European Journal of Technique

journal homepage: https://dergipark.org.tr/en/pub/ejt

Vol. 14, No.1, 2024

Research Article

Performance Comparison of Machine Learning Methods in Discovery of BACE-1 Inhibitors in Alzheimer's Disease Therapy

Bihter Das[1], **Suat Toraman**[2*]

[1]Firat University, Technology Faculty, Software Engineering Department, 23119, Elazig, Turkey. (e-mail: bihterdas@firat.edu.tr).
[2*] Firat University, School of Civil Aviation Department, 23119, Elazig, Turkey. (e-mail: storaman@firat.edu.tr). +904242370000-8978

ARTICLE INFO

Received: March, 20, 2024
Revised: May., 06. 2024
Accepted: May, 08. 2024

Keywords:
Drug Interaction, Alzheimer Disease, BACE-1 inhibitors, Machine Learning, Gradient Boosting Machine (GBM)

Corresponding author: *Suat Toraman*

ISSN: *2536-5010* | e-ISSN: *2536-5134*

DOI: https://doi.org/10.36222/ejt.1455786

ABSTRACT

Alzheimer's disease poses a significant challenge in the realm of neurodegenerative disorders, necessitating effective therapeutic interventions. One promising approach involves the discovery of β-secretase 1 (BACE-1) inhibitors, pivotal in mitigating amyloid-β peptide accumulation, a hallmark of AD pathology. In this study, we compare the performance of three prominent machine learning methods, namely Gradient Boosting Machine, Random Forest, and Support Vector Machine in the discovery of BACE-1 inhibitors. Leveraging the BACE dataset sourced from MoleculeNet, comprising quantitative and qualitative binding results of compounds, we explored the classification efficacy of these methods. Our experimental results reveal distinct precision, recall, and accuracy metrics for each method, showcasing RF with precision and accuracy scores of 1.00 and 99.67%, respectively, followed by GBM and SVM. Furthermore, feature importance analysis underscores pIC50 as the most influential attribute across all methods, emphasizing its pivotal role in classifying BACE-1 inhibitors. Additionally, RF prioritizes Estate as the second most important feature, while AlogP emerges as GBM's secondary significant attribute. These findings shed light on the efficacy of machine learning techniques in identifying potential therapeutics for AD, offering insights into feature importance variations among methods and highlighting the significance of diverse molecular descriptors in drug discovery.

1. INTRODUCTION

Alzheimer's disease (AD) is a prevalent chronic neurodegeneration-related condition affecting the elderly, characterized by the gradual decline of cognitive abilities [1]. The disease is characterized by symptoms such as memory loss, cognitive impairments, and personality changes, and seriously affects patients' quality of life. Alzheimer's disease, which affects millions of people worldwide, has been one of the priority areas of research for the medical and scientific community [2,3]. Abnormal accumulation of beta-amyloid peptide is a key component of the pathophysiology of AD. One of the enzymes enabling this peptide's formation is a protease known as human β-secretase 1 (BACE-1) [4]. BACE-1 performs one of the cuts on the amyloid precursor protein (APP), which is the starting material for the beta-amyloid peptide. Therefore, BACE-1 inhibitors have garnered considerable interest as a promising therapeutic target for addressing AD [5-8]. There are various studies in the literature regarding the use of BACE1 inhibitors for the Alzheimer's disease. Wang et al. studied how β-saron affects the proteins expression associated with AD. Findings indicate that β-Asarone can reduce Aβ accumulation and improve the autophagy process by regulating the protein expression of AD. This approach can be considered a potential strategy for treating AD [9]. Chang et al. devised a sensor to detect β-Secretase (BACE1), an enzyme pivotal in generating β-amyloid, a key factor in the onset of AD. The sensor has a design containing a ferrocene probe and an aldehyde group, thus producing an electrochemical signal by the N terminus released when the APP peptide is cleaved by BACE-1. This developed sensor can detect BACE1 activity by providing a sensitive, low-cost, and easy detection method [10]. Kalaimathi et al. revealed that four marine-derived compounds derived from cyanobacteria could be considered potential BACE1 inhibitors in treating Alzheimer's disease.

These findings may open a promising avenue for the development of new therapeutics [11]. Bhanukiran et. al. performed various physicochemical, reactivity, and stability assays to evaluate the vaccine derivative VA10, a potential BACE1 inhibitor that could be used to treat Alzheimer's disease. Their studies revealed that VA10 was effective as a BACE1 inhibitor and showed no toxicity in mice. These

findings support the evaluation of VA10 as a potential agent in the treatment of AD [12]. Nakano et al. discovered that smoking increases the risk of AD and that nicotinic acetylcholine receptor (nAChR) activation in neurons increases the production of amyloid β (Aβ) in their study. They found that nAChR activation triggers Aβ production by increasing the transcription of the BACE1 gene and increases Aβ accumulation [13]. Abraham et. al. used a multiscale modeling approach to identify minimal congenital characters β-secretase (BACE1) and α-secretase (ADAM10) transmembrane regions, which participate significantly in the amyloid cascade of AD. Our results showed that membrane composition affects the character of the transmembrane domains of BACE1 and ADAM10, providing support for speculations about the role of membrane domains in the etiology of Alzheimer's disease [14]. Vincent et al. investigated the potential link between AD and schizophrenia, highlighting the shared frontotemporal anomaly and increased risk of co-morbid dementia and psychosis. They focused on the molecular level, specifically, the metabolism of β-amyloid precursor protein and neuregulin 1 by β-site APP cleaving enzyme 1, proposing a model to explain the accompanying symptoms [15]. Dominko et al. investigated the proteolysis and distribution of BACE1 substrates, sex6 and Sex6L, in Alzheimer's disease (AD) using the 5xFAD mouse model. they found that while bace1 accumulation in ad brains did not affect the proteolysis of sez6 and sez6l, their distribution was altered in the peri-plaque regions. this suggests different localization and/or function of these substrates compared to app and bace1. The study highlights the potential role of aß-targeted therapies in mitigating the accumulation and modified distribution of BACE-1 and its substrates, along with APP in AD [16].

This study will examine how BACE-1 inhibitors can be evaluated and designed as potential drug candidates targeting Alzheimer's disease. First, we will describe the biochemical and structural properties of BACE-1 and its role in AD pathogenesis. Next, we discuss the various approaches used in the design and identification of BACE-1 inhibitor candidates. Finally, we will summarize current research findings and focus on future studies, highlighting the promise of BACE-1 inhibitors for AD therapy.

This paper aims to contribute to the development of innovative and effective therapeutic strategies for the treatment of Alzheimer's disease. Deepening current knowledge of the role of BACE-1 inhibitors in drug development and promoting progress in this field could be an important step in the fight against AD.

The primary contributions of this paper are as follows:

- Evaluation of BACE-1 inhibitors for identifying potential Alzheimer's disease treatment drug candidates.
- Discussion of various approaches used in the design and identification of BACE-1.
- Analysis of various molecular properties used in the classification of BACE-1 inhibitors and determination of which property has the most significant impact on classification.
- Comparison of the results of classification using machine learning methods such as Gradient Boosting Machine (GBM), Random Forest (RF), and Support Vector Machine (SVM).
- The aim of the study to contribute to the development of innovative therapeutic strategies for Alzheimer's disease treatment and to deepen the understanding of the role of BACE-1 inhibitors in the drug development process.

The remainder of the paper is structured as follows: Section 2 elaborates on the Materials and Methods employed in the experiment. Section 3 delineates the experimental results and engages in discussion. Finally, section 4 provides concluding remarks in the form of the Conclusion.

2. MATERIALS AND METHODS

This paper conducts a study on the evaluation and classification of potential drug candidates for the in the treatment of AD. The performances of methods such as GBM, RF, and SVM are discussed for the design of BACE-1 inhibitors that can be used in the treatment of this disease. Ultimately, the study summarizes current research findings, aiming to underscore the potential of BACE-1 inhibitors in treatment of AD. The steps of the experimental study to achieve this goal will be explained in this section.

2.1. Dataset

This study utilizes a dataset containing quantitative (IC50) and qualitative (binary label) binding results of inhibitors aimed at human β-secretase 1 (BACE-1).In addition to molecular properties, the data set also includes pIC50 values, which indicate the binding strength of inhibitors. The BACE dataset was retrieved from the MolecularNet database and contains 1513 compounds [17]. The dataset was split 80/10/10 to create training, validation, and testing subsets. This division is made to ensure that the model is evaluated correctly during the training, validation and testing phases. Table 1 illustrates a sample of the dataset utilized in the study.

2.2. Machine Learning Methods for Classification

This section will introduce the machine learning (ML) methods used to solve the classification problem on the BACE dataset. The data set contains the properties of each compound, along with a label indicating whether each compound is a BACE-1 inhibitor. This is considered a classification problem and various ML algorithms can be used to solve this problem. In this section, we will explain the basic principles and applications of popular classification algorithms such as SVM, RF and GBM. We will focus on how

2.2. Machine Learning Methods for Classification

This section will introduce the machine learning (ML) methods used to solve the classification problem on the BACE dataset. The data set contains the properties of each compound, along with a label indicating whether each compound is a BACE-1 inhibitor. This is considered a classification problem and various ML algorithms can be used to solve this problem. In this section, we will explain the basic principles and applications of popular classification algorithms such as SVM, RF and GBM. We will focus on how each algorithm works, what types of data it performs best with, and the metrics used to evaluate classification results.

TABLE I
A SAMPLE OF THE USED DATASET

Mol	CID	Class	pIC50	MW	AlogP	HBA	HBD	RB	Heavy Atom Count	Chiral Center Count	Chiral Center CountAll Possible	Ring Count	PSA	Estate	MR	Polar
O1CC[C@@H](NC(=O)[C@@H](Cc2cc3cc(ccc3nc2N)-c2ccccc2C)C)CC1(C)C	BACE_1	1	9.1549015	431.56979	4.4014001	3	2	5	32	2	2	4	77.239998	67.251999	129.9039	58.397999
Fc1cc(cc(F)c1)C[C@H](NC(=O)[C@@H](N1CC[C@](NC(=O)C)(CC(C)C)C1=O)CCc1ccccc1)[C@H](O)[C@@H]1[NH2+]C[C@H](OCCC)C1	BACE_2	1	8.8538723	657.81073	2.6412001	5	4	16	47	6	6	4	124.58	115.417	173.6176	76.254997
S1(=O)(=O)N(c2cc(cc3c2n(cc3CC)CC1)C(=O)N[C@H]([C@H](O)C[NH2+]Cc1cc(OC)ccc1)Cc1ccccc1)C	BACE_3	1	8.6989698	591.74091	2.5499001	4	3	11	42	2	3	5	125.86	96.585999	160.12421	75.639
O=C(NCC1CCCCC1)[C@@H](Cc1cc2cc(ccc2nc1N)-c1ccccc1C)CCC	BACE_4	1	7.9586072	443.6236	7.0788999	2	2	8	33	1	1	4	68.010002	66.001999	137.3194	61.431
S1(=O)(=O)C[C@@H](Cc2cc3c([nH]cc3CC(F)F)cc2)[C@H](O)[C@@H]([NH2+]Cc2cc(ccc2)C(C)(C)C)C1	BACE_5	1	7.2596374	505.6402	2.7595999	2	3	8	35	3	4	4	95.150002	87.500999	132.3071	60.943001
O1[C@@H]2COCC[C@@]2(N=C1N)c1cc(ccc1)-c1cncnc1	BACE_6	0	4.0030508	296.32379	0.87709999	5	0	2	22	2	2	4	82.620003	47.750999	80.689301	37.137001
[NH+]=1[C@](N=C(c2ccccc2)C=1N)(C)c1cc(ccc1)-c1cccnc1	BACE_7	0	4	327.4024	1.4339	2	0	3	25	1	1	4	65.239998	49.251999	97.589897	48.525002
n1cccc(NCc2cc(ccc2)-c2cccnc2)c1N	BACE_8	0	4	276.33569	2.2346001	2	2	4	21	0	0	3	63.830002	44.334999	86.700699	41.577
O1CCC(OC(=O)[C@@H]2[NH2+]C[C@]3(C2)c2c(NC3=O)cccc2)CC1	BACE_9	0	3.9430952	317.35959	-1.0807	4	2	3	23	2	2	4	81.239998	53.084	80.653801	36.096001
O=C1NC(=NC(=C1)CCc1cc2[nH]ccc2cc1)N	BACE_10	0	3.8860567	254.2872	1.7381001	2	3	3	19	0	0	3	87.559998	44.001999	75.263802	34.915001

2.2. Machine Learning Methods for Classification

This section will introduce the machine learning (ML) methods used to solve the classification problem on the BACE dataset. The data set contains the properties of each compound, along with a label indicating whether each compound is a BACE-1 inhibitor. This is considered a classification problem and various ML algorithms can be used to solve this problem. In this section, we will explain the basic principles and applications of popular classification algorithms such as SVM, RF and GBM. We will focus on how each algorithm works, what types of data it performs best with, and the metrics used to evaluate classification results. We will also cover important

topics such as the advantages and disadvantages of each algorithm, their application areas, and parameter tuning. Finally, we will compare the performance of these algorithms on the BACE-1 inhibitor classification problem and analyze the results. This will summarize the paper's main findings and evaluate the effectiveness of ML methods for BACE-1 inhibitor detection.

2.2.1 Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a classification and regression machine learning algorithm utilized for solving classification and regression tasks [18]. SVM creates a decision boundary to classify data points, and this provides the best separation of data points (Figure 1). The main purpose of SVM is to accurately classify new observations by creating a decision boundary between different classes. SVM can solve linear and nonlinear classification problems and generally performs well with high-dimensional datasets [19]. SVM has various important applications in drug discovery and development, such as biological activity prediction, drug mechanism of action analysis, drug toxicity prediction, and drug resistance analysis. Models based on this algorithm can be used for important decisions such as classification and prioritization of potential drug candidates [20]. Additionally, it has been shown that SVM can be used effectively in critical areas such as determining the mechanisms of action of drugs on target proteins and evaluating the safety profiles of drugs. Finally, SVM can also be used as an important tool for predicting the treatment response of medically resistant diseases, which can help develop personalized medicine approaches and optimize treatment strategies [21].

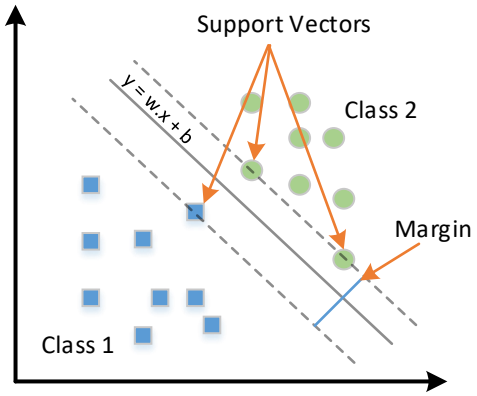

Figure 1. Support Vector Machine

2.2.2 Random Forest (RF)

RF: Random Forest (RF) is a popular ensemble learning technique employed for classification and regression tasks by employing a group of decision trees, known as a forest of trees [22] (Figure 2). Each decision tree splits the dataset by building trees on randomly selected features. Then, each tree makes a prediction and the results are combined to produce the final prediction. RF combats overfitting, works effectively with high-dimensional datasets and generally provides high accuracy [23]. Random Forest (RF) is a potent machine-learning technique extensively employed in drug discovery and development. RF has a variety of important applications, from predicting the biological activity of a given compound to drug discovery and design, drug mechanism of action analysis, and drug toxicity prediction. Therefore, RF is a valuable tool for researchers in the pharmaceutical industry and plays an important role in discovering and evaluating new drug candidates [24,25].

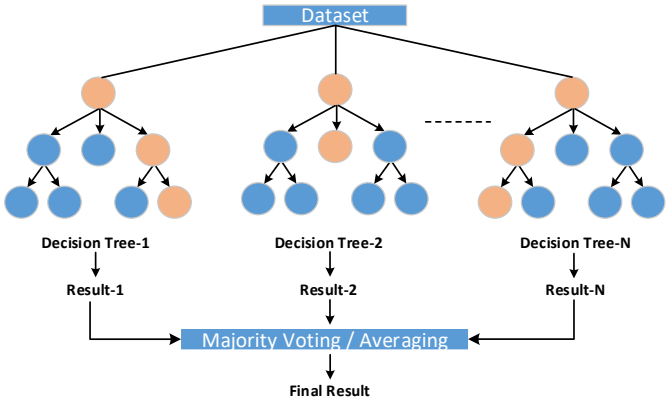

Figure 2. Random Forest

2.2.3 Gradient Boosting Machine (GBM)

Gradient Boosting Machine (GBM) is an ensemble learning technique that creates a strong predictor by combining many weak predictors (commonly known as decision trees) (Figure 3). GBM successively adds weak predictors, and each predictor focuses on correcting the errors of previous predictors. In this way, GBM can adapt to the complexity of the dataset and provide high accuracy. GBM also reduces overfitting and generally produces competitive performance [26,27]. GBM is less known or less used in drug studies than some other algorithms.

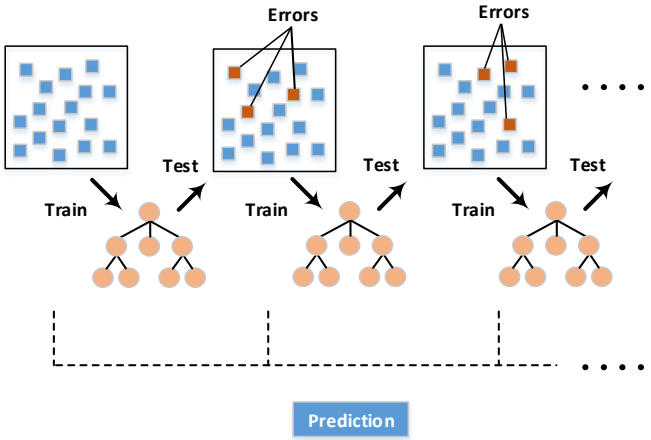

Figure 3. Gradient Boosting Machine

3. RESULTS AND DISCUSSION

This section presents the experimental results of three different machine learning methods used for the classification of the BACE-1 drug dataset and a discussion of these results. Classification results obtained using RF, SVM, and GBM methods were examined. The performance of each method was evaluated in terms of precision, sensitivity (Recall), and accuracy (Accuracy) criteria. Additionally, an in-depth discussion of possible factors that may affect the success of these methods is presented. This discussion will help us better understand the results obtained and their impact on drug discovery and development processes. In this study, a significant analysis was also performed to determine the most important features in classifying BACE-1 drug data. This analysis focused on identifying the features that contribute most to the classification performance of each machine learning method. The results obtained evaluate the contribution

of each feature to the classification process and reveal which features are more decisive in predicting drug effectiveness.

TABLE II.
HYPERPARAMETERS FOR ALL THREE METHODS

RF	"max_depth": [5, 8, 10], "max_features" : [2, 5, 10], "n_estimators" : [200, 500, 1000], "min_samples_split" : [2, 10, 80]
SVM	"C": [0.1, 10, 100, 1000], "kernel": ["linear", "rbf", "polynomial"]
GBM	"max_depth" : [3, 5, 8], "n_estimators" : [50, 100, 200], "subsample" : [.5, 1, 3]

Identifying these important properties can help identify priority targets and identify more effective drug candidates in drug discovery and design processes. Table 2 gives the parameters used and the ranges examined for all three models. Table 3 presents the performance results of all three methods in classifying BACE-1 data. As seen in Table 3, the RF method shows a very effective performance with high precision and high recall for classification. The SVM method also performs quite well, with high precision and recall, but slightly lower compared to the RF method. The GBM method shows a very successful performance with high sensitivity, recall, and accuracy for classification, and achieved slightly higher accuracy than the other two methods. These results show that machine learning methods such as RF, SVM, and GBM are effective in BACE drug data classification. Figure 4 shows the results of the confusion matrix for three methods.

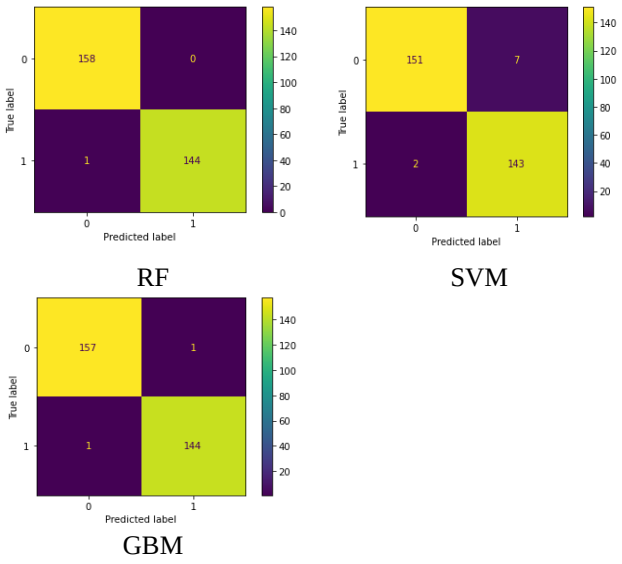

Figure 4. Results of the confusion matrix for three methods

TABLE III.
THE PERFORMANCE RESULTS FOR THREE METHODS

	Precision (%)	Recall (%)	Accuracy (%)
RF	100.00	99.31	99.67
SVM	95.33	98.62	97.03
GBM	99.31	99.31	99.34

Figures 5 and 6 give the most important chemical properties in classification for GBM and RF, respectively. These results show that it is meaningful to determine the PlC50 feature as the most important feature in classifying BACE drug data. Since PlC50 is a metric that measures the biological activity of a compound, this property is thought to directly reflect drug efficacy. Therefore, PlC50 emerged as the most important feature, indicating that it is a determining factor in the classification of drug candidates. The second important feature for RF was identified as Estate, indicating that this feature may play a decisive role in evaluating drug effectiveness. Estate is a property that represents the electronic properties of a compound, and this property can have an impact on the biological activity of the compound. For GBM, the second important feature is determined as AlogP, a feature that reflects the hydrophobicity and lipophilicity of the compound. The emergence of AlogP as a significant feature indicates that it may influence critical factors such as the cellular penetration of the compound and its ability to bind to the target protein. These results suggest that different machine learning methods may emphasize different features and that different mechanisms may be important in determining drug efficacy. Weights assigned to features can only be used in the case of linear kernel in SVM. This cannot be used in other kernel functions (rbf, polynomial). For this reason, feature importance could not be calculated for SVM.

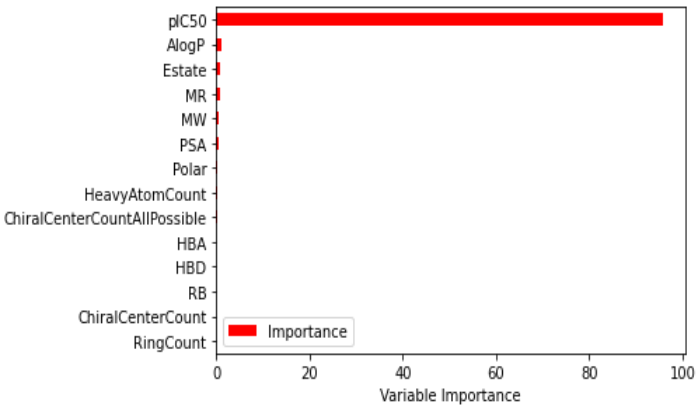

Figure 5. GBM feature importance

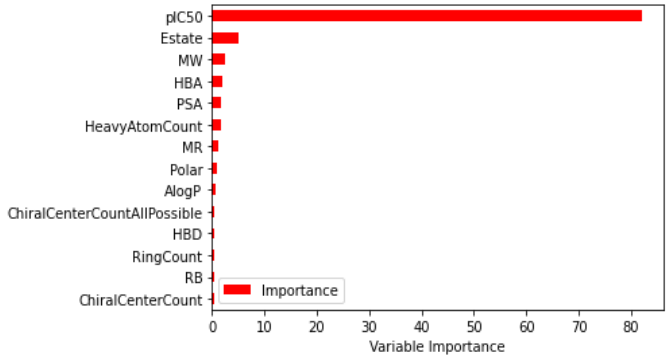

Figure 6. RF feature importance

4. CONCLUSION

This study thoroughly assessed the effectiveness of Gradient Boosting Machine (GBM), Random Forest (RF), and Support Vector Machine (SVM) algorithms in identifying potential BACE-1 inhibitors for the treatment of Alzheimer's disease. Through the analysis of precision, recall, and accuracy metrics, we observed that RF exhibited the highest classification

efficacy among the three methods, followed closely by GBM and SVM. Feature importance analysis revealed pIC50 as the most influential attribute across all methods, underscoring its significance in classifying BACE-1 inhibitors. . Moreover, RF prioritized Estate as the second most important feature, while AlogP emerged as GBM's secondary significant attribute. These findings contribute to our understanding of machine learning methods' utility in drug discovery and emphasize the importance of diverse molecular descriptors in identifying potential therapeutics for Alzheimer's disease. Future studies are planned to explore additional machine learning algorithms and feature selection strategies to increase the effectiveness of drug discovery efforts targeting Alzheimer's disease.

ACKNOWLEDGEMENT

This work is supported by the Turkish Scientific and Technical Research Council (TUBITAK) project named "Development of a New Model for the Discovery of Anti-HIV Effective Molecules Using Geometric Deep Network Approaches" and project code 123E098.

REFERENCES

[1] M. Thambisetty, L. Beason-Held, M. Kraut, R. Desikan, S. Resnick, & Y. “An. The Entorhinal Cortex-Hippocampal System Is An Early Target Of Clusterin-Related Neurodegeneration In Alzheimer’s Disease”. Alzheimer’s & Dementia, 10(4), Supplement, P160. doi: 10.1016/j.jalz.2014.04.145.

[2] E. Ruether, H. Moessler, & M. Windisch. “The MAD-B study — A randomized, double-blind, placebo-controlled trial with cerebrolysin in Alzheimer’s disease”. European Neuropsychopharmacology, 10, 355, 2000, doi: 10.1016/S0924-977X(00)80463-0.

[3] D. Al-Jumeily, S. Iram, F. Vialatte, & P. Fergus. “A novel method to analyze EEG synchrony for the early diagnosis of Alzheimer's disease in optimized frequency bands”. In 2014 IEEE 11th Consumer Communications and Networking Conference (CCNC) (pp. 1-4). Las Vegas, NV, USA. doi: 10.1109/CCNC.2014.6866646.

[4] M. M. Mishra & P. Kumar. “Crocin: A Potent Secondary Metabolite As BACE1 Inhibitor In Alzheimer’s Disease”. 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), s. 1486-1490. Coimbatore, Hindistan. doi: 10.1109/ICACCS57279.2023.10112776.

[5] S. Sharma ve Y. Hasija. “Identification and screening of BACE1 inhibitors using Drug Repurposing: A Computational Approach”. 2023 3rd International Conference on Innovative Sustainable Computational Technologies (CISCT), pp. 1-5. Dehradun, Hindistan. doi: 10.1109/CISCT57197.2023.10351386.

[6] A. F. Nugroho, R. R. Septiawan, ve I. Kurniawan. “Prediction of Human β-secretase 1 (BACE-1) Inhibitors for Alzheimer Therapeutic Agent by Using Fingerprint-based Neural Network Optimized by Bat Algorithm”. 2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE), pp. 257-261. Jakarta, Endonezya. doi: 10.1109/ICCoSITE57641.2023.10127718.

[7] Das, B., Yan, R. A Close Look at BACE1 Inhibitors for Alzheimer’s Disease Treatment. *CNS Drugs* **33**, 251–263 (2019). https://doi.org/10.1007/s40263-019-00613-7

[8] F. H. Bazzari and A. H. Bazzari, “BACE1 Inhibitors for Alzheimer’s Disease: The Past, Present and Any Future?,” *Molecules*, vol. 27, no. 24, Art. no. 24, Jan. 2022, doi: 10.3390/molecules27248823.

[9] Z. Wang et al. “Inhibitory effects of β-asarone on lncRNA BACE1-mediated induction of autophagy in a model of Alzheimer’s disease. Behavioural Brain Research, vol 463, pp. 114896, 2024.

[10] Z. Chang et al. “Signal-on' electrochemical detection of BACE1 for early detection of Alzheimer’s disease”. Cell Reports Physical Science, pp. 101632, 2023.

[11] K. Kalaimathi et al. “Cyanobacterial metabolites as novel inhibitors of BACE1 implicated in Alzheimer’s disease through in silico approaches”. Intelligent Pharmacy, vol. 2, no. 1, pp. 144–149, 2024.

[12] B. Louis, V. K. Agrawal, ve P. V. Khadikar. “Single crystal X-ray, DFT, molecular dynamic simulations, and biological evaluation of 3-OH pyrrolidine derivative VA10 from alkaloid vasicine for BACE1 inhibition”. Journal of Molecular Structure, vol. 1300, pp. 137196, 2024.

[13] M. Nakano, T. Tsuchida, Y. Mitsuishi, ve M. Nishimura. “Nicotinic acetylcholine receptor activation induces BACE1 transcription via the phosphorylation and stabilization of nuclear SP1”. Neuroscience Research, 2023.

[14] C. B. Abraham, L. Xu, G. A. Pantelopulos, ve J. E. Straub. “Characterizing the transmembrane domains of ADAM10 and BACE1 and the impact of membrane composition”. Biophysical Journal, cilt. 122, no. 19, pp. 3999–4010,2023.

[15] B. Vincent ve S. Maitra. “BACE1-dependent metabolism of neuregulin 1: Bridging the gap in explaining the occurrence of schizophrenia-like symptoms in Alzheimer’s disease with psychosis?”, Ageing Research Reviews, vol. 89, pp. 101988. doi: 10.1016/j.arr.2023.101988, 2023.

[16] K. Dominko et. al. “Amyloid-ß plaque formation and BACE1 accumulation in the brains of a 5xFAD Alzheimer’s disease mouse model is associated with altered distribution and not proteolysis of BACE1 substrates Sez6 and Sez6L”. Mechanisms of Ageing and Development, vol. 207, pp. 111726. doi: 10.1016/j.mad.2022.111726, 2022.

[17] Z. Wu et. al. “MoleculeNet: A Benchmark for Molecular Machine Learning”. arXiv preprint, arXiv: 1703.00564, 2017.

[18] M. M. Inuwa and R. Das, “A comparative analysis of various machine learning methods for anomaly detection in cyber attacks on IoT networks,” *Internet of Things*, vol. 26, p. 101162, Jul. 2024, doi: 10.1016/j.iot.2024.101162.

[19] J. Lin et. al. “Accurate prediction of potential druggable proteins based on genetic algorithm and Bagging-SVM ensemble classifier”. Artificial Intelligence in Medicine, vol. 98, pp. 35–47. doi: 10.1016/j.artmed.2019.07.005, 2019

[20] J. I. B. Janairo. “Chapter 6 - Support vector machine in drug design. In Cheminformatics, QSAR and Machine Learning Applications for Novel Drug Development”, K. Roy (Ed.), Academic Press, pp. 161–179. doi: 10.1016/B978-0-443-18638-7.00021-9, 2023.

[21] B. Das, S. Toraman, I. Turkoglu, "A novel genome analysis method with the entropy-based numerical techniqueusing pretrained convolutional neural networks," *Turkish Journal of Electrical Engineering and Computer Sciences*: Vol. 28: No. 4, Article 9, 2020. https://doi.org/10.3906/elk-1909-119

[22] A. F. Amiri, H. Oudira, A. Chouder, ve S. Kichou. “Faults detection and diagnosis of PV systems based on machine learning approach using random forest classifier”. Energy Conversion and Management, vol. 301, pp. 118076, 2024, doi: 10.1016/j.enconman.2024.118076.

[23] A. Alloubani, B. Abuhaija, M. Almatari, G. Jaradat, ve B. Ihnaini. “Predicting vitamin D deficiency using optimized random forest classifier”. Clinical Nutrition ESPEN, vol. 60, pp. 1–10, 2024. doi: 10.1016/j.clnesp.2023.12.146.

[24] G. Cano et. al. “Automatic selection of molecular descriptors using random forest: Application to drug discovery”. Expert Systems with Applications, vol. 72, pp. 151–159, 2017. doi: 10.1016/j.eswa.2016.12.008.

[25] Das B. An implementation of a hybrid method based on machine learning to identify biomarkers in the Covid-19 diagnosis using DNA sequences. Chemometr Intell Lab Syst. 2022 Nov 15;230:104680. doi: 10.1016/j.chemolab.2022.104680.

[26] P. B. Yang, Y. J. Chan, S. K. Yazdi, and J. W. Lim, “Optimisation and economic analysis of industrial-scale anaerobic co-digestion (ACoD) of palm oil mill effluent (POME) and decanter cake (DC) using machine learning models: A comparative study of Gradient Boosting Machines (GBM), K-nearest neighbours (KNN), and random forest (RF),” *Journal of Water Process Engineering*, vol. 58, p. 104752, Feb. 2024, doi: 10.1016/j.jwpe.2023.104752.

[27] P. B. Yang, Y. J. Chan, S. K. Yazdi, and J. W. Lim, “Optimisation and economic analysis of industrial-scale anaerobic co-digestion (ACoD) of palm oil mill effluent (POME) and decanter cake (DC) using machine learning models: A comparative study of Gradient Boosting Machines (GBM), K-nearest neighbours (KNN), and random forest (RF),” *Journal of Water Process Engineering*, vol. 58, p. 104752, Feb. 2024, doi: 10.1016/j.jwpe.2023.104752.

BIOGRAPHIES

Bihter Das graduated B.S. and M.S. degrees from the Department of Computer Science at the Firat University in 2004 and 2007 respectively. Then she received Ph.D. degree at the Department of Software Engineering at the same university in 2018. She also worked between September 2017 and June 2018 as a visiting scholar at the Department of Computing Science at the University of Alberta, Edmonton, Canada. Her current research areas include data science,

big data, data analytics, bioinformatic, digital signal processing, genome data analysis.

Suat Toraman graduated B.S. and M.S. degrees from the Department of Computer Science at the Firat University in 2002 and 2006 respectively. Then she received Ph.D. degree at the Department of Software Engineering at the same university in 2016. Her current research areas include data science, big data, machine learning, deep learning, digital signal processing.