out Cross-Validation (LOO-CV) method.

2.1. Dataset

A dataset used in this work consists of 134 compounds as DPP IV inhibitors together with their half maximal inhibitory concentration values, IC_{50}, in nano-molar (nM), collected from works of literature [12]. The unit of IC_{50} is converted from nano-molar (nM) to molar (M), then the values of IC_{50} in molar are converted into pIC_{50} which is a negative logarithmic of IC_{50}, so now it shows a more potent inhibitor characteristic as the value of pIC_{50} increases. After that, molecular descriptors from all compounds are calculated by using PaDEL-descriptor software, resulting in 1875 molecular descriptors. Later all compounds are divided randomly into a training dataset and a test dataset with a ratio of 70:30 (107 compounds in a training dataset and 27 compounds in a test dataset). Such conversion steps with a descriptors calculation step and a random division step for training and test dataset are commonly done in QSAR calculation, such as in [19]. The distribution of the pIC_{50} value of all compounds can be seen in Figure 1.

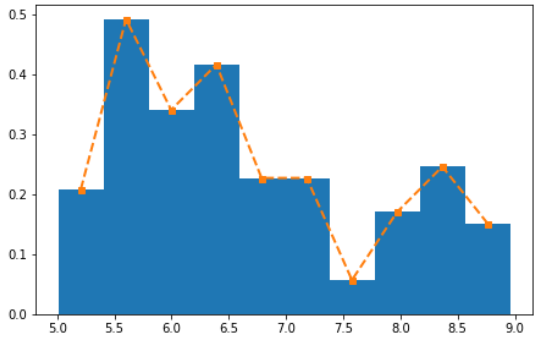

Figure 1. The Distribution of pIC50 Activities

2.2. Features Selection

In the features selection step, the Pearson correlation coefficient (PCC) is used to reduce the number of descriptors. PCC is a metric in statistics that is used to measure the linear correlation between two sets of data. In this work, PCC is calculated as a criterion that determines the optimal reduction filter for descriptors [20]. Before calculating the PCC value of compounds, some descriptors with zero variance and standard deviation less than 0.95 are removed. After that, PCC analysis is used to remove descriptors that bring similar information to other descriptors. Descriptors with weak correlation with target (PCC value < 0.1) and strong correlation with other descriptors (PCC value > 0.9) are removed [21]. From 1875 descriptors, 100 descriptors with the highest correlation are chosen. After that, to select the best descriptor, a Particle Swarm Optimization (PSO) algorithm is used. PSO algorithm is an optimization algorithm which is invented by Kennedy and Eberhart based on the behavioral actions of a swarm of birds [22]. Here, the performance of a swarm of particles is evaluated on every iteration by using equations (1) and (2) as follows [23]:

$$p_b(i,t) = \arg\min[f(\Pi_i)], i \in \{1,2,\dots,N_p\}, \quad (1)$$

$$g_b(t) = \arg\min[f(p_i(t))], i \in \{1,2,\dots,N_p\}, \quad (2)$$

where $p_b(i,t)$ is the best-known position for particle i at an iteration step t, $g_b(t)$ is the best position of the entire swarm at an iteration step t, Π_i is a set of arguments for fitting function with $\Pi_i \in \{p_b(i,t-1), P_i(t)\}$, f is a fitting function, and $P_i(t)$ is a position of particle i at a given iteration t. The position and velocity of particle i are updated by using the following equations:

$$V_i(t+1) = \omega V_i(t) + c_1 r_1 \Delta p_i(t) + c_2 r_2 \Delta g_i(t), \quad (3)$$

$$P_i(t+1) = P_i(t) + V_i(t+1), \quad (4)$$

where $V_i(t)$ is a velocity of particle i at a given iteration t, ω is an inertia weight, c_1 and c_2 are positive constants called cognitive coefficient and social coefficient, respectively, r_1 and r_2 are random-generated numbers

DOI: https://doi.org/10.29207/resti.v6i6.4470

with $r_1, r_2 \in [0,1]$, Δp_i and Δg_i are calculated by $\Delta p_i(t) = p_b(i,t) - P_i(t)$ and $\Delta g_i(t) = g_b(t) - P_i(t)$. The first term of equation (3) represents the inertia-weighted velocity from the previous iteration. The second term of (3), called a cognitive term, provides momentum for each particle to move guided by the best-known position in its own search space. The third term of (3), called a social term, guides the movement of each particle by the swarm's best-known position [24]. Equation (4) updates the position of each particle by using their velocity from equation (3).

2.3. Support Vector Machine (SVM)

The Support Vector Machine (SVM) method is a learning algorithm that is based on statistical learning frameworks on some given samples to construct a hyperplane that can be used for classification or regression [25]. There are two types of SVM: Support Vector Classification (SVC) and Support Vector Regression (SVR) [26]. The goal of SVM is to construct the best hyperplane that can divide a given set of samples into two different classes in the n-dimensions space. The best hyperplane maximizes the distance (margin) between the hyperplane with the nearest training data points (called support vectors) from both sides [27]. An example of the classification of data by using a linear hyperplane can be seen in Figure 2.

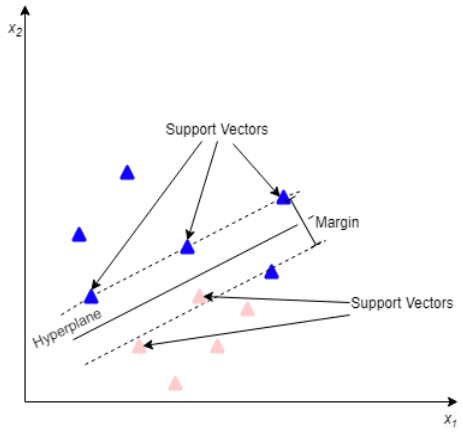

Figure 2. Example of Linear Classification of Data into Two Classes

To find the ownership of point x can be calculated by using the following equation:

$$f(x) = (\sum_{i=1}^{n} \alpha_i y_i \langle x_i, x \rangle) + b, \quad (5)$$

where x_i is a training data point, $y_i = 1$ or $y_i = -1$ shows the ownership of point x_i belongs to which class, n is a number of training data points, α_i is a Lagrange multiplier for point i, and $\langle , \rangle$ is an inner product operator. The sign of $f(x)$ shows the ownership of point x. In many cases, the classification process cannot be done correctly in a limited dimension of the space so the classification process needs to be done in a higher dimension. In that cases, the ownership of point x is calculated by the following equation:

$$f(x) = (\sum_{i=1}^{n} \alpha_i y_i \langle \phi(x_i), \phi(x) \rangle) + b, \quad (6)$$

where ϕ is a mapping function from the original dimension to the higher dimension. To add another dimension to a data point, several types of functions – which are called kernel functions – can be used. Several popular kernel functions are radial basis function (RBF), polynomial, and linear [28]. In Support Vector Regression (SVR), the goal is to build a model so that no output falls outside a specified margin ϵ from the model [29]. Three already mentioned kernel functions that are commonly used in SVR can be written as:

$$K(X,Y) = X^T Y, \quad (7)$$

$$K(X,Y) = (\Upsilon . X^{\mathrm{T}} Y + r)^d; \Upsilon > 0; d = (1,2,\dots), \quad (8)$$

$$K(X,Y) = \exp\left[\frac{\|X-Y\|^2}{2\sigma^2}\right], \quad (9)$$

where (7)-(9) are linear kernel functions, polynomial kernel functions, and RBF kernel functions, respectively. The most optimum model is the one with the smallest value of Root Mean Square Error (RMSE) [18].

2.4 Hyperparameter Tuning

Hyperparameter tuning is used to improve the performance of the model [21]. In SVR, hyperparameter tuning on the dataset and feature selection is used to maximize the performance of the prediction [29]. To optimize parameters in SVR, a Particle Swarm Optimization (PSO) method is used [30]. A list of all parameters on SVR that need to be optimized is shown in Table 1.

Table 1. A List of Parameters on SVR that Need to be Optimized and Their Range of Values

Parameter	Range of value
Kernel	[RBF, Linear, Polynomial]
C	[0.1, 1, 10, 100, 1000]
Gamma	['auto', 'scale']
Degree	[1, 2, 3, 4, 5]
Epsilon	[0.1, 1, 10, 100, 1000]

The kernel parameter determines the prediction model used by the SVR method. The options for the kernel are RBF, linear, and polynomial functions. The C parameter is the cost value which determines the penalty value for data located outside the margin area. The gamma parameter is the value of the coefficient in a kernel function. The degree parameter is a degree coefficient in a polynomial kernel. The epsilon parameter is the error margin allowed between the data and the regression line [31].

2.5 Model Validation

In this work, the generated model will be validated by using a Leave-One-Out Cross-Validation (LOO-CV) method. LOO-CV works by removing one molecule from the original training dataset and then generating the QSAR model again based on the remaining dataset.

DOI: https://doi.org/10.29207/resti.v6i6.4470

Then the activity of the removed molecule can be measured from equations produced by QSAR. This cycle is repeated until all molecules from the training dataset are already removed once and the activities of all molecules in the training data set are already calculated which are used in calculations of internal validation parameters [32]. This model is used to predict the pIC50 of all molecules in the training dataset [33]. R^2 value represents a correlation level between observed and predicted activities data, shown in (10) as $(y_i - \bar{y})$ and $(\hat{y}_i - \bar{\hat{y}})$, respectively. $\bar{y}$ is an average of molecular activities in the training dataset, and $\bar{\hat{y}}$ is an average of molecular activities in the test dataset. y_i and $\hat{y}_i$ show an experimental and predicted pIC_{50} value of a molecule. k shows the slope of the regression data, shown in (11). r_o^2 shown in (12) represents the correlation between the quadratic coefficient and predicted activity value without intercept. In [34], the calculation of Q^2 which is based on the test prediction shown in (13). Equation (14) shows the correlation between the quadratic coefficient with predicted activity data. Equations (15) – (16) show parameters that represent overall internal and external contributions to validation techniques to check the external predictability of the QSAR model. Equation (17) shows a difference of an average of randomized quadratic coefficient correlation.

$$R^2 = \frac{\left(\Sigma(y_i-\bar{y})(\hat{y}_i-\bar{\hat{y}})\right)^2}{\Sigma(y_i-\bar{y})^2 \times \Sigma(\hat{y}_i-\bar{\hat{y}})^2}. \quad (10)$$

$$k = \frac{\Sigma(y_i \times \hat{y}_i)}{\Sigma(\hat{y}_i)^2}. \quad (11)$$

$$r_o^2 = 1 - \frac{\Sigma(y_i - k \times \hat{y}_i)^2}{\Sigma(y_i-\bar{y})^2}. \quad (12)$$

$$Q^2 = 1 - \frac{\Sigma_{i=1}^{n}(\hat{y}_i - y_i)^2}{\Sigma_{i=1}^{n}(y_i-\bar{y})^2}. \quad (13)$$

$$r_m^2 = r^2 \times (1 - \sqrt{r^2 - r_o^2}. \quad (14)$$

$$\overline{r_m^2} = \frac{r_m^2 + {r'_m}^2}{2}. \quad (15)$$

$$\Delta r_m^2 = \left| r_m^2 + {r'_m}^2 \right|. \quad (16)$$

$$^{c}R_p^2 = R \times \sqrt{R^2 - R_r^2} \quad (17)$$

Thresholds for each validation parameter for a model to be accepted are shown in equations (18) – (23).

$$Q^2 > 0.5. \quad (18)$$

$$r^2 > 0.6. \quad (19)$$

$$\frac{r^2 - r_0^2}{r^2} < 0.1 \quad (20)$$

$$\left| r_0^2 - r_0'^2 \right| < 0.3. \quad (21)$$

$$\overline{r_m^2} > 0.5. \quad (22)$$

$$\Delta r_m^2 < 0.2. \quad (23)$$

3. Results and Discussions

To determine the best model, the accuracy of the model is used as the main criterion. In this work, QSAR modeling is done with several different numbers of descriptors (5, 10, 15, 20, and 25 descriptors) on each RBF, polynomial, and linear kernel. The mean-squared error (MSE) on each model for each descriptor can be seen in Figure 3.

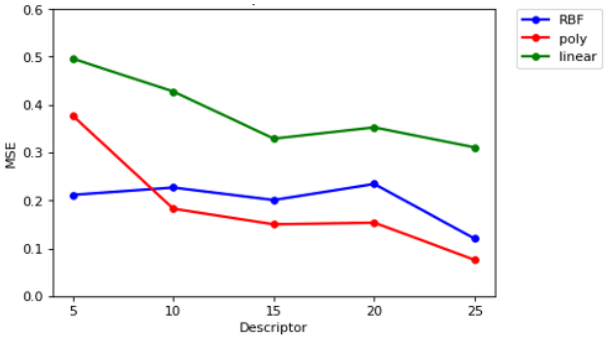

Figure 3. MSE of Each Model for Each Number of Descriptors.

From Figure 3, we can see that all three models show a tendency to have a smaller value of MSE, which is good as the number of descriptors increases. All three models show the smallest number of MSE when the number of descriptors is set to 25. This indicates that the increase in descriptor number corresponds to the improvement of the model performance. However, we limit the number of descriptors to 25 descriptors to avoid too complex a model.

The profile of the feature selection process presented in the plot of MSE corresponds to iteration, as shown in Figure 5. We found that the MSE in the first six iterations significantly decreased. Then, the error gradually decreases in the next iteration. Figure 5 also points out that the optimization process is done as expected, which is indicated by the decreasing of MSE during the iteration.

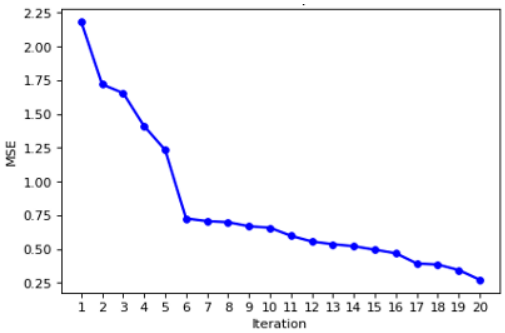

Figure 4. The Graph of MSE vs Iteration

After performing feature selection, we optimize the model by conducting a hyperparameter tuning process. The optimal parameter of the SVM model for each kernel is presented in Table 2. We found that the value of the C parameter of the polynomial and linear kernel is

DOI: https://doi.org/10.29207/resti.v6i6.4470

similar. Besides, the optimal value of the epsilon parameter is similar for all kernels.

Table 2. The Results of the Hyperparameter Tuning Process

Kernel	C	gamma	degree	epsilon
RBF	10	scale	-	0.1
Polynomial	1	-	3	0.1
Linear	1	-	1	0.1

We plot the predicted value of pIC50 against the actual one to get an insight into the model performance, as shown in Figure 6. The deviation between the plot with the diagonal line indicates the magnitude of the error. We found that the deviation of data in the RBF kernel is relatively smaller than in other kernels. Meanwhile, the deviation of data in polynomial dan linear kernels is quite similar. This deviation will directly correspond to the validation parameter value.

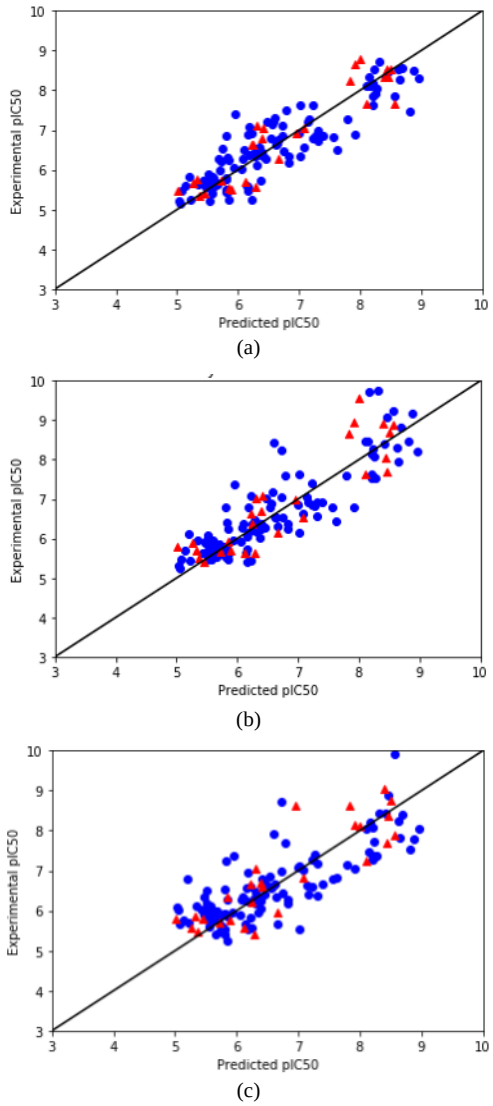

Figure 6. The Graph of Experimental pIC_{50} vs Predicted pIC_{50} for (a) RBF, (b) Polynomial, and (c) Linear Model

We calculated several validation parameters to evaluate the model performance and compared them with the threshold value, as shown in Table 3. As for the train set, we found that all models satisfied the threshold, in which SVM with RBF kernel gives the best value of the Q^2 parameter. This indicates that the RBF kernel is suitable for transforming the train set into a new dimension that is more linearly separable. However, by considering the test set, we found that SVM with a polynomial kernel is not valid because $\frac{r^2-r_0^2}{r^2}$ parameters do not satisfy the threshold. In this study, we consider the R^2 value of the test set to determine the best model. By comparing the R^2 value, we found that SVM with RBF kernel also gives the best performance in the test set. This point out the general ability of the RBF kernel to map out the dimension of both the train and test set. The outperform of the RBF kernel is related to the flexibility of this kernel to transform the data set.

Table 3. Testing Result of QSAR Model

Parameter	RBF		Poly		Linear		Thre
	Train	Test	Train	Test	Train	Test	shold
R^2	**0.79**	**0.85**	**0.75**	**0.80**	**0.62**	**0.76**	> 0.6
Q^2	**0.79**	-	**0.72**	-	**0.61**	-	> 0.5
$\frac{r^2-r_0^2}{r^2}$	**0.008**	**0.05**	**0.06**	0.10	**0.04**	**0.09**	< 0.1
$\lvert r^2-r_0^2\rvert$	**0.003**	**0.03**	**0.04**	**0.07**	**0.01**	**0.06**	< 0.3
$\overline{r_m^2}$	**0.72**	**0.73**	**0.67**	**0.65**	**0.51**	**0.64**	> 0.5
Δr_m^2	**0.01**	**0.11**	**0.15**	**0.15**	**0.02**	**0.17**	< 0.2
$^cR_p^2$	**0.65**	-	**0.73**	-	**0.67**	-	> 0.5

Regarding the applicability of the model, we analyzed the applicability domain (AD) by evaluating leverage values, as shown in Figure 7. The rectangle in the Figure indicates the domain of model applicability. According to the Figure, we found that only one train data and two test data are lying outside the region for the RBF kernel. This indicated that the model is applicable to almost all of the data set. Finally, we evaluated the probability of a systematical error occurring in the RBF model by presenting a plot of residual error, as shown in Figure 8. We confirmed that is no systematical error found in the model according to the pattern presented in the Figure.

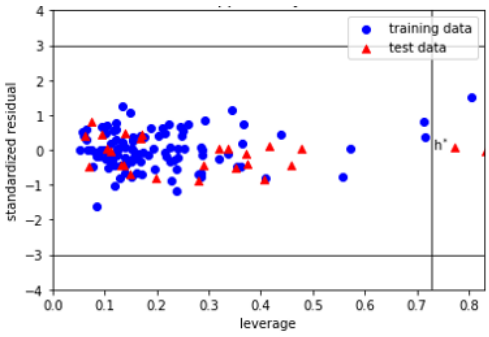

Figure 7. Applicability Domain for RBF Kernel

DOI: https://doi.org/10.29207/resti.v6i6.4470

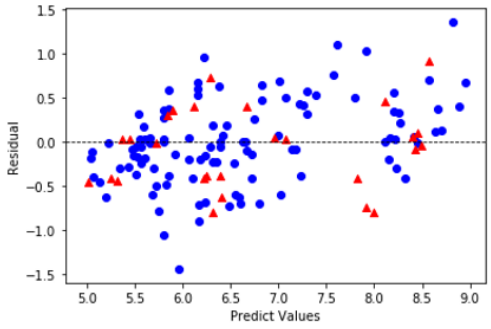

Figure 8. Residual Graph of RBF Model

4. Conclusion

We have developed prediction models to predict the activities of DPP IV inhibitors as an anti-diabetic agent using the Particle Swarm Optimization-Support Vector Machine (PSO-SVM). According to the results, we found that the PSO algorithm can be used to obtain the optimal number of features. The performance of the model was improved after conducting a hyperparameter tuning procedure. Based on the validation results, we found that the SVM model with RBF kernel gives the best results, with the R2 score of the train and test set being 0.79 and 0.85, respectively.

References

[1] American Diabetes Association, "Diagnosis and classification of diabetes mellitus," *Diabetes Care*, vol. 27 Suppl 1, pp. S5–S10, Jan. 2004. doi: 10.2337/diacare.27.2007.s5.

[2] "Standards of Medical Care in Diabetes—2015 Abridged for Primary Care Providers," *Clin. Diabetes Publ. Am. Diabetes Assoc.*, vol. 33, no. 2, pp. 97–111, Apr. 2015. doi: 10.2337/diaclin.33.2.97.

[3] S. Anazawa, "[Gestational diabetes mellitus]," *Nihon Rinsho Jpn. J. Clin. Med.*, vol. 73, no. 12, pp. 2015–2021, Dec. 2015.

[4] B. D. Green, P. R. Flatt, and C. J. Bailey, "Dipeptidyl peptidase IV (DPP IV) inhibitors: a newly emerging drug class for the treatment of type 2 diabetes," *Diab. Vasc. Dis. Res.*, vol. 3, no. 3, pp. 159–165, Dec. 2006. doi: 10.3132/dvdr.2006.024.

[5] D. Kirpichnikov, S. I. McFarlane, and J. R. Sowers, "Metformin: an update," *Ann. Intern. Med.*, vol. 137, no. 1, pp. 25–33, Jul. 2002. doi: 10.7326/0003-4819-137-1-200207020-00009.

[6] E. Kristin, "DIPEPTIDYL PEPTIDASE 4 (DPP-4) INHIBITORS FOR THE TREATMENT OF TYPE 2 DIABETES MELLITUS," *J. Med. Sci. Berk. Ilmu Kedokt.*, vol. 48, no. 2, Art. no. 2, Dec. 2016. doi: 10.19106/JMedSci004802201606.

[7] J. J. Holst and C. F. Deacon, "Inhibition of the activity of dipeptidyl-peptidase IV as a treatment for type 2 diabetes," *Diabetes*, vol. 47, no. 11, pp. 1663–1670, Nov. 1998. doi: 10.2337/diabetes.47.11.1663.

[8] P. R. Flatt, C. J. Bailey, and B. D. Green, "Dipeptidyl peptidase IV (DPP IV) and related molecules in type 2 diabetes," *Front. Biosci.*, vol. 13, no. 10, pp. 3648–3660, May 2008. doi: 10.2741/2956.

[9] C. F. Deacon, "Dipeptidyl peptidase-4 inhibitors in the treatment of type 2 diabetes: a comparative review," *Diabetes Obes. Metab.*, vol. 13, no. 1, pp. 7–18, Jan. 2011. doi: 10.1111/j.1463-1326.2010.01306.x.

[10] X. Yang, M. Li, Q. Su, M. Wu, T. Gu, and W. Lu, "QSAR studies on pyrrolidine amides derivatives as DPP-IV inhibitors for type 2 diabetes," *Med. Chem. Res.*, vol. 22, no. 11, pp. 5274–5283, Nov. 2013. doi: 10.1007/s00044-013-0527-2.

[11] E. Estrada, "On the topological sub-structural molecular design (TOSS-MODE) in QSPR/QSAR and drug design research," *SAR QSAR Environ. Res.*, vol. 11, no. 1, pp. 55–73, 2000. doi: 10.1080/10629360008033229.

[12] 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 QSAR Environ. Res.*, vol. 30, no. 6, pp. 403–416, Jun. 2019. doi: 10.1080/1062936X.2019.1607899.

[13] M. C. Sharma, S. Jain, and R. Sharma, "Trifluorophenyl-based inhibitors of dipeptidyl peptidase-IV as antidiabetic agents: 3D-QSAR COMFA, CoMSIA methodologies," *Netw. Model. Anal. Health Inform. Bioinforma.*, vol. 7, no. 1, p. 1, Dec. 2017. doi: 10.1007/s13721-017-0163-8.

[14] C. Jiang, S. Han, T. Chen, and J. Chen, "3D-QSAR and docking studies of arylmethylamine-based DPP IV inhibitors," *Acta Pharm. Sin. B*, vol. 2, no. 4, pp. 411–420, Aug. 2012. doi: 10.1016/j.apsb.2012.06.007.

[15] B. D. Patel and M. D. Ghate, "3D-QSAR studies of dipeptidyl peptidase-4 inhibitors using various alignment methods," *Med. Chem. Res.*, vol. 24, no. 3, pp. 1060–1069, Mar. 2015. doi: 10.1007/s00044-014-1178-7.

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

[17] Z. Wang, G. L. Durst, R. C. Eberhart, D. B. Boyd, and Z. B. Miled, "Particle swarm optimization and neural network application for QSAR," in *18th International Parallel and Distributed Processing Symposium, 2004. Proceedings.*, Apr. 2004, pp. 194-. doi: 10.1109/IPDPS.2004.1303214.

[18] H. Nguyen, "Support vector regression approach with different kernel functions for predicting blast-induced ground vibration: a case study in an open-pit coal mine of Vietnam," *SN Appl. Sci.*, vol. 1, no. 4, p. 283, Mar. 2019. doi: 10.1007/s42452-019-0295-9.

[19] I. Kurniawan, D. Tarwidi, and Jondri, "QSAR modeling of PTP1B inhibitor by using Genetic algorithm-Neural network methods," *J. Phys. Conf. Ser.*, vol. 1192, p. 012059, Mar. 2019. doi: 10.1088/1742-6596/1192/1/012059.

[20] J. Benesty, J. Chen, and Y. Huang, "On the Importance of the Pearson Correlation Coefficient in Noise Reduction," *IEEE Trans. Audio Speech Lang. Process.*, vol. 16, no. 4, pp. 757–765, May 2008. doi: 10.1109/TASL.2008.919072.

[21] I. Kurniawan, M. Rosalinda, and N. Ikhsan, "Implementation of ensemble methods on QSAR Study of NS3 inhibitor activity as anti-dengue agent," *SAR QSAR Environ. Res.*, vol. 31, no. 6, pp. 477–492, Jun. 2020. doi: 10.1080/1062936X.2020.1773534.

[22] M. Zamani, M. Karimi-Ghartemani, N. Sadati, and M. Parniani, "Design of a fractional order PID controller for an AVR using particle swarm optimization," *Control Eng. Pract.*, vol. 17, no. 12, pp. 1380–1387, Dec. 2009. doi: 10.1016/j.conengprac.2009.07.005.

[23] Hindawi, "Artificial Intelligence and Its Applications 2014." https://www.hindawi.com/journals/mpe/2016/3871575/ (accessed Sep. 15, 2022).

[24] Hindawi, "A Comprehensive Survey on Particle Swarm Optimization Algorithm and Its Applications." https://www.hindawi.com/journals/mpe/2015/931256/ (accessed Sep. 15, 2022).

[25] "Photonic neural networks and learning machines | IEEE Journals & Magazine | IEEE Xplore." https://ieeexplore.ieee.org/document/163674 (accessed Sep. 15, 2022).

DOI: https://doi.org/10.29207/resti.v6i6.4470

[26] S. Shamshirband *et al.*, "Support vector regression methodology for wind turbine reaction torque prediction with power-split hydrostatic continuous variable transmission," *Energy*, vol. 67, pp. 623–630, Apr. 2014. doi: 10.1016/j.energy.2014.01.111.

[27] P. Sihag, P. Jain, and M. Kumar, "Modelling of impact of water quality on recharging rate of storm water filter system using various kernel function based regression," *Model. Earth Syst. Environ.*, vol. 4, no. 1, pp. 61–68, Apr. 2018. doi: 10.1007/s40808-017-0410-0.

[28] F. Wang, Z. Zhen, B. Wang, and Z. Mi, "Comparative Study on KNN and SVM Based Weather Classification Models for Day Ahead Short Term Solar PV Power Forecasting," *Appl. Sci.*, vol. 8, no. 1, Art. no. 1, Jan. 2018. doi: 10.3390/app8010028.

[29] S. Xu, B. Lu, M. Baldea, T. F. Edgar, and M. Nixon, "An improved variable selection method for support vector regression in NIR spectral modeling," *J. Process Control*, vol. 67, pp. 83–93, Jul. 2018. doi: 10.1016/j.jprocont.2017.06.001.

[30] Z. Zhong and T. R. Carr, "Application of mixed kernels function (MKF) based support vector regression model (SVR) for CO2 – Reservoir oil minimum miscibility pressure prediction," *Fuel*, vol. 184, pp. 590–603, Nov. 2016. doi: 10.1016/j.fuel.2016.07.030.

[31] S. M. S. Nugroho, I. A. Budiastuti, and M. Hariadi, "Predicting daily consumer price index using support vector regression method based cloud computing," in *2017 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, Aug. 2017, pp. 313–318. doi: 10.1109/ISITIA.2017.8124101.

[32] K. Roy and I. Mitra, "On various metrics used for validation of predictive QSAR models with applications in virtual screening and focused library design," *Comb. Chem. High Throughput Screen.*, vol. 14, no. 6, pp. 450–474, Jul. 2011. doi: 10.2174/138620711795767893.

[33] B. Sepehri and R. Ghavami, "Design of new CD38 inhibitors based on CoMFA modelling and molecular docking analysis of 4-amino-8-quinoline carboxamides and 2,4-diamino-8-quinazoline carboxamides," *SAR QSAR Environ. Res.*, vol. 30, no. 1, pp. 21–38, Jan. 2019. doi: 10.1080/1062936X.2018.1545695.

[34] G. Schüürmann, R.-U. Ebert, J. Chen, B. Wang, and R. Kühne, "External Validation and Prediction Employing the Predictive Squared Correlation Coefficient — Test Set Activity Mean vs Training Set Activity Mean," *J. Chem. Inf. Model.*, vol. 48, no. 11, pp. 2140–2145, Nov. 2008. doi: 10.1021/ci800253u.

DOI: https://doi.org/10.29207/resti.v6i6.4470