WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Biophysical approach of modeling inhibition potency of acetylcholinesterase for possible Alzheimer disease treatment using genetic algorithm optimized support vector regression

James Ibitoye Agbi*

Physics and Electronics Department, Adekunle Ajasin University, Akungba Akoko, Ondo State, Nigeria
jamesagbiibitoye@gmail.com ; +2347043273991

ABSTRACT :

For the treatment of Alzheimer's diseases, Acetylcholinesterase (AChE) is considered a therapeutic target. As a result, modelling inhibition potency of acetylcholinesterase of different compounds for possible Alzheimer disease treatment is important in the process of finding better drugs for the treatment of Alzheimer diseases. In this work, a genetic algorithm optimized support vector regression (GA-SVR) was deployed to model inhibition potency of diverse compounds which comprise oximes, N-hydroxy-iminoacetamides, 4-aminoquinolines and flavonoids. The model was developed using the following descriptors; topological descriptor- the valence molecular connectivity index of the zero-order $^0\chi v$,) as well as two constitutional descriptors: nOH (total number of hydroxyl groups) and nR10 (number of 10-membered rings). A correlation coefficient of 97.71 per cent was obtained between the GA-SVR and the measured values of the inhibition potency of the human AChE, pK_i. For the same compounds, the QSAR results from the literature achieved a correlation coefficient of 94.70 per cent. In terms of prediction accuracy, the root mean square error was determined to be 0.3428 and 0.5120 for the GA-SVR and QSAR, respectively. This translates to over 49.34 per cent improvement of the proposed GA-SVR over the existing QSAR model. This demonstrates that the developed GA- SVR model could reliably predict the inhibition potency of oximes, N-hydroxy-iminoacetamides, 4-aminoquinolines and flavonoids better than the popular QSAR technique existing in the literature.

Keywords: Alzheimer, support vector regression, Acetylcholinesterase, genetic algorithm and inhibition potency

1.0 Introduction

The progressive loss of neural cells has been identified as the main factor contributing to the dysfunctional nervous system which may manifest as neurodegenerative diseases like Alzheimer's disease (AD) and Parkinson's disease[1]. Specifically, approximately 36 million people have been reported to have suffered from AD globally and it is projected that the figure may approach 66 million by 2030. This places AD as an emergency public health challenge amongst the aged population[2][1]. Unfortunately, the race to develop an efficacious cure or preventive therapy for AD has not been won yet. Consequently, research into drug discovery for AD is currently a fascinating research path [1]. AD's symptoms include progressive loss of memory, reduction in language and cognitive skills, behavioral anomaly and eventually death. The pathological effects of AD correlate with a huge neurons loss in different parts of the central nervous system which consequently leads to a substantial reduction in the numbers of neurotransmitters [3]. Acetylcholine (ACh) is a neurotransmitter with a function relating to vital muscular and cognitive actions. For example, in the peripheral nervous system, the ACh is located at the neuromuscular junction and is responsible for muscle contraction. The ACh found in the central nervous system is responsible for learning, thought, and memory [1]. The cholinergic hypothesis suggests that the drastic reduction of ACh in AD patients is attributable to activities of Acetylcholinesterases (AChE), an enzyme, that speeds up the breakdown of ACh to choline and acetic acid [4]. Therefore, a promising therapeutic strategy is to stabilize the level of ACh by inhibiting the activities of the enzyme breaking down the ACh. Acetylcholinesterase inhibitors (AChEIs) have proven effective in treating Alzheimer's disease [3]. A few of the drugs that are used for treating Alzheimer's disease include donepezil, rivastigmine, galanthamine, and tacrine. These drugs are generally referred to as AChE inhibitors (AChEIs). However, some of the concerns with these drugs include the question of the longevity of the inhibitors and side effects of drugs such as causing cardiovascular diseases. These concerns revealed that further research efforts in terms of discovering new and safe drugs are still required on AChE inhibitors [5].

In alleviating the concerns raised, the Quantitative Structure-Activity Relationship (QSAR) and other data-driven techniques have been used prominently to study the interaction of organic compounds and various biological activities [2], [3], [5]–[7]. The main purpose of the QSAR is to develop a mathematical model that will elucidate the relationship between organic structures and different biological activities. Thus providing a means to predict the binding affinity of new molecules based on the known inhibitors and other relevant molecular parameters. Several authors have suggested that the QSAR approach can aid rapid design and discovery of drugs for various diseases and eventually reduce the routine and time-consuming procedure often encountered in drug development [2]. As a result, numerous data-driven models have been developed for predicting the activity of AChEIs[2], [6]–[8]. For example, in a recent study [9], a QSAR model was formulated for estimating the acetylcholinesterase inhibitory activity of a new set of 72 tertiary

amine derivatives of cinnamic acids. The model was based on four descriptors such as HVcpx, VEA2, HATS5u, and qpmax indices, selected by the Enhanced Replacement Method (ERM). Their results exhibited a good correlation between the predicted and actual values acetylcholinesterase inhibitory activity, to the factor of $(R^2)_{train}$ = 0.913 and $(R^2)_{test}$ =0.856. Similarly, [2] developed a multilayer perceptron network (MLP) based model for predicting the acetylcholinesterase inhibitory activity for N-benzylpiperidines using 10 descriptors. They achieved a correlation coefficient of (R^2) = 0.882. [10] Studied the inhibitory activity towards cholinesterase disease from a series of N-aryl derivatives using the QSAR approach. The inhibitory activity of the N-aryl derivatives was described using the following descriptors MATS4m, nHDon, MLOGP2, and EEig10r. Their model achieved a squared correlation coefficient and cross –validated correlation coefficient of 94.5 % and 91.9 % with respect to acetylcholinesterase inhibitory activity and the predicted values. Estimation of the inhibitory activity (IC50) of different Coumarin compounds using a number of techniques such as ANN, SVM, PLS, MLR, and RF has also been reported [10]. They found that the ANN tuned with GA has the best predicted accuracy as measured by squared correlation coefficient of 96 %. As revealed from the literature survey, the QSAR is the most famous technique used for predicting the activity of AChEIs and it has so far achieved satisfactory results, albeit at a higher computational cost. However, machine learning techniques such as support vector regression (SVR) has been deemed to have remarkable generalizability with lower computational time. Unfortunately, there are few studies to have explored the powerful generalizability of SVR in predicting the activity of AChEIs. In this contribution, we explore the use of a hybrid model of genetic algorithm (GA) and support vector regression model to predict the inhibitory activity of structurally different inhibitors which comprise oximes, N-hydroxyiminoacetamides, 4-aminoquinolines and flavonoids, for the first time. To highlights the superior performance of the developed model, we benchmark the predictive accuracy of our model with the results of QSAR recently developed. The analysis revealed that the developed hybrid genetic algorithm and support vector regression is more accurate for predicting the activity of AChEIs eith correlation coefficient of over 97%.

Support vector regression (SVR) is an intelligent machine learning modeling method based on structural risk minimization principle. The algorithm is characterized with unique properties of data transformation to feature space with the aid of mapping function. This uniqueness coupled with its strong mathematical formulation has strengthen the algorithm in solving many real life problems in science , medical and engineering fields [11]–[15]. The hyper-parameters contained in SVR algorithm have strong influence on the performance of the algorithm and can be tuned and selected through grid-search, heuristic or manual search approaches. The optimization method employed in this present work is genetic algorithm due to its significant features which include fast convergence, avoidance of premature convergence and easy convergence to global solution.

The rest of the manuscript is arranged as follows: section two presents and discusses the mathematical formulation of the hybridized algorithm while section three presents the computational methodology coupled with data set description as well as acquisition. Section four discusses and presents the results of the developed hybrid model as well as the comparison between the developed and existing QSAR model in the literature. Section five presents the conclusion.

2.0. Mathematical foundations of the hybridized algorithms

This section details the mathematical formulation of the developed support vector regression model, along with steps on how the genetic algorithm optimizes the user-defined parameters of the model.

2.1. Support vector regression mathematical derivation

Support vector regression (SVR) is powerful non-linear modeling technique that finds its root in statistical learning theory [16], [17]. The goal of the algorithm is to establis approximate functional relationship $g(x)$ between inhibition potency of acetylcholinesterase and its descriptors, including the molecular connectivity index of zero-order, the number of 10-membered rings and the total number of hydroxyl group. One key condition for the function generated within the SVR framework is that the resulting deviation for all the compounds under study should not go beyond a specific threshold. At the same time, the function needs to maintain its flatness condition [18], [19]. These goals are achieved by employing the kernel trick and introducing a mapping function to transform the complex nonlinear problems to linear ones. For a standard $\varepsilon-\text{insensitive}$ loss function based SVR algorithm implemented in this work in which the inhibition potency of acetylcholinesterase (represented as I in the mentioned set of data) of all the studied compounds along with their associated descriptors (D) are included in a set $\mathrm{A}=\{\mathrm{D_j}, I_j\}$, where $I_j \in \square$ and $j = 1,2,3,.....,M$, M is the number of investigated acetylcholinesterase compounds. The algorithm's approximated function is presented in equation (1) [20].

$$g(D) = \omega^T \chi(D) + b \quad (1)$$

Where $\omega \in \square^M$ and $b \in \square$

The flatness of the approximated function, which links acetylcholinesterase inhibition potency to its descriptors, is achieved by solving a convex optimization problem. This involves minizing equation 2 while adhering to the contrainst outlined in equation 3

$$\frac{\|\omega\|^2}{2}+C\sum_{j=1}^{N}\left(\xi_j+\xi_j^*\right) \tag{2}$$

$$\begin{cases} \langle \omega, \chi(D)\rangle + b - g_j \le \varepsilon + \xi_j \\ g_j - \left(\langle \omega, \chi(D)\rangle + b\right) \le \varepsilon + \xi_j^* \\ \xi_j^*, \xi_j \ge 0 \end{cases} \tag{3}$$

The penalty factor, denoted as C in equation (2), controls both the tolerance for deviations exceeding the predefined epsilon threshold and the flatness of the approximated function $g(D)$. Slack variables (ξ_j^* and ξ_j) are included to prevent external constraints from hindering the achievement of a flat function within the SVR framework [21, 22]. When performing the dual formulation, eliminating dual variables leads to the minimization of equation (4), subject to the conditions specified in equation (5) [19], [23].

$$\frac{1}{2}\sum_{j,i=1}^{N}\left(\gamma_j^*-\gamma_j\right)\left(\gamma_i^*-\gamma_i\right)\delta(D_j.D_i) + \varepsilon\sum_{j=1}^{N}\left(\gamma_j^*+\gamma_j\right)-\sum_{j=1}^{N}g_j\left(\gamma_j^*+\gamma_j\right) \tag{4}$$

$$\begin{cases} \sum_{j=1}^{N} g_j\left(\gamma_j^*-\gamma_j\right)=0 \\ 0 \le \gamma_j^*, \gamma_j \le C \end{cases} \tag{5}$$

Solving the optimization problem entails invoking Lagrange multipliers multipliers (γ_j and γ_j^*). Additionally, the Mercer condition is met when implementing the kernel function $\delta(D_j, D_i) = \vartheta^T(D_j)\vartheta(D_i)$. The parameter b is computed using the Karush-Kuhn-Tucker conditions, with the computational details provided in equations (6) and (7).

$$\omega = \sum_{j=1}^{N}\left(\gamma_j^*-\gamma_j\right)D_j \tag{6}$$

$$\begin{aligned} b &= g_j - \langle \omega, \chi(D)\rangle - \varepsilon \text{ for } 0 \le \gamma_j \le C \\ b &= g_j - \langle \omega, \chi(D)\rangle + \varepsilon \text{ for } 0 \le \gamma_j^* \le C \end{aligned} \tag{7}$$

The resulting regression function estimated by SVR algorithm is presented in equation (8) while the mathematical derivation for the optimum kernel function is detailed in equation (9) [24], [25].

$$g(D) = \sum_{j=1}^{N}\left(\gamma_j-\gamma_j^*\right)\delta(D_j, D_i)+b \tag{8}$$

$$\delta(D, D') = \exp\left(\frac{-\|D-D'\|}{2\beta^2}\right) \tag{9}$$

Where β is the kernel parameter

Although the SVR algorithm has a strong mathematical basis, its performance can suffer dramatically from poorly chosen penalty factor, epsilon values and kernel parameter. To overcome this, we've employed a genetic algorithm for parameter selection. This choice is due to the algorithm's fast

convergence and its reduced likelihood of settling into a local minimum.

2.2.Genetic algorithm operation procedures

A genetic algorithm is a prominent evolutionary population-based optimization algorithm that addresses many engineering and biomedical problems when implemented in continuous real-coded or binary form [2], [26]. Three different operations govern the optimization operational principle of genetic algorithm. The operators controlling these operations include selection, crossover and mutation operators. The selection procedure follows a random generation of population matrix that contains candidate solutions [27]–[29]. Each of the candidate solutions in this work encodes three parameters which include the penalty factor, kernel parameter and epsilon. The initially generated candidate solutions are known as the parents while the offsprings are generated from the parents for generation replacement with the aid of a crossover operator. Avoidance of local minima convergence and enhancement of genetic diversity is ensured through mutation operator. Genetic algorithm procedures are repeated until a stopping condition is met [30]. Elitism in genetic algorithm operational procedure leads the optimization to the optimal solution and further prevents mutation of best solutions. This promotes the transition of best solutions from one generation to another unaltered.

3.0. Computational methodology

The computational methods of this contribution are presented in this section. The description and acquisition of data employed for modeling is also presented.

3.1. Data acquisition and description

The accuracy of data-driven models such as QSAR or machine learning-based models is dependent on the descriptors chosen for the model development. Over the years, several descriptors have been advanced in the modeling and estimation of acetylcholinesterase inhibitors potency. These descriptors can be grouped into the following classes; constitutional, Gálvez topological charge indices, atom-centred fragments, topological, empirical, molecular walk counts, 2D autocorrelations, and functional group descriptor [31]. The descriptors used in this study consist of topological descriptor- the valence molecular connectivity index of the zero-order $^{0}\chi^{v}$,) as well as two constitutional descriptors: nOH (total number of hydroxyl groups) and nR10 (number of 10-membered rings) as reported by [9]. Using fewer descriptors to develop the model often contribute to the model's simplicity. This study used fewer descriptors to estimate the potency of acetylcholinesterase inhibitors. Our study is based on a group of 94 diverse compounds which comprise oximes, N-hydroxy-iminoacetamides, 4-aminoquinolines and flavonoids as previously reported by Sinko [9]. These compounds were initially studied using the QSAR model [9]. In this study, a GA-SVR model for the prediction of the inhibition potency of the human AChE, pK_i is investigated. A detailed description of the procedure employed for the measurement of the activity of AChE and subsequently Ki was based on Ellman method as explained in [9]. The statistical description of the dataset is presented in Table 1.

Table 1: Statistical description of the dataset used in this study

Variables	N	Mean	Minimum	Maximum	Range
PKi	94	5.023	2.190	8.220	6.030
$^{0}\chi^{v}$	94	13.968	5.960	26.070	20.110
nOH	94	1.50	0.000	10.000	10.000
nR10	94	0.4787	0.000	1.000	1.000

The presented values of the mean for each of the descriptors give insight about the nature and the content of the employed dataset. The presented maximum, minimum and range give information about the boundaries of the employed dataset from measurement to measurement.

3.2. Computational methodology

The algorithm development was conducted in MATLAB® 2020 environment. The dataset is randomly split, with 80% allocated and 20% for testing. The randomization of the data was done to avoid sample bias. The algorithm uses the training dataset (n=75) to rigorously learn the relationship between the mentioned descriptors and the target (inhibition potency of AChE, pKi) using cross-validation technique. This learning stage requires that the right set of SVR parameters is selected. These parameters consists of the kernel option, the kernel function, the epsilon and the regularization factor. Due to the range of options of each of the parameters, using the manual procedure to tune these parameters is often difficult and there is no guarantee that the optimum solution will be obtained. In order to address this, a genetic algorithm was employed to search for the optimal values of the SVR parameters. After the optimal values of the SVR parameters have been obtained, the performance of the trained model was assessed with three different performance measuring parameters using the testing dataset which was reserved for evaluating the robustness of the developed model.

4.0. Results and discussion

This section discusses the result of the developed hybrid GA-SVR model is presented in this section. The section also presents the convergence of the optimization algorithm for various population matrix. Likewise, the present model is compared against existing model using various performance

measuring parameters.

4.1. Optimization of user-defined hyper-parameters in SVR algorithm using GA

The performance influence of the number of population matrix in GA during optimization of the penalty factor, kernel parameter and epsilon on the convergence of the model is shown in Fig.1. Global convergence was attained with fifty candidate solutions while the increase in the initial number of candidate solutions to one hundred leads to poor and local minimum convergence as shown in the figure. Optimal selection of the number of probable candidate solutions controls the exploration and exploitation strengths of the optimization algorithm as the exploitation capacity becomes poorer as the number of possible solutions in the population matrix increases.

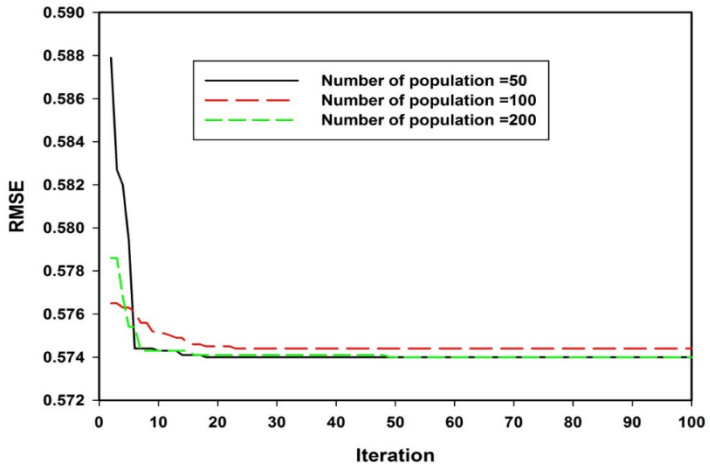

Fig.1: Convergence of GA-SVR model with increase in the population number

The convergence of the regularization (penalty) factor at every stage of optimization is presented in Fig.2 with a different population. The regularization factor for fifty number of the population shows the lowest convergence value as compared with others. Similar trends are observed for the variation of kernel option and epsilon at different stages of iteration as respectively presented in Fig.3 and Fig.4.

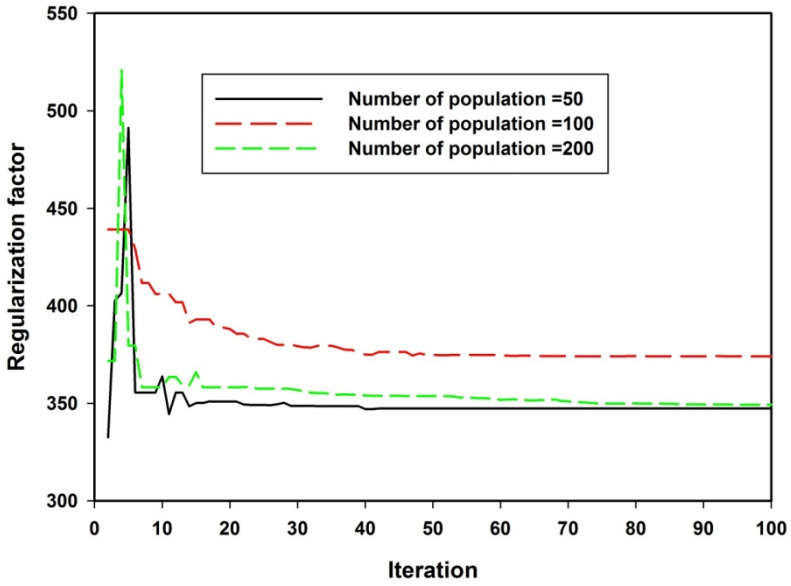

Fig.2: Convergence of GA-SVR model with support vector regression regulaization factor

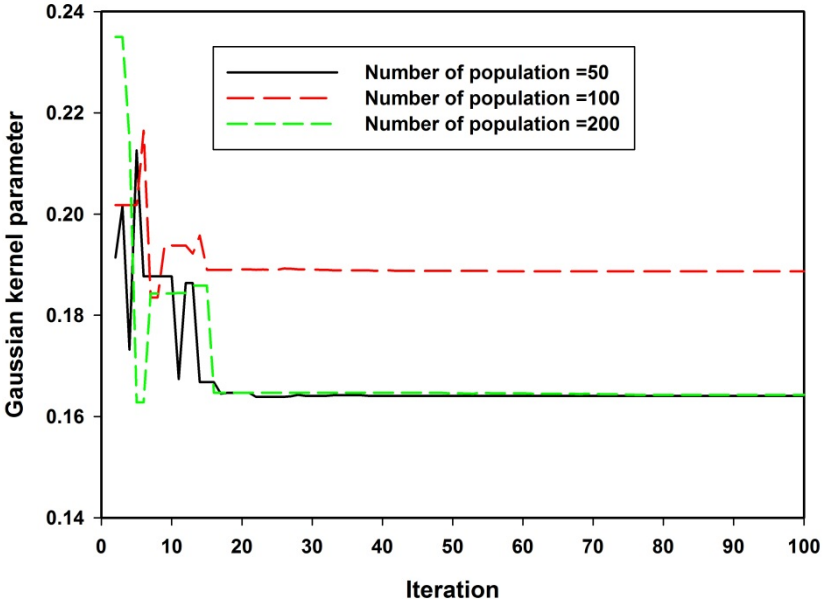

Fig.3: Convergence of GA-SVR model with support vector regression kernel option

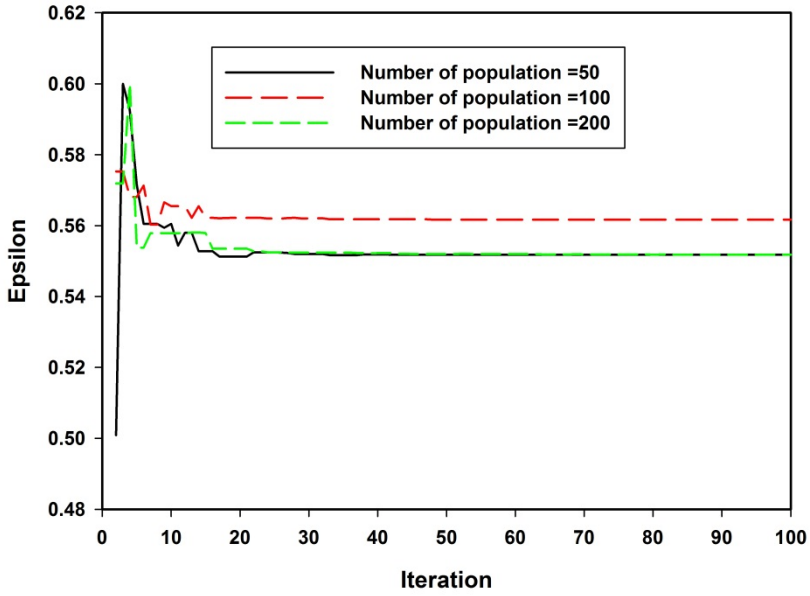

Fig.4: Convergence of GA-SVR model with support vector regression epsilon

The values of the user-defined parameters in the SVR algorithm as obtained from GA are presented in Table 2. The presented values in the table ensure the reproducibility of the presented results.

Table 2: SVR parameters and their optimum values obtained from GA

Hyper-parameters	Optimum value
Kernel option	0.1641
Regularization factor	347.2849
Epsilon	0.5518

Number of population	50
Hyper-parameter lambda	E-7
Kernel function	Gaussian

4.2. Performance comparison between the present and existing models

The performance of the hybrid GA-SVR model is assessed by employing root mean square error (RMSE), correlation coefficient (CC) and means absolute error (MAE). The performances of the model at the training and testing stage are presented in Table 3. Although the testing performance of the developed hybrid model is 93.97% however, the training performance is better than the testing results with a performance improvement of 4.83% on the basis of CC. Training performance improvements of 54.8% (RMSE)and 54.24% (MAE) were obtained over the testing set of data while comparing the two stages.

Table 3: Performance of the developed hybrid model at every developmental stage

	CC	RMSE	MAE
Training dataset	0.9874	0.2594	0.1968
Testing dataset	0.9397	0.574	0.4301

A correlation cross-plot between the experimentally measured acetylcholinesterase inhibition potency (AIP) and the estimated values using the presently developed GA-SVR and QSAR (2021) model [9] is presented in Fig.5. The superiority of the present GA-SVR model over the existing model is inferred from the alignment and closeness of the data points for GA-SVR estimates.

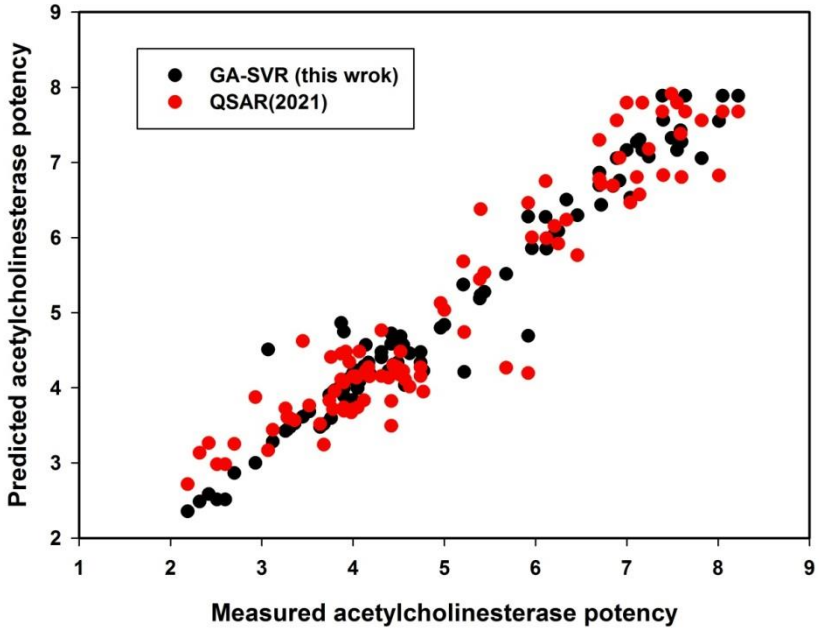

Fig.5:Correlation cross-plot between the measured and estimated acetylcholinesterase inhibition potency

Another performance comparison between present and developed models is presented in Fig.6, Fig.7 and Fig.8, respectively using RMSE, CC and MAE as performance evaluation metrics. The present GA-SVR model performs better than the existing QSAR (2021) [9] on the basis of RMSE presented in Fig.6, with a performance improvement of 49.34%. Using CC as performance evaluation yardstick presented in Fig.7, the developed GA-SVR model shows performance improvement of 3.08% over the existing model while percentage improvement of 65.2% was obtained for the comparison on the basis of MAE presented in Fig.8. Table 4 presents the values of each of the performance measuring parameters and the percentage improvement of the present model over the existing QSAR (2021) model.

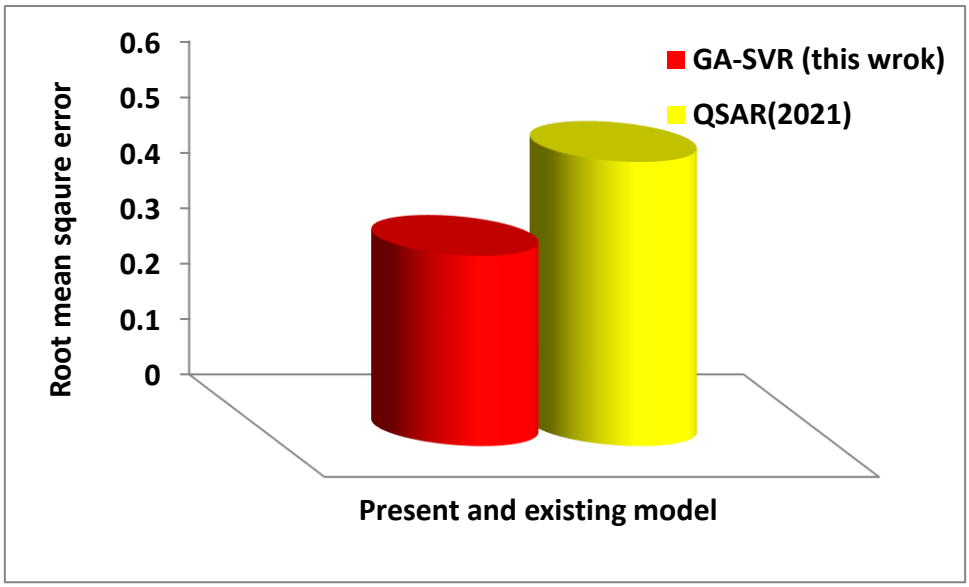

Fig.6: Performance improvement of the present GA-SVR model over the existing model on the basis of RMSE

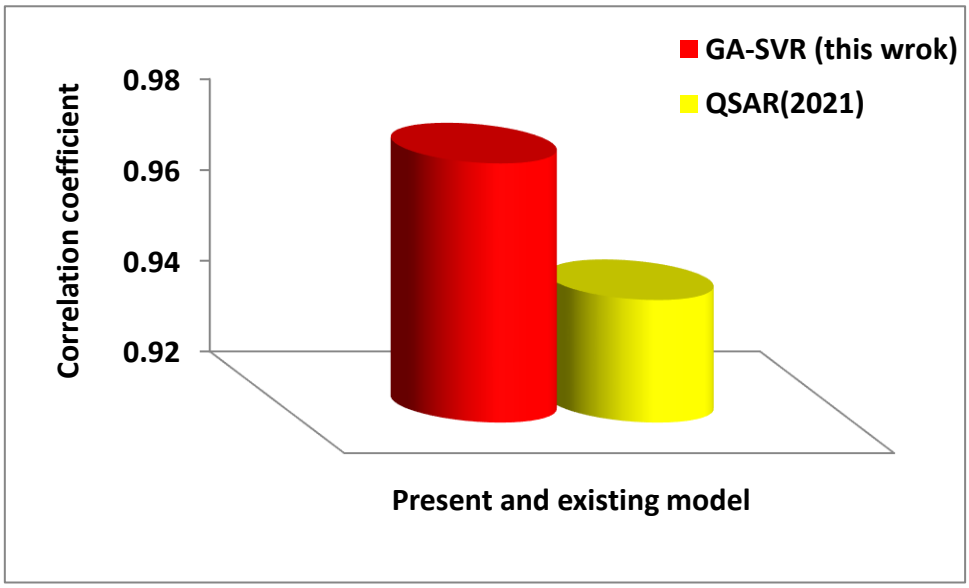

Fig.7: Performance improvement of the present GA-SVR model over the existing model on the basis of CC

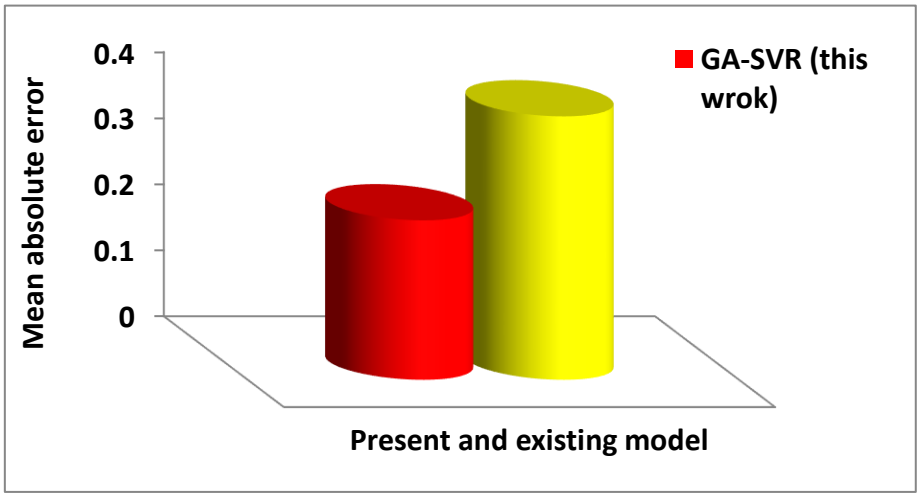

Fig.8: Performance improvement of the present GA-SVR model over the existing model on the basis of MAE

Table 4: Evaluation parameters of the present and existing model and the corresponding percentage improvement

	CC	% improvement	RMSE	% improvement	MAE	% improvement
GA-SVR	0.9771		0.3428		0.2415	
QSAR(2021) [9]	0.9470	3.08%	0.5120	49.34	0.3990	65.2

The comparison of the estimates of the present and existing model is presented in Table 5. The table also contains the percentage deviation of each of the compound from the measured values. The mean absolute percentage deviation (MAPD) for each of the model is also presented. On the basis of the presented MAPD, the present GA-SVR model outperforms the existing QSAR(2021) [9] model with a performance improvement of 71.76%. The obtained performance enhancement of the present model compared to existing model can be explained by the robust mathematical background of the implemented SVR algorithm, the low chance of local convergence possibility and excellent optimization genetic algorithm for the right combinatory choice of the model parameters.

Table 5: Comparison of the estimates of the present and existing model and their percentage error

S/N	AIC	GA_SVR	% error	QSAR	% error
1	4.55	4.57	0.41	4.22	7.19
2	4.51	4.63	2.75	4.17	7.48
3	4.62	4.46	3.55	4.01	13.12
4	4.14	4.57	10.36	4.22	2.00
5	3.96	4.12	4.14	4.34	9.63
6	4.77	4.23	11.40	3.95	17.25
7	3.74	3.90	4.39	3.83	2.37
8	4	3.84	4.10	3.71	7.21
9	3.31	3.47	4.96	3.59	8.55
10	4.74	4.33	8.56	4.27	9.89
11	5.92	4.69	20.77	4.19	29.15
12	5.68	5.52	2.89	4.26	24.92
13	3.76	3.60	4.36	4.41	17.24
14	4.44	4.31	3.01	4.30	3.20
15	2.93	3.00	2.36	3.87	32.21
16	3.45	3.61	4.76	4.62	33.92
17	4.31	4.40	2.17	4.76	10.49
18	3.87	4.86	25.63	4.45	15.03
19	4.96	4.80	3.31	5.13	3.37
20	5.39	5.19	3.75	5.45	1.04
21	5.44	5.28	3.02	5.53	1.65
22	6.34	6.50	2.59	6.24	1.61
23	6.21	6.05	2.64	6.15	0.90
24	6.11	6.27	2.69	6.75	10.48
25	5.4	5.24	3.04	6.38	18.08
26	5	4.84	3.28	5.03	0.67
27	4.07	4.08	0.35	4.48	10.08
28	4.03	4.19	4.07	4.15	3.01
29	4.42	4.58	3.71	3.82	13.53
30	4.42	4.72	6.82	3.49	20.98
31	4.18	4.19	0.34	4.15	0.69
32	4	4.19	4.85	4.15	3.78
33	3.92	4.08	4.19	4.48	14.30
34	3.52	3.68	4.66	3.76	6.89

35	2.51	2.51	0.09	2.98	18.72
36	2.32	2.48	7.07	3.13	35.07
37	3.07	4.51	46.82	3.17	3.15
38	2.19	2.35	7.49	2.72	24.02
39	2.6	2.51	3.37	2.98	14.61
40	3.68	3.52	4.46	3.24	11.90
41	2.7	2.86	6.08	3.25	20.45
42	4.52	4.68	3.63	4.48	0.86
43	3.64	3.48	4.51	3.52	3.41
44	3.36	3.52	4.88	3.57	6.14
45	3.9	3.89	0.29	3.69	5.31
46	4.12	4.28	3.98	3.83	7.00
47	3.78	3.94	4.34	3.71	1.73
48	3.98	3.82	4.12	3.67	7.75
49	4.05	3.99	1.43	3.74	7.70
50	3.9	3.99	2.36	3.74	4.15
51	3.9	4.75	21.69	4.07	4.33
52	7.82	7.05	9.79	7.56	3.34
53	8.22	7.89	4.06	7.68	6.62
54	8.05	7.89	2.04	7.68	4.65
55	7.55	7.16	5.11	7.79	3.23
56	7.17	7.16	0.08	7.79	8.70
57	7.49	7.33	2.19	7.91	5.64
58	6.89	7.05	2.38	7.56	9.70
59	7.64	7.89	3.22	7.68	0.47
60	7.39	7.89	6.71	7.68	3.86
61	7	7.16	2.34	7.79	11.34
62	2.42	2.58	6.78	3.26	34.89
63	3.12	3.28	5.26	3.44	10.25
64	3.28	3.44	5.00	3.61	9.91
65	3.26	3.42	5.03	3.72	14.22
66	5.22	4.21	19.35	4.74	9.22
67	3.8	3.96	4.32	3.96	4.10
68	4.49	4.33	3.47	4.27	4.87
69	4.17	4.33	3.94	4.27	2.43
70	4.39	4.23	3.74	4.13	5.86
71	4.04	4.23	4.60	4.13	2.30
72	4.57	4.03	11.73	4.11	10.08
73	3.87	4.03	4.24	4.11	6.18
74	4.74	4.47	5.61	4.15	12.35
75	4.31	4.47	3.81	4.15	3.61
76	5.21	5.37	3.15	5.68	9.06
77	6.25	6.09	2.63	5.92	5.30
78	6.72	6.43	4.26	6.71	0.14
79	8.01	7.55	5.72	6.83	14.78
80	5.92	6.28	6.05	6.46	9.12

81	7.04	6.53	7.25	6.47	8.14
82	6.7	6.86	2.45	7.30	8.93
83	7.14	7.30	2.30	6.57	7.96
84	7.24	7.08	2.27	7.18	0.88
85	7.4	7.56	2.22	6.83	7.72
86	7.6	7.27	4.29	6.80	10.47
87	7.11	7.27	2.31	6.80	4.30
88	6.92	6.76	2.37	7.06	2.04
89	7.59	7.43	2.16	7.38	2.74
90	6.46	6.30	2.54	5.77	10.75
91	5.96	5.86	1.76	6.00	0.71
92	6.12	5.85	4.40	5.99	2.14
93	6.7	6.69	0.08	6.78	1.15
94	6.85	6.69	2.40	6.69	2.28
	MAPD		**5.17**		**8.88**

5.0 Conclusion

In this study, a support vector regression optimized with a genetic algorithm was developed to predict the inhibition potency of acetylcholinesterase for diverse compounds which comprise oximes, N-hydroxy-iminoacetamides, 4-aminoquinolines and flavonoids. The GA/SVR proposed in this study was developed using the following descriptors; topological descriptor- the valence molecular connectivity index of the zero-order $^0\chi_v$,) as well as two constitutional descriptors: nOH (total number of hydroxyl groups) and nR10 (number of 10-membered rings. The GA/SVR developed for modelling the inhibition potency of the compounds showed a strong correlation of 97.71 % with the measured values. The root mean square values for the developed GA/SVR and existing QSAR model are 0.3428 and 0.5120, respectively. The shows an improvement of 49.34 per cent improvement for the GA/SVR model over the existing QSAR model. It can be inferred that the GA/SVR model proposed in this study demonstrate better predictive accuracy for the modelling inhibition potency of acetylcholinesterase for oximes, N-hydroxy-iminoacetamides, 4-aminoquinolines and flavonoids compared to the existing QSAR.

Data availability statement: The raw data required to reproduce these findings are available in the cited references in section 3.1 of the manuscript.

Competing interest

Author declares that there is no competing interest

REFERENCES

[1] K. K. Roy, A. Dixit, and A. K. Saxena, "An investigation of structurally diverse carbamates for acetylcholinesterase (AChE) inhibition using 3D-QSAR analysis," *J. Mol. Graph. Model.*, vol. 27, no. 2, pp. 197–208, Sep. 2008.

[2] S. Bitam, M. Hamadache, and S. Hanini, "QSAR model for prediction of the therapeutic potency of N-benzylpiperidine derivatives as AChE inhibitors," *SAR QSAR Environ. Res.*, vol. 28, no. 6, pp. 471–489, Jun. 2017.

[3] M. Recanatini, A. Cavalli, and C. Hansch, "A comparative QSAR analysis of acetylcholinesterase inhibitors currently studied for the treatment of Alzheimer's disease," *Chem. Biol. Interact.*, vol. 105, no. 3, pp. 199–228, Aug. 1997.

[4] S. Gharaghani, T. Khayamian, and M. Ebrahimi, "Molecular dynamics simulation study and molecular docking descriptors in structure-based QSAR on acetylcholinesterase (AChE) inhibitors," *SAR QSAR Environ. Res.*, vol. 24, no. 9, pp. 773–794, Sep. 2013.

[5] S. Abuhamdah, M. Habash, and M. O. Taha, "Elaborate ligand-based modeling coupled with QSAR analysis and in silico screening reveal new potent acetylcholinesterase inhibitors," *J. Comput. Aided. Mol. Des.*, vol. 27, no. 12, pp. 1075–1092, Dec. 2013.

[6] K. Y. Wong, A. G. Mercader, L. M. Saavedra, B. Honarparvar, G. P. Romanelli, and P. R. Duchowicz, "QSAR analysis on tacrine-related acetylcholinesterase inhibitors," *J. Biomed. Sci.*, vol. 21, no. 1, p. 84, Sep. 2014.

[7] M. Nekoeinia and S. Yousefinejad, "QSAR analysis of the acetylcholinesterase inhibitory activity of some tertiary amine derivatives of cinnamic acid," *Struct. Chem.*, pp. 1–10, Jan. 2021.

[8] L. AiLin, G. HongMei, Z. LiYa, D. GuanHua, L. M. Simon Y, and W. YiTao, "3D-QSAR analysis of a new type of acetylcholinesterase inhibitors," 2007.

[9] G. Sinko, "Development of a simple QSAR model for reliable evaluation of acetylcholinesterase inhibitor potency," vol. 160, no. February, 2021.

[10] M. S. M. Ridzuan, M. Z. Jaafar, and M. M. Zain, "Quantitative structure-activity relationship (QSAR) modelling of N-aryl derivatives as cholinesterase inhibitors," *SHUSER 2012 - 2012 IEEE Symp. Humanit. Sci. Eng. Res.*, no. March 2016, pp. 907–912, 2012.

[11] H. Zhang, Y. Zhang, D. Dai, M. Cao, and W. Shen, "Modelling and optimization of the superconducting transition temperature," *Mater. Des.*, vol. 92, no. 2016, pp. 371–377, 2015.

[12] N. Science, C. Phenomena, M. Sabzekar, S. Mohammad, and H. Hasheminejad, "Chaos , Solitons and Fractals," *Chaos, Solitons Fractals Interdiscip. J. Nonlinear Sci. Nonequilibrium Complex Phenom.*, vol. 144, p. 110738, 2021.

[13] S. Li, H. Fang, and B. Shi, "Remaining useful life estimation of Lithium-ion battery based on interacting multiple model particle filter and

support vector regression ☆," *Reliab. Eng. Syst. Saf.*, vol. 210, no. February, p. 107542, 2021.
[14] S. O. Olatunji and T. O. Owolabi, "Modeling superconducting transition temperature of doped MgB 2 superconductor from structural distortion and ambient temperature resistivity measurement using hybrid intelligent approach," *Comput. Mater. Sci.*, vol. 192, no. March, p. 110392, 2021.
[15] A. Balogun, F. Rezaie, Q. Bao, and L. Gigovi, "Geoscience Frontiers Spatial prediction of landslide susceptibility in western Serbia using hybrid support vector regression (SVR) with GWO , BAT and COA algorithms," vol. 12, 2021.
[16] V. N. Vapnik, "The Nature of Statistical Learning Theory," *Springer-Verlag New York, Inc*, vol. 70–92, Jun. 1995.
[17] T. O. Owolabi, M. Amiruddin, and A. Rahman, "Energy Band Gap Modeling of Doped Bismuth Ferrite Multifunctional Material Using Gravitational Search Algorithm Optimized Support Vector Regression," pp. 1–15, 2021.
[18] V. N. Vapnik, *Statistical Learning Theory*. New York, New York, USA: Wiley-Interscience, 1998.
[19] S. M. I. Shamsah and T. O. Owolabi, "Newtonian mechanics based hybrid machine learning method of characterizing energy band gap of doped zno semiconductor," *Chinese J. Phys.*, vol. 68, no. October, pp. 493–506, 2020.
[20] E. Dodangeh, M. Panahi, F. Rezaie, S. Lee, and D. Tien, "Novel hybrid intelligence models for flood-susceptibility prediction : Meta optimization of the GMDH and SVR models with the genetic algorithm and harmony search," *J. Hydrol.*, vol. 590, no. July, p. 125423, 2020.
[21] X. Ju, F. Liu, L. Wang, and W. J. Lee, "Wind farm layout optimization based on support vector regression guided genetic algorithm with consideration of participation among landowners," *Energy Convers. Manag.*, vol. 196, no. July, pp. 1267–1281, 2019.
[22] J. Rui, H. Zhang, D. Zhang, F. Han, and Q. Guo, "Journal of Petroleum Science and Engineering Total organic carbon content prediction based on support-vector-regression machine with particle swarm optimization," *J. Pet. Sci. Eng.*, vol. 180, no. June, pp. 699–706, 2019.
[23] M. W. Ahmad, J. Reynolds, and Y. Rezgui, "Predictive modelling for solar thermal energy systems : A comparison of support vector regression , random forest , extra trees and regression trees," *J. Clean. Prod.*, vol. 203, pp. 810–821, 2018.
[24] B. Wang, L. Zhou, K. Xu, and Q. Wang, "Prediction of minimum ignition energy from molecular structure using quantitative structure-property relationship (QSPR) models," *Ind. Eng. Chem. Res.*, vol. 56, no. 1, pp. 47–51, 2017.
[25] A. A. Adewunmi, S. Ismail, T. O. Owolabi, A. S. Sultan, S. O. Olatunji, and Z. Ahmad, "Modeling the thermal behavior of coal fly ash based polymer gel system for water reduction in oil and gas wells," *J. Pet. Sci. Eng.*, vol. 157, 2017.
[26] B. S. Obadawo, O. E. Oyeneyin, M. M. Anifowose, K. H. Fagbohungbe, and J. S. Amoko, "QSAR Evaluation of C-8-Tert-Butyl Substituted as Potent Anti-enterovirus Agents," *Sci. Lett.*, vol. 8, no. 1, pp. 28–35, 2020.
[27] Y. Wang and C. Wei, "Design optimization of office building envelope based on quantum genetic algorithm for energy conservation," *J. Build. Eng.*, vol. 35, no. December 2020, p. 102048, 2021.
[28] T. O. Owolabi, "Modeling the magnetocaloric effect of manganite using hybrid genetic and support vector regression algorithms," *Phys. Lett. Sect. A Gen. At. Solid State Phys.*, vol. 383, no. 15, 2019.
[29] L. Gharsalli and Y. Guérin, "Composites Part C : Open Access Mechanical sizing of a composite launcher structure by hybridizing a genetic algorithm with a local search method," *Compos. Part C Open Access*, vol. 5, no. February, p. 100125, 2021.
[30] A. Pourrajabian, M. Dehghan, and S. Rahgozar, "Genetic algorithms for the design and optimization of horizontal axis wind turbine (HAWT) blades : A continuous approach or a binary one ?," *Sustain. Energy Technol. Assessments*, vol. 44, no. February, p. 101022, 2021.
[31] E. B. Asadabadi, P. Abdolmaleki, S. M. H. Barkooie, S. Jahandideh, and M. A. Rezaei, "A combinatorial feature selection approach to describe the QSAR of dual site inhibitors of acetylcholinesterase," *Comput. Biol. Med.*, vol. 39, no. 12, pp. 1089–1095, Dec. 2009.