Original Article

Check for updates

Available online at
www.heca-analitika.com/malacca_pharmaceutics

Malacca Pharmaceutics

Vol. 1, No. 2, 2023

Integrating Genetic Algorithm and LightGBM for QSAR Modeling of Acetylcholinesterase Inhibitors in Alzheimer's Disease Drug Discovery

Teuku Rizky Noviandy [1], Aga Maulana [1], Ghazi Mauer Idroes [2,*], Nur Balqis Maulydia [3], Mohsina Patwekar [4], Rivansyah Suhendra [5] and Rinaldi Idroes [6]

[1] Department of Informatics, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; trizkynoviandy@gmail.com (T.R.N.); agamaulana@usk.ac.id (A.M.)
[2] Department of Occupational Health and Safety, Faculty of Health Sciences, Universitas Abulyatama, Aceh Besar 23372, Indonesia; idroesghazi_k3@abulyatama.ac.id (G.M.I.)
[3] Graduate School of Mathematics and Applied Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; Nur_balq1s@mhs.unsyiah.ac.id (N.B.M)
[4] Department of Pharmacology, Luqman College of Pharmacy, Karnataka 585102, India; mohsina.patwekar@gmail.com (M.P.)
[5] Department of Information Technology, Faculty of Engineering, Universitas Teuku Umar, Aceh Barat 23681, Indonesia; rivansyahsuhendra@utu.ac.id (R.S.)
[6] Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; rinaldi.idroes@usk.ac.id (R.I.)

***** Correspondence: idroesghazi_k3@abulyatama.ac.id

Article History

Received 23 May 2023
Revised 2 July 2023
Accepted 15 July 2023
Available Online 20 July 2023

Keywords:
AChE
Feature selection
Genetic algorithm
LightGBM
Machine learning
QSAR

Abstract

This study explores the use of Quantitative Structure-Activity Relationship (QSAR) studies using genetic algorithm (GA) and LightGBM to search for acetylcholinesterase (AChE) inhibitors for Alzheimer's disease. The study uses a dataset of 6,157 AChE inhibitors and their IC_{50} values. A LightGBM model is trained and evaluated for classification performance. The results show that the LightGBM model achieved high performance on the training and testing set, with an accuracy of 92.49% and 82.47%, respectively. This study demonstrates the potential of GA and LightGBM in the drug discovery process for AChE inhibitors in Alzheimer's disease. The findings contribute to the drug discovery process by providing insights about AChE inhibitors that allow more efficient screening of potential compounds and accelerate the identification of promising candidates for development and therapeutic use.

1. Introduction

Alzheimer's disease is a devastating neurodegenerative disorder that exacts a heavy toll on individuals, families, and society [1]. Marked by the gradual deterioration of cognitive functions, including memory loss, impaired reasoning, and behavioral changes, it poses significant challenges to those affected [2]. As one of the most prevalent forms of dementia, Alzheimer's disease affects millions of people worldwide, with its prevalence expected to rise due to the aging population [3]. The profound impact of this disease extends beyond the individual, affecting their loved ones and caregivers and placing a considerable burden on healthcare systems and society's resources. Addressing the complexities of Alzheimer's disease, finding effective treatments, and supporting those affected are necessary to improve the

well-being and quality of life of individuals facing this challenging condition. [4].

One of the solutions formulated to tackle this problem is the inhibition of acetylcholinesterase (AChE), an enzyme responsible for breaking down the neurotransmitter acetylcholine (ACh) [5]. By inhibiting AChE, the concentration of ACh in the inter-synaptic space can be significantly increased, leading to enhanced activity within the cholinergic system in the central nervous system. This approach holds promise for improving the functioning of neural pathways involved in memory and cognitive processes, offering a potential avenue for managing the symptoms of Alzheimer's disease and enhancing the quality of life for affected individuals [6, 7].

There is a fast and cost-effective approach to searching for AChE inhibitors, which involves conducting Quantitative Structure-Activity Relationship (QSAR) studies. QSAR examine the relationship between the chemical structure of compounds and their biological activities, so the researchers can predict the activity of AChE inhibitors, aiding in the development of more potent therapeutic agents for Alzheimer's disease [8–10].

QSAR studies is a computational drug discovery method [11, 12] that uses molecular descriptors to analyze the relationship between chemical structure and biological activity. These descriptors provide quantitative information about various molecular properties, such as size, shape, and electronic characteristics [13, 14]. However, since there can be thousands of molecular descriptors available for a given compound, it becomes crucial to select the most optimal descriptors for model simplicity and efficiency [15]. To simplify and speed up the process, genetic algorithm (GA), inspired by Darwin's theory, are commonly used [16, 17]. They select the most effective descriptors by simulating natural selection, gradually refining the set through iterations. This approach helps researchers predict the activity of AChE inhibitors, aiding the development of potent therapeutic agents for Alzheimer's disease.

In recent years, several studies have focused on QSAR for AChE inhibitors [18–21]. These investigations have employed various statistical and machine learning methods, which have shown promising results in achieving good performance. However, there still have room for improvement in terms of accuracy. One approach that can be done is to use recent machine learning algorithms such as LightGBM [22].

In this study, we propose an approach to QSAR using GA for molecular descriptors selection and LightGBM to train the QSAR model. With GA, we aim to identify the most relevant and informative molecular descriptors that contribute significantly to the inhibitory activity of AChE inhibitors. Subsequently, the selected descriptors are utilized in conjunction with LightGBM, a powerful and efficient machine learning algorithm, to build a robust QSAR model. The combination of GA and LightGBM holds the potential to improve the accuracy and predictive performance of the QSAR models for AChE inhibitors. This approach holds great promise for enhancing our understanding of structure-activity relationships and facilitating the drug discovery process for Alzheimer's disease.

2. Materials and Methods

2.1. Dataset

We collected data on 8832 AChE inhibitors from the ChEMBL database and their IC_{50} values [23]. Next, we removed duplicate data and left 6157 compounds. To carry out the classification process, we construct a class variable by converting the IC50 value to pIC50, and if the pIC_{50} value < 6, then the compound is assigned to an inactive class, and if $pIC_{50} \geq$, then the compound is active [24]. Among the 6157 compounds, 3591 of them (58.32%) were classified as inactive, and 2566 compounds (41.68%) were categorized as active.

2.2. Molecular Descriptors

In QSAR studies, molecular descriptors are used as features to build models. We used Mordred to derive 1661 2D-molecular descriptors for each AChE inhibitor compound. Molecular descriptors with high correlation (>0.95) and low variance (<0.1) were eliminated, leaving 280 molecular descriptors [24].

2.3. Feature Selection

Many molecular descriptors may be unimportant or duplicate for the classification task, thus necessitating the use of a feature selection technique. Feature selection involves choosing a subset of pertinent features from the initial feature set to enhance the performance of machine learning models. This process aids in reducing data dimensionality, preventing overfitting, enhancing model interpretability, and decreasing the computational burden of training models [25].

In this study, we use GA to select the most optimal molecular descriptor due to its ability to efficiently explore the extensive space of possible descriptors and identify concise and relevant representations of chemical features [26, 27]. GA works by iteratively generating heuristic solutions that represent different subsets of molecular descriptors. It then evaluates the fitness of each individual using a predefined fitness function that measures how well the individual solves the problem or

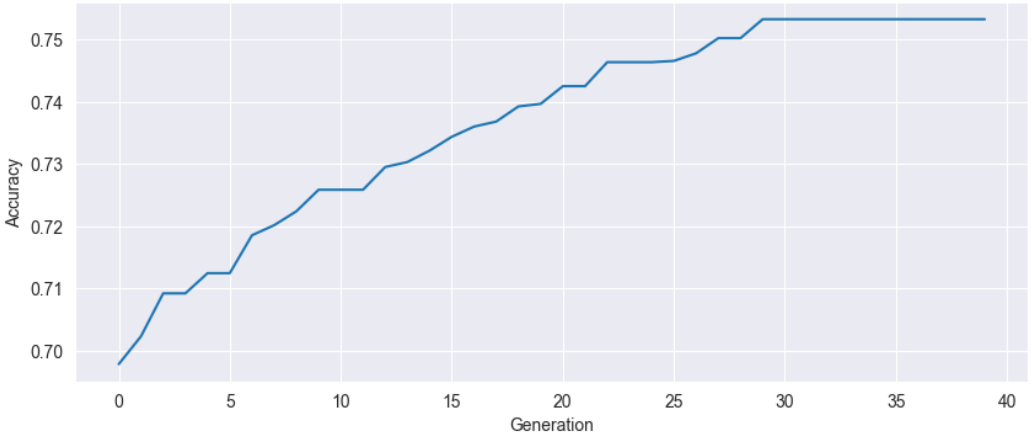

Figure 1. The result of GA for each generation.

represents the quality of the selected molecular descriptors [28, 29].

To initiate the GA-based feature selection process, we generated an initial population consisting of 50 random molecular descriptor subsets. The GA then utilized crossover and mutation operations to create new combinations of features. Crossover occurred with a 90% probability between selected subsets, while mutation took place with a 5% probability for each subset [30].

The GA process ran for 200 generations, evaluating the fitness of each subset with accuracy as the fitness function. Logistic regression was used as an estimator because of its simplicity and was run through 10-fold cross-validation to minimize overfitting and provide a more accurate estimation of the model's generalization ability. The best subsets became parents for the next generation, ensuring the gradual replacement of weaker subsets. To prevent the GA from getting stuck, we implemented a stopping criterion. If the best subset remained unchanged for ten consecutive generations, we terminated the algorithm, indicating convergence had been achieved [31].

2.4. LightGBM

LightGBM is an efficient gradient-boosting decision tree method known for its exceptional computational efficiency and remarkable accuracy [22, 32, 33]. It effectively tackles complex problems by efficiently handling large datasets, leading to faster training and prediction times. With its advanced algorithms and optimization techniques, LightGBM consistently achieves superior performance, enabling precise predictions across various domains [34, 35]. Because of that, LightGBM has gained widespread popularity among data scientists and machine learning practitioners for its ability to deliver accurate results efficiently.

2.5. Performance Evaluation

To assess the effectiveness of the proposed model, a comprehensive performance evaluation was conducted using four metrics: accuracy, precision, recall, and F1-score. These metrics provide insights into the model's classification performance across different aspects [36].

3. Results and Discussion

3.1. Genetic Algorithm Results

In this study, the GA was used to select the most optimal subset of molecular descriptors. Prior to running the GA, the data were divided randomly into two subsets, namely the training set and the testing set, in an 80:20 ratio.

The GA reached its termination point after the 40th generation, as there was no improvement in the fitness value for ten consecutive generations. The progression of the GA fitness value for each iteration is presented in Figure 1. The initial generation of GA yielded a fitness value of 0.698, which progressively increased to 0.753 by the 40th generation.

Table 1 presents the selected molecular descriptors, categorized into different types. The most abundant type of molecular descriptor selected was Estate, with five molecular descriptors. EState is a molecular descriptor that characterizes the electronic properties of molecules, providing valuable insights into their chemical behavior. In contrast, the type with the fewest descriptors includes AATS and information content, each having only one descriptor. AATS represents atom-type autocorrelation, analyzing the spatial distribution and connectivity of specific atom types. The information content descriptor encompasses information-theoretic measures related to molecular structures. Despite their limited number, these descriptors provide valuable information to the GA algorithm-selected feature subset.

Table 1. Selected molecular descriptors.

Type	Names	Definition
AATS	AATS4dv	Averaged moreau-broto autocorrelation of lag 4 weighted by valence electrons
Carbon Types	C1SP2	SP2 carbon bound to 1 other carbon
	C4SP3	SP3 carbon bound to 4 other carbons
EState	NssCH2	Number of ssCH2
	NdsCH	Number of dsCH
	NssssC	Number of ssssC
	NdO	Number of dO
	StCH	Sum of tCH
Fragment Complexity	fragCpx	Fragment Complexity
	ZMIC1	1-ordered Z-modified information content
Information Content	FilterItLogS	Predicted logarithm of solubility in water (LogS) using Filter-it model
Moe Type	PEOE_VSA2	MOE Charge VSA Descriptor 2
	PEOE_VSA12	MOE Charge VSA Descriptor 12
	SlogP_VSA10	MOE logP VSA Descriptor 10
	EState_VSA1	EState VSA Descriptor 1
Ring Count	nRing	Number of rings
	n6aHRing	Number of aromatic 6-membered rings
	nG12FRing	Number of 12- or more-membered rings (including fused rings)
	n9FaRing	Number of aromatic 9-membered rings

Table 2. Model performance.

Metrics	Training Set (%)	Testing Set (%)
Accuracy	92.49	82.47
Precision	91.66	80.71
Recall	90.04	77.65
F1-score	90.84	79.15

Table 3. Confusion matrix of the testing set.

Actual	Predicted	
	Inactive	Active
Inactive	606	98
Active	118	410

3.2. QSAR Model

We trained the LightGBM model using the selected molecular descriptors and default hyperparameters. The model's performance was evaluated separately on the training set and the testing set, and the results are presented in Table 2.

It can be seen that the LightGBM model achieved an accuracy of 92.49%, precision of 91.66%, recall of 90.04%, and F1-score of 90.84% for the training set. On the testing set, the LightGBM model achieved an accuracy of 82.47%, precision of 80.71%, recall of 77.65%, and F1-score of 79.15%. Compared to the results from the training set, there is a slight decrease in performance, indicating that the model might be overfitting and encountering some difficulty in generalizing to new, unseen data.

The confusion matrix of the testing set is presented in Table 3. This result shows that there were 606 compounds correctly classified as inactive and 98 instances incorrectly classified as active. Furthermore, there were 118 instances incorrectly classified as inactive and 410 instances correctly classified as active. The LightGBM model correctly classified 77.65% active compounds and 86.08% inactive compounds.

For further analysis, we explored the feature importance of our LightGBM model using the split method, which is presented in Figure 2. This method involves counting the number of times each feature is used to split the data across all the trees in the ensemble. Features with higher counts are considered more important, as they play a more frequent role in the decision-making process of the model. Among the descriptors, AATS4dv emerged as the most important, with a high importance score of 405, which indicates that AATS4dv has a significant impact on the model's output and plays a crucial role in the predictive capabilities of the model. AATS4dv represents the Averaged Moreau-Broto Autocorrelation of lag 4 weighted by valence electrons. It captures specific spatial arrangements and electron distributions within the molecule. On the other hand, the descriptor C4SP3 exhibited the lowest importance score of 15, suggesting that it has relatively less influence on the model's predictions compared to other molecular descriptors.

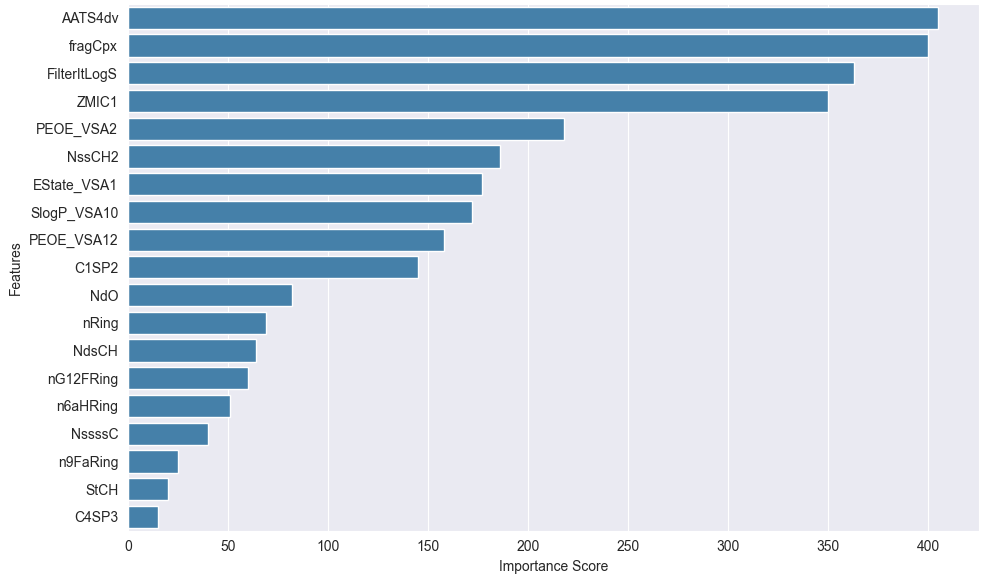

Figure 2. Feature importance of the LightGBM model.

C4SP3 refers to a carbon atom bound to four other carbon atoms with a tetrahedral (SP3) geometry.

3.3. Limitation of This Study

While this study provides valuable insights into the application of GA and LightGBM in the search for AChE inhibitors for Alzheimer's disease, it is important to acknowledge certain limitations. The generalizability of the LightGBM model's performance is dependent on the dataset used, which may not fully represent the diversity of AChE inhibitors. Additionally, the study lacks experimental validation of the predicted inhibitory activity, necessitating further experimental studies for confirmation and assessment of actual potency and efficacy. These limitations emphasize the need for future research to address these challenges and broaden our understanding of effective AChE inhibitors for Alzheimer's disease.

4. Conclusions

This study demonstrates the effectiveness of combining GA and LightGBM in the search for AChE inhibitors for Alzheimer's disease. By using GA for feature selection, the study identifies the most relevant molecular descriptors to predict AChE inhibitory activity. The LightGBM model trained on these selected descriptors achieves high performance. However, there is a possibility of overfitting, as the model's performance on the testing data is slightly lower than that on the training data. For future studies, it is recommended to fine-tune hyperparameters and enhance regularization techniques to achieve a more balanced trade-off between model complexity and generalization. This approach can help mitigate overfitting issues and improve the model's ability to perform well on unseen data. Overall, the findings highlight the potential of these methods for drug discovery efforts for Alzheimer's disease.

Author Contributions: Conceptualization, T.R.N., A.M. and G.M.I.; methodology, T.R.N. and G.M.I.; software, T.R.N., A.M. and R.S.; validation, G.M.I., N.B.M. and R.I.; formal analysis, T.R.N., M.P. and R.I.; resources, G.M.I. and R.I.; data curation, M.P. and R.I.; writing—original draft preparation, T.R.N. A.M., N.B.M. and R.S..; writing—review and editing, G.M.I., M.P. and R.I.; visualization, T.R.N., N.B.M. and R.S.; supervision, G.M.I. and R.I.; project administration, G.M.I.; All authors have read and agreed to the published version of the manuscript.

Funding: This study received no external funding.

Ethical Clearance: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data utilized in this study is accessible through the ChEMBL database.

Acknowledgments: The authors extend their appreciation to their respective institutions and universities

Conflicts of Interest: All the authors declare that there are no conflicts of interest.

References

1. Castellani, R. J., Rolston, R. K., and Smith, M. A. (2010). Alzheimer Disease, *Disease-a-Month*, Vol. 56, No. 9, 484–546. doi:10.1016/j.disamonth.2010.06.001.
2. Braak, H., and Braak, E. (1991). Neuropathological stageing of Alzheimer-related changes, *Acta Neuropathologica*, Vol. 82, No. 4, 239–259. doi:10.1007/BF00308809.
3. Knopman, D. S., Amieva, H., Petersen, R. C., Chételat, G., Holtzman, D. M., Hyman, B. T., Nixon, R. A., and Jones, D. T. (2021). Alzheimer disease, *Nature Reviews Disease Primers*, Vol. 7, No. 1, 33. doi:10.1038/s41572-021-00269-y.

4. Yiannopoulou, K. G., and Papageorgiou, S. G. (2020). Current and Future Treatments in Alzheimer Disease: An Update, *Journal of Central Nervous System Disease*, Vol. 12, 117957352090739. doi:10.1177/1179573520907397.

5. Talesa, V. N. (2001). Acetylcholinesterase in Alzheimer's disease, *Mechanisms of Ageing and Development*, Vol. 122, No. 16, 1961–1969. doi:10.1016/S0047-6374(01)00309-8.

6. Peitzika, S.-C., and Pontiki, E. (2023). A Review on Recent Approaches on Molecular Docking Studies of Novel Compounds Targeting Acetylcholinesterase in Alzheimer Disease, *Molecules*, Vol. 28, No. 3, 1084. doi:10.3390/molecules28031084.

7. Dai, R., Sun, Y., Su, R., and Gao, H. (2022). Anti-Alzheimer's disease potential of traditional chinese medicinal herbs as inhibitors of BACE1 and AChE enzymes, *Biomedicine & Pharmacotherapy*, Vol. 154, 113576. doi:10.1016/j.biopha.2022.113576.

8. Huang, T., Sun, G., Zhao, L., Zhang, N., Zhong, R., and Peng, Y. (2021). Quantitative Structure-Activity Relationship (QSAR) Studies on the Toxic Effects of Nitroaromatic Compounds (NACs): A Systematic Review, *International Journal of Molecular Sciences*, Vol. 22, No. 16, 8557. doi:10.3390/ijms22168557.

9. Noviandy, T. R., Maulana, A., Emran, T. B., Idroes, G. M., and Idroes, R. (2023). QSAR Classification of Beta-Secretase 1 Inhibitor Activity in Alzheimer's Disease Using Ensemble Machine Learning Algorithms, *Heca Journal of Applied Sciences*, Vol. 1, No. 1, 1–7. doi:10.60084/hjas.v1i1.12.

10. Pingaew, R., Prachayasittikul, V., Worachartcheewan, A., Thongnum, A., Prachayasittikul, S., Ruchirawat, S., and Prachayasittikul, V. (2022). Anticancer activity and QSAR study of sulfur-containing thiourea and sulfonamide derivatives, *Heliyon*, Vol. 8, No. 8, e10067. doi:10.1016/j.heliyon.2022.e10067.

11. Maulydia, N. B., Khairan, K., and Noviandy, T. R. (2023). Prediction of Pharmacokinetic Parameters from Ethanolic Extract Mane Leaves (Vitex pinnata L.) in Geothermal Manifestation of Seulawah Agam Ie-Seu'um, Aceh, *Malacca Pharmaceutics*, Vol. 1, No. 1, 16–21. doi:10.60084/mp.v1i1.33.

12. Idroes, G. M., Tallei, T. E., Idroes, R., Muslem, Riza, M., and Suhendrayatna. (2021). The study of Calotropis Gigantea leaf metabolites from Ie Brouk geothermal area Lamteuba-Aceh Besar using molecular docking, *IOP Conference Series: Earth and Environmental Science*, Vol. 667, No. 1, 012072. doi:10.1088/1755-1315/667/1/012072.

13. Ponzoni, I., Sebastián-Pérez, V., Martínez, M. J., Roca, C., De la Cruz Pérez, C., Cravero, F., Vazquez, G. E., Páez, J. A., Díaz, M. F., and Campillo, N. E. (2019). QSAR Classification Models for Predicting the Activity of Inhibitors of Beta-Secretase (BACE1) Associated with Alzheimer's Disease, *Scientific Reports*, Vol. 9, No. 1, 9102. doi:10.1038/s41598-019-45522-3.

14. Xue, L., and Bajorath, J. (2000). Molecular Descriptors in Chemoinformatics, Computational Combinatorial Chemistry, and Virtual Screening, *Combinatorial Chemistry & High Throughput Screening*, Vol. 3, No. 5, 363–372. doi:10.2174/1386207003331454.

15. Remeseiro, B., and Bolon-Canedo, V. (2019). A review of feature selection methods in medical applications, *Computers in Biology and Medicine*, Vol. 112, 103375.

16. Liman, W., Oubahmane, M., Hdoufane, I., Bjij, I., Villemin, D., Daoud, R., Cherqaoui, D., and El Allali, A. (2022). Monte Carlo Method and GA-MLR-Based QSAR Modeling of NS5A Inhibitors against the Hepatitis C Virus, *Molecules*, Vol. 27, No. 9, 2729. doi:10.3390/molecules27092729.

17. Jawarkar, R. D., Bakal, R. L., Zaki, M. E. A., Al-Hussain, S., Ghosh, A., Gandhi, A., Mukerjee, N., Samad, A., Masand, V. H., and Lewaa, I. (2022). QSAR based virtual screening derived identification of a novel hit as a SARS CoV-229E 3CLpro Inhibitor: GA-MLR QSAR modeling supported by molecular Docking, molecular dynamics simulation and MMGBSA calculation approaches, *Arabian Journal of Chemistry*, Vol. 15, No. 1, 103499. doi:10.1016/j.arabjc.2021.103499.

18. Simeon, S., Anuwongcharoen, N., Shoombuatong, W., Malik, A. A., Prachayasittikul, V., Wikberg, J. E. S., and Nantasenamat, C. (2016). Probing the origins of human acetylcholinesterase inhibition via QSAR modeling and molecular docking, *PeerJ*, Vol. 4, e2322. doi:10.7717/peerj.2322.

19. Hammoudi, N.-E.-H., Sobhi, W., Attoui, A., Lemaoui, T., Erto, A., and Benguerba, Y. (2021). In silico drug discovery of Acetylcholinesterase and Butyrylcholinesterase enzymes inhibitors based on Quantitative Structure-Activity Relationship (QSAR) and drug-likeness evaluation, *Journal of Molecular Structure*, Vol. 1229, 129845. doi:10.1016/j.molstruc.2020.129845.

20. El Khatabi, K., El-Mernissi, R., Aanouz, I., Ajana, M. A., Lakhlifi, T., Khan, A., Wei, D.-Q., and Bouachrine, M. (2021). Identification of novel acetylcholinesterase inhibitors through 3D-QSAR, molecular docking, and molecular dynamics simulation targeting Alzheimer's disease, *Journal of Molecular Modeling*, Vol. 27, No. 10, 302. doi:10.1007/s00894-021-04928-5.

21. López, A. F. F., Martínez, O. M. M., and Hernández, H. F. C. (2021). Evaluation of Amaryllidaceae alkaloids as inhibitors of human acetylcholinesterase by QSAR analysis and molecular docking, *Journal of Molecular Structure*, Vol. 1225, 129142. doi:10.1016/j.molstruc.2020.129142.

22. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree, *Advances in Neural Information Processing Systems*, Vol. 30.

23. Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B., and Overington, J. P. (2012). ChEMBL: a large-scale bioactivity database for drug discovery, *Nucleic Acids Research*, Vol. 40, No. D1, D1100–D1107. doi:10.1093/nar/gkr777.

24. Yu, T., Nantasenamat, C., Kachenton, S., Anuwongcharoen, N., and Piacham, T. (2023). Cheminformatic Analysis and Machine Learning Modeling to Investigate Androgen Receptor Antagonists to Combat Prostate Cancer, *ACS Omega*, Vol. 8, No. 7, 6729–6742. doi:10.1021/acsomega.2c07346.

25. Khaire, U. M., and Dhanalakshmi, R. (2022). Stability of feature selection algorithm: A review, *Journal of King Saud University - Computer and Information Sciences*, Vol. 34, No. 4, 1060–1073. doi:10.1016/j.jksuci.2019.06.012.

26. Islam, M. L., Shatabda, S., Rashid, M. A., Khan, M. G. M., and Rahman, M. S. (2019). Protein structure prediction from inaccurate and sparse NMR data using an enhanced genetic algorithm, *Computational Biology and Chemistry*, Vol. 79, No. September 2018, 6–15. doi:10.1016/j.compbiolchem.2019.01.004.

27. Ramaswamy, R., Kandhasamy, P., and Palaniswamy, S. (2023). Feature Selection for Alzheimer's Gene Expression Data Using Modified Binary Particle Swarm Optimization, *IETE Journal of Research*, Vol. 69, No. 1, 9–20. doi:10.1080/03772063.2021.1962747.

28. Idroes, R., Maulana, A., Noviandy, T. R., Suhendra, R., Sasmita, N. R., Lala, A., and Irvanizam. (2020). A Genetic Algorithm to Determine Research Consultation Schedules in Campus Environment, *IOP Conference Series: Materials Science and Engineering*, Vol. 796, 012033. doi:10.1088/1757-899X/796/1/012033.

29. Paplomatas, P., Krokidis, M. G., Vlamos, P., and Vrahatis, A. G. (2023). An Ensemble Feature Selection Approach for Analysis and Modeling of Transcriptome Data in Alzheimer's Disease, *Applied Sciences*, Vol. 13, No. 4, 2353. doi:10.3390/app13042353.

30. Idroes, R., Noviandy, T. R., Maulana, A., Suhendra, R., Sasmita, N. R., Muslem, M., Idroes, G. M., Kemala, P., and Irvanizam, I. (2021). Application of Genetic Algorithm-Multiple Linear Regression and Artificial Neural Network Determinations for Prediction of Kovats Retention Index, *International Review on Modelling and Simulations (IREMOS)*, Vol. 14, No. 2, 137. doi:10.15866/iremos.v14i2.20460.

31. Noviandy, T. R., Maulana, A., Sasmita, N. R., Suhendra, R., Irvanizam, I., Muslem, M., Idroes, G. M., Yusuf, M., Sofyan, H., Abidin, T. F., and Idroes, R. (2022). The Prediction of Kovats Retention Indices of Essential Oils at Gas Chromatography Using Genetic Algorithm-Multiple Linear Regression and Support Vector Regression, *Journal of Engineering Science and Technology*, Vol. 17, No. 1, 306–326.

32. Yang, H., Chen, Z., Yang, H., and Tian, M. (2023). Predicting Coronary Heart Disease Using an Improved LightGBM Model: Performance Analysis and Comparison, *IEEE Access*, Vol. 11, 23366–23380. doi:10.1109/ACCESS.2023.3253885.

33. Sinha, B. B., Ahsan, M., and Dhanalakshmi, R. (2023). LightGBM empowered by whale optimization for thyroid disease detection, *International Journal of Information Technology*, Vol. 15, No. 4, 2053–2062. doi:10.1007/s41870-023-01261-3.

34. Rufo, D. D., Debelee, T. G., Ibenthal, A., and Negera, W. G. (2021). Diagnosis of Diabetes Mellitus Using Gradient Boosting Machine (LightGBM), *Diagnostics*, Vol. 11, No. 9, 1714. doi:10.3390/diagnostics11091714.

35. Wen, X., Xie, Y., Wu, L., and Jiang, L. (2021). Quantifying and comparing the effects of key risk factors on various types of roadway segment crashes with LightGBM and SHAP, *Accident Analysis & Prevention*, Vol. 159, 106261. doi:10.1016/j.aap.2021.106261.

36. Idroes, G. M., Maulana, A., Suhendra, R., Lala, A., Karma, T., Kusumo, F., Hewindati, Y. T., and Noviandy, T. R. (2023). TeutongNet: A Fine-Tuned Deep Learning Model for Improved Forest Fire Detection, *Leuser Journal of Environmental Studies*, Vol. 1, No. 1, 1–8.