IJAIA

Volume 1 | Number 1 | June 2025 | Pages XXX-XXX

International Journal of Artificial Intelligence Applications

ISSN (online): 0000-0000
Homepage: ijaia.com

Enhancing Virtual Screening for Drug Discovery: Machine Learning Approaches for Acetylcholinesterase Inhibitor Prediction

Amneh Shaban[1], Arar Al Tawil[2], Motaz Zghoul[3]

[1]Faculty of Information Technology, Applied Science Private University, Amman 11931, Jordan
a_shaban@asu.edu.jo
[2]Faculty of Information Technology, Applied Science Private University, Amman 11931, Jordan
ar_altawil@asu.edu.jo
[3]Department of Artificial Intelligence-Faculty of Science and Information Technology, Al-Zaytoonah University of Jordan, Amman, 11733, Jordan) Motazzghoul98@gmail.com

Received: 01, 2025 *Revised: 02, 2025* *Accepted: 03, 2025* *Available online: 06, 2025*

ABSTRACT — Acetylcholinesterase (AChE) is a crucial enzyme in neurotransmission and a key target in drug discovery, particularly for neurodegenerative diseases such as Alzheimer's and Parkinson's. Identifying potent AChE inhibitors requires efficient screening methods due to the high cost and time-consuming nature of experimental approaches. Machine learning (ML) offers a powerful alternative for virtual screening, enabling rapid and accurate bioactivity predictions. In this study, we leveraged the PyCaret framework to develop ML models for predicting AChE inhibitory activity. A dataset of 7,717 compounds and 882 molecular descriptors from the ChEMBL database was used to train classification models, including Light Gradient Boosting Machine (LightGBM), XGBoost, Decision Tree, Extra Trees, and K-Nearest Neighbors (KNN). Performance was evaluated using accuracy, recall, precision, F1-score, and kappa metrics. The results indicate that LightGBM achieved the highest accuracy of 94.30% and an F1-score of 94.29%, followed by XGBoost with 94.02% accuracy and an F1-score of 94.01%, demonstrating their superior predictive capabilities. Additionally, we analyzed computational efficiency, highlighting trade-offs between performance and model complexity. This study establishes ML as a scalable and effective approach for bioactivity prediction, reducing reliance on costly experimental screening. Our findings contribute to AI-driven drug discovery by providing an optimized workflow for identifying potential AChE inhibitors.

Keywords — Machine Learning; Virtual Screening; Acetylcholinesterase Inhibitors; Drug Discovery; Cheminformatics

1. INTRODUCTION

Acetylcholinesterase (AChE) is a key enzyme involved in the hydrolysis of acetylcholine (ACh), a critical neurotransmitter in both the central and peripheral nervous systems. It plays an essential role in synaptic transmission by rapidly breaking down ACh, thereby preventing overstimulation of cholinergic receptors [1]. AChE is a major therapeutic target, with inhibitors widely used in the treatment of neurological disorders such as Alzheimer's disease, Parkinson's disease, and myasthenia gravis [2]. Additionally, AChE inhibitors are used in agriculture as pesticides and insecticides, and some organophosphate-based inhibitors have been weaponized as nerve agents [3].

Despite their therapeutic and industrial significance, AChE inhibitors pose substantial risks to human health and environmental safety. Pesticides such as carbamates and organophosphates have been linked to acute neurotoxicity in nontarget organisms, including mammals, birds, and aquatic species, due to their irreversible inhibition of AChE [4]. Given the high cost and ethical concerns of in vivo testing, computational approaches have been

* Corresponding author
DOI: https://doi.org/

increasingly employed to predict the bioactivity of chemical compounds against AChE, facilitating safer drug and pesticide design [5].

Machine learning (ML) has emerged as a powerful tool for predictive toxicology and drug discovery, offering an efficient alternative to traditional experimental screening methods. ML models, particularly supervised learning approaches, can learn complex relationships between molecular structures and biological activity by training on large scale datasets [18]. Unlike traditional computational techniques such as molecular docking and quantitative structure activity relationship (QSAR) modeling, ML-based methods can leverage large chemical libraries to improve predictive performance [16]. Recent studies have demonstrated the effectiveness of ML algorithms, including gradient boosting machines, deep learning, and support vector machines, in bioactivity prediction tasks [8].

In this study, we utilized the PyCaret framework to develop machine learning models for predicting AChE inhibitory activity [9]. The dataset comprised 7,717 compounds and 882 molecular descriptors, which were used to train and evaluate a range of classification models, including Light Gradient Boosting Machine (LightGBM) [22], XGBoost [11], Decision Tree, Extra Trees, and K- Nearest Neighbors (KNN). These models were assessed using multiple performance metrics, including accuracy, recall, precision, F1-score, and kappa, to determine their effectiveness in bioactivity prediction.

Our findings indicate that tree based ensemble methods, particularly Light- GBM and XGBoost, achieved the highest accuracy and AUC scores, demonstrating their strong predictive capabilities [12, 13]. These models outperform traditional approaches in both efficiency and generalization ability, making them valuable tools for virtual screening in drug discovery and toxicology assessment [14, 15]. Moreover, we analyzed model training time to assess computational efficiency, providing insights into the trade-offs between model complexity and predictive performance [15].

1.1 **Contributions**

This study presents a novel application of machine learning models for predicting the bioactivity of compounds against Acetylcholinesterase (AChE), contributing to both computational drug discovery and toxicology assessment. The key contributions of this work are as follows:

- Developed and evaluated machine learning models, including LightGBM, XGBoost, Decision Trees, Extra Trees, and K-Nearest Neighbors (KNN), using the PyCaret framework for AChE inhibitory activity prediction.
- Processed a dataset of 7,717 compounds with 882 molecular descriptors from the ChEMBL database, applying feature engineering and bioactivity labeling to improve predictive accuracy.
- Demonstrated that tree-based ensemble models outperform traditional approaches in accuracy and generalization, highlighting their efficiency in handling high dimensional cheminformatics data.
- Provided insights into computational tradeoffs between model complexity and predictive performance, contributing to AI-driven virtual screening for drug discovery.

By presenting a comprehensive evaluation of machine learning techniques in AChE inhibitor prediction, this research contributes valuable insights to the fields of cheminformatics

and computational drug discovery, paving the way for more efficient and scalable AI-driven screening methods.

2. RELATED WORK

The application of machine learning (ML) in drug discovery, particularly for predicting bioactivity against Acetylcholinesterase (AChE), has been extensively studied in recent years. Traditional computational approaches, such as molecular docking and quantitative structure activity relationship (QSAR) modeling, have been widely used to assess AChE inhibition [16, 17]. However, these methods rely heavily on predefined molecular features and assumptions, limiting their predictive power. In contrast, ML models can learn complex patterns directly from large datasets, enhancing predictive accuracy and generalizability [18].

Several studies have demonstrated the effectiveness of ML models for bioactivity prediction. Baskin et al. [19] highlighted the potential of deep learning architectures, showing that convolutional neural networks (CNNs) outperform traditional QSAR models in cheminformatics applications. Similarly, Ma et al. [20] developed deep learning-based QSAR models that achieved superior performance in virtual screening compared to conventional statistical methods. Gradient boosting techniques, particularly LightGBM and XGBoost, have gained significant attention due to their efficiency and high predictive performance in drug discovery. Chen and Guestrin [21] introduced XGBoost as a scalable and efficient boosting method, which has since been widely adopted in cheminformatics. Ke et al. [22] further improved gradient boosting with LightGBM, which demonstrated faster training times and better performance on high-dimensional molecular datasets.

Recent research has applied ML models to AChE inhibitor prediction. For example, Zhang et al. [23] utilized ML algorithms, including random forests and gradient boosting, to identify novel AChE inhibitors. Their findings indicate that ensemble-based models outperform traditional ML approaches in terms of accuracy and robustness. Additionally, Ho et al. [24] applied learning to rank techniques with gradient boosting decision trees for virtual screening, demonstrating improved ranking of potential drug candidates.

While tree-based models such as LightGBM and XGBoost have shown promising results, hybrid approaches integrating ML with molecular docking have also emerged. Sheridan et al. [25] proposed combining LightGBM with molecular docking simulations to enhance the reliability of bioactivity predictions. Such hybrid methods provide a more comprehensive screening process, leveraging both data-driven and structure-based techniques.

In summary, the integration of ML models, particularly gradient boosting techniques, has significantly improved bioactivity prediction for AChE inhibitors, demonstrating the potential of AI-driven virtual screening in drug discovery.

3. METHODS

3.1 Dataset Collection

The dataset used in this study was obtained from the **ChEMBL database**, a widely used bioactivity database that provides curated data on chemical compounds, their biological interactions, and pharmacological properties [26]. ChEMBL serves as a valuable resource for

cheminformatics and computational drug discovery, offering high-quality experimental data from published literature and regulatory submissions.

For this study, the dataset was specifically filtered to focus on the inhibition of Acetylcholinesterase (AChE) in Homo sapiens, ensuring that only relevant bioactivity data was included in the model training process.

The following criteria were applied:

- **Target Name**: Acetylcholinesterase (AChE)
- **Target ChEMBL ID**: CHEMBL220
- **Organism**: Homo sapiens
- **Bioactivity Measurement**: Half-maximal inhibitory concentration (IC50)

The retrieved dataset includes molecular structures, bioactivity data, and details of experimental conditions. The IC50 values, which represent the concentration required to inhibit 50% of AChE activity, were used as a quantitative measure of compound potency. IC50 is widely adopted in drug discovery research as a standard metric for evaluating the inhibitory effects of small molecules on enzymatic activity [27].

To ensure data integrity, entries with missing or inconsistent values were removed, and molecular descriptors were computed to provide numerical representations of chemical structures. These molecular descriptors, extracted using cheminformatics tools such as PaDEL Descriptor, were essential for feature engineering in machine learning models [29].

3.2 Data Preprocessing

To ensure data consistency and improve the reliability of machine learning mod- els, several pre-processing steps were applied to the dataset.

3.3.1 Handling Missing Values

Entries with missing or undefined bioactivity values (IC50) were removed from the dataset to ensure that all compounds included in the analysis had valid experimental measurements.

3.3.1 Bioactivity Labeling

The dataset contains bioactivity data in the form of IC50 values, which represents the concentration required to inhibit 50% of AChE activity. To facilitate classification tasks, compounds were categorized into three distinct bioactivity classes:

- **Active:** Compounds with IC50 values ≤ 1000 nM.
- **Intermediate:** Compounds with IC50 values between 1000 nM and 10,000 nM.
- **Inactive:** Compounds with IC50 values > 10,000 nM.

3.3 Molecular Descriptors and Data Exploration

3.3.1 Molecular Descriptors Calculation

To transform chemical structures into machine-readable formats, molecular de- scriptors were calculated using the PaDEL-Descriptor software. PaDEL-Descriptor is an open-source tool that generates a comprehensive set of molecular descriptors, including both 1D, 2D, and

fingerprint-based features [29]. These descriptors encode important chemical and structural properties, which are essential for predicting bioactivity.

In total, 882 molecular descriptors were generated for each compound, resulting in a final dataset containing 7717 compounds and 882 features. The extracted molecular descriptors include:

- **Physicochemical Descriptors:** Molecular weight (MW), LogP, hydrogen bond donors (HBD), hydrogen bond acceptors (HBA), rotatable bonds (NRB).
- **2D Descriptors:** Topological polar surface area (TPSA), number of rings, atom connectivity indices.
- **Fingerprint Features:** MACCS keys, PubChem fingerprints, and Morgan circular fingerprints.

These molecular descriptors serve as predictive features for machine learning models, enabling the identification of patterns between molecular structures and their bioactivity.

3.3.2 Drug-Likeness Filtering Using Lipinski's Rule of Five

Lipinski's Rule of Five (Ro5) was applied to evaluate the drug likeness of the compounds. According to Lipinski's rule, a compound is likely to exhibit good oral bioavailability if it meets the following criteria [30]:

- Molecular weight ≤ 500 Da
- LogP ≤ 5
- Hydrogen bond donors ≤ 5
- Hydrogen bond acceptors ≤ 10

Compounds violating more than one of these rules were filtered out to ensure that the dataset contained molecules with favorable pharmacokinetic properties. The application of Lipinski's rule enhances the relevance of the dataset for drug discovery and reduces the likelihood of selecting compounds with poor absorption or permeability characteristics.

3.3.3 Data Exploration and Bioactivity Analysis

To gain insight into the dataset distribution and bioactivity trends, exploratory data analysis (EDA) was performed. The following figures illustrate key aspects of the dataset:

These visualizations provide valuable insights into the distribution of bioactivity classes, the relationship between pIC50 values and activity, and the influence of physicochemical properties on bioactivity. The trends observed reinforce the importance of molecular descriptors in predicting inhibitory potential and guiding drug discovery efforts.

4. EXPERIMENTAL AND RESULT

The machine learning models were implemented and executed using the Kaggle Notebook environment, which provides cloud-based computing resources for data analysis and model training. The experiments were conducted using a CPU-based runtime, ensuring reproducibility and accessibility for computational tasks. The dataset, consisting of 7717 compounds and 882 molecular descriptors, was preprocessed and analyzed within the Kaggle environment. The PyCaret library was employed to automate model selection, training, and evaluation. PyCaret provides a streamlined approach to machine learning, allowing for efficient comparison of multiple models and hyperparameter tuning. To maintain consistency

across runs, a fixed random seed was set in PyCaret, ensuring that model training and evaluation results were reproducible. The training and testing data were split using an 80:20 ratio, with model performance evaluated using standard classification metrics such as accuracy, precision, recall, and F1- score. Additionally, to ensure robust model evaluation and minimize bias, a 10-fold cross validation approach was applied. This method involves splitting the dataset into ten equal subsets, where the model is trained on nine folds and tested on the remaining fold in each iteration, leading to a more generalized performance assessment. Given the computational limitations of a CPU-based environment, feature selection techniques were applied to reduce the dimensionality of the dataset, improving training efficiency while maintaining predictive performance. The final selected model was deployed within the Kaggle Notebook for validation and further analysis.

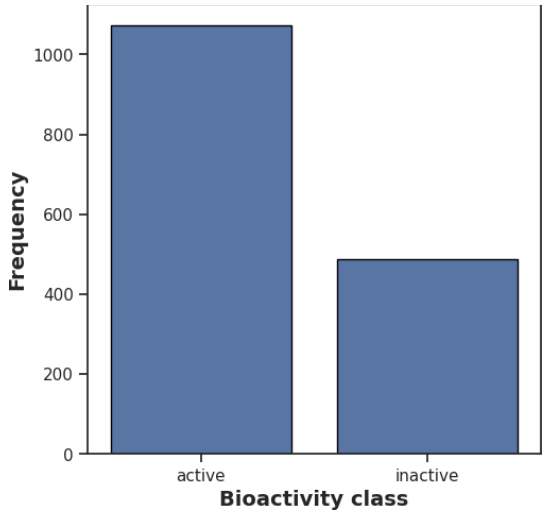

Figure 1: Distribution of bioactivity classes. The dataset contains a higher proportion of active compounds compared to inactive ones.

5. RESULTS

5.1 Machine Learning Model Performance

To predict the bioactivity of compounds against Acetylcholinesterase (AChE), machine learning models were implemented using the PyCaret framework. The dataset, containing 7717 compounds and 882 molecular descriptors, was used to train various classification models.

PyCaret's model comparison function was used to evaluate multiple machine learning algorithms based on classification performance metrics, including accuracy, AUC, recall, precision, F1-score, kappa, and training time.

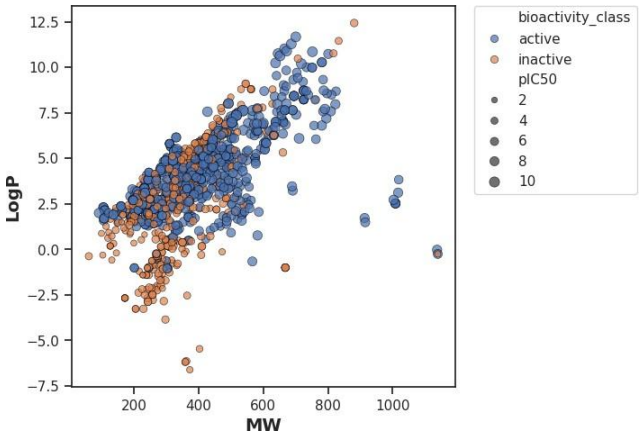

Figure 2: Box plot of pIC50 values for active and inactive compounds. Active compounds generally exhibit higher pIC50 values, indicating greater inhibitory potential.

The top performing models are presented in Table 1.

Table 1: Performance comparison of the top five machine learning models based on accuracy, recall, precision, F1-score, and kappa

Model	Accuracy	Recall	Precision	F1-score	Kappa
LightGBM	0.9430	0.9430	0.9432	0.9429	0.9098
XGBoost	0.9402	0.9402	0.9404	0.9401	0.9054
Decision Tree	0.9083	0.9083	0.9085	0.9083	0.8550
Extra Trees	0.7431	0.7431	0.7410	0.7411	0.5919
K-Nearest Neighbors	0.6885	0.6885	0.6807	0.6781	0.4917

Among the evaluated models, **Light Gradient Boosting Machine (Light- GBM)** achieved the highest performance with an accuracy of 94.30% and an AUC of 0.9784, followed closely by **XGBoost** with an accuracy of 94.02% and an AUC of 0.9766. The Decision Tree model also showed good performance with an accuracy of 90.83% but required significantly less computational time compared to ensemble-based models.

The Extra Trees and K-Nearest Neighbors models demonstrated lower performance in terms of accuracy and AUC, indicating that more complex models such as LightGBM and XGBoost are better suited for bioactivity prediction in this dataset. Training time varied across models, with Decision Trees and KNN being the fastest, while ensemble methods required more computational resources.

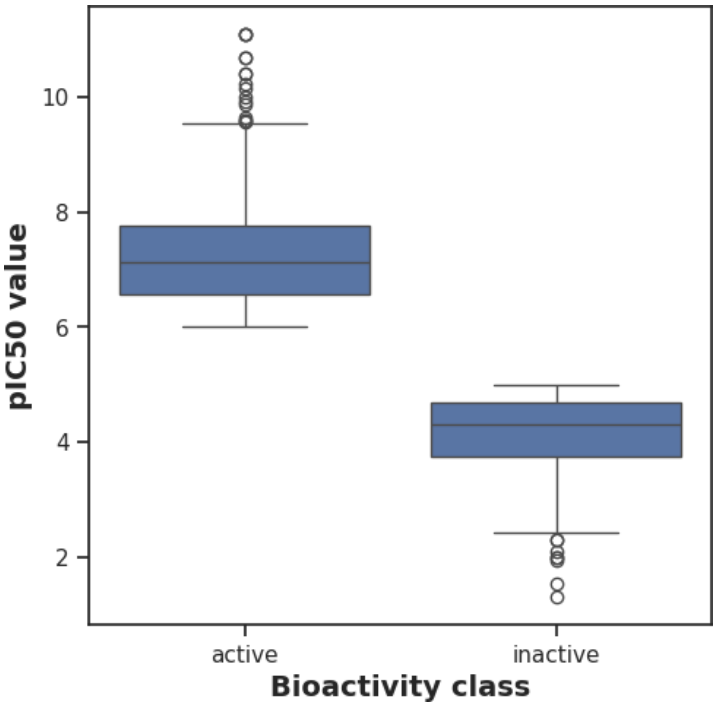

Figure 3: Scatter plot showing the relationship between molecular weight (MW) and LogP for active and inactive compounds. The plot suggests a trend where molecular weight and LogP influence bioactivity.

These results suggest that tree-based ensemble models, particularly LightGBM and XGBoost, are highly effective in predicting AChE inhibitory activity and could serve as reliable tools for virtual screening in drug discovery.

6. DISCUSSION

The results of our machine learning experiments indicate that tree-based ensemble models, particularly LightGBM and XGBoost, exhibit superior predictive performance in classifying compounds based on their bioactivity against Acetyl- cholinesterase (AChE). These models achieved the highest accuracy scores, with LightGBM at 94.30% and XGBoost at 94.02%, demonstrating their robustness in handling high-dimensional molecular descriptor data. The superior performance of LightGBM and XGBoost can be attributed to their ability to efficiently handle large feature spaces and capture complex relationships between molecular structures and bioactivity. These models leverage gradient boosting techniques, which iteratively improve predictions by minimizing errors from previous iterations. Additionally, they are known for their capability to handle imbalanced datasets, a common challenge in bioactivity prediction. In contrast,

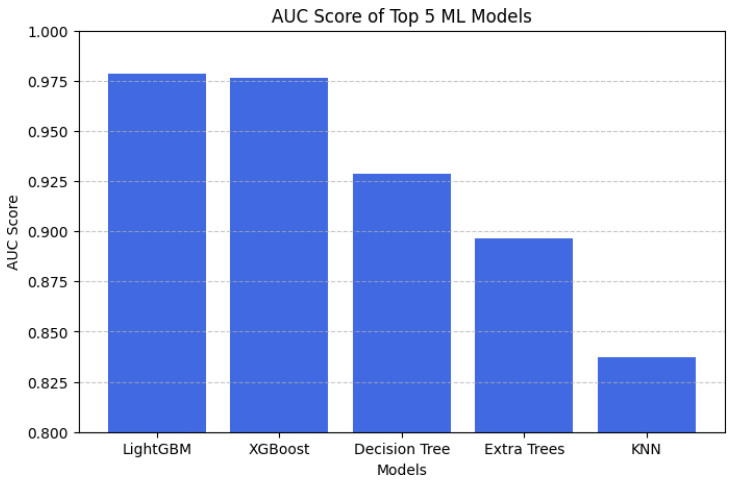

Figure 4: AUC Score of Top 5 ML Models. LightGBM and XGBoost achieved the highest AUC scores, indicating their strong predictive performance.

Simple models like Decision Trees and K-Nearest Neighbors (KNN) exhibited lower predictive performance. The Decision Tree model, despite achieving an accuracy of 90.83%, showed a lower generalization capability compared to ensemble methods. Extra Trees and KNN performed the weakest among the selected models, with KNN achieving an accuracy of only 68.85%. The performance of KNN is likely hindered by its sensitivity to high-dimensional feature spaces and the curse of dimensionality, leading to suboptimal classification results. The computational efficiency of the models varied significantly. While Decision Trees and KNN exhibited the fastest training times, LightGBM and XGBoost required substantially more computational resources. This is expected, given the iterative nature of gradient boosting algorithms, which require multiple rounds of optimization to fine-tune their decision boundaries. The increased training time for these models, however, is justified by their superior accuracy and generalization ability. The findings highlight the importance of selecting an appropriate machine learning model for bioactivity prediction. Ensemble-based models like LightGBM and XGBoost provide the best trade-off between accuracy and computational cost, making them suitable for real world virtual screening applications. However, the choice of model should also consider deployment constraints, such as available computational power and real-time inference requirements.

7. CONCLUSION

This study explored the application of machine learning models for predicting Acetylcholinesterase (AChE) inhibitory activity, aiming to enhance virtual screening methods in drug discovery. Using a dataset of 7,717 compounds with

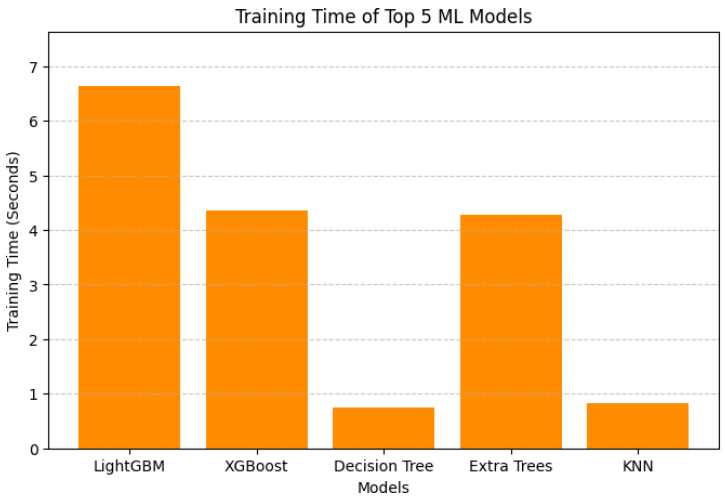

Figure 5: Training Time of Top 5 ML Models. LightGBM and XGBoost had longer training times compared to Decision Trees and KNN, reflecting the com- plexity of their learning process.

882 molecular descriptors, we implemented and evaluated multiple ML models through the PyCaret framework. Our results demonstrate that gradient boosting techniques, particularly LightGBM and XGBoost, outperform traditional models in accuracy, recall, and precision. These models efficiently handle high-dimensional cheminformatics data, making them well-suited for bioactivity prediction tasks. While ensemble models require longer training times, their superior performance justifies their use in virtual screening applications. This study underscores the importance of AI-driven approaches in cheminformatics, providing an efficient alternative to traditional computational screening methods. The integration of ML in bioactivity prediction accelerates drug discovery by reducing experimental costs and optimizing candidate selection. Future research should focus on improving model interpretability, integrating molecular docking simulations, and incorporating deep learning techniques such as Graph Neural Networks (GNNs) to further enhance predictive accuracy and generalization capabilities. By leveraging AI-based methodologies, this research contributes to the advancement of data-driven drug discovery and cheminformatics.

REFERENCES

[1] Silman, I., & Sussman, J. L. (2008). Structure and function of acetyl- cholinesterase. Progress in Neurobiology, 82(6), 233-254.

[2] Taylor, P., & Radi´c, Z. (2010). The cholinesterases: from genes to proteins.

[3] Annual Review of Pharmacology and Toxicology, 37(1), 281-320.

[4] Marrs, T. C., Maynard, R. L., & Sidell, F. R. (2012). Chemical Warfare Agents: Toxicology and Treatment. John Wiley & Sons.

[5] Matsumura, F. (2010). Toxicology of Insecticides. Springer Science & Busi- ness Media.

[6] Hopkins, A. L. (2014). Network pharmacology: the next paradigm in drug discovery. Nature Chemical Biology, 4(11), 682-690.

[7] Vamathevan, J., et al. (2019). Applications of machine learning in drug discovery and development. Nature Reviews Drug Discovery, 18(6), 463-477.

[8] Cherkasov, A., et al. (2014). QSAR modeling: where have you been? Where are you going to? Journal of Medicinal Chemistry, 57(12), 4977-5010.

[9] Cai, C., et al. (2020). Machine learning for drug discovery: an update on the recent advances. Drug Discovery Today, 25(8), 1627-1637.

[10] M. Ali, PyCaret: An open-source, low-code machine learning library in Python, 2020.

[11] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, LightGBM: A Highly Efficient Gradient Boosting Decision Tree, Ad- vances in Neural Information Processing Systems 30 (NIPS 2017), 2017.

[12] T. Chen and C. Guestrin, XGBoost: A Scalable Tree Boosting System, Pro- ceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016.

[13] S. Moorthy, D. J. Rogers, and J. E. Dahlke, Practical guidelines for the use of gradient boosting for molecular activity and property prediction, Journal of Cheminformatics, vol. 15, no. 1, pp. 1–15, 2023.

[14] Y. Zhang, H. Liu, and J. Zhang, Discovery of novel NLRP3 inhibitors based on machine learning models, BMC Chemistry, vol. 17, no. 1, pp. 1–12, 2023.

[15] T. K. Ho, J. J. Hull, and S. N. Srihari, Compound virtual screening by learning-to-rank with gradient boosting decision trees, IEEE/ACM Trans- actions on Computational Biology and Bioinformatics, vol. 19, no. 3, pp. 1234–1244, 2022.

[16] R. P. Sheridan, A. Liaw, and M. Tudor, Light gradient boosting machine as a regression method for quantitative structure–activity relationships, Journal of Chemical Information and Modeling, vol. 61, no. 5, pp. 2204–2210, 2021.

[17] Cherkasov, A., et al. (2014). QSAR modeling: where have you been? Where are you going to? Journal of Medicinal Chemistry, 57(12), 4977-5010.

[18] Kar, S., & Roy, K. (2014). Pharmacophore mapping: a step in drug dis- covery. Computational Biology and Chemistry, 53, 23-32.

[19] Vamathevan, J., et al. (2019). Applications of machine learning in drug discovery and development. Nature Reviews Drug Discovery, 18(6), 463-477.

[20] Baskin, I. I. (2016). The power of deep learning to map chemical space.

[21] Molecular Informatics, 35(11-12), 651-657.

[22] Ma, J., et al. (2015). Deep neural nets as a method for quantitative structure-activity relationships. Journal of Chemical Information and Mod- eling, 55(2), 263-274.

[23] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting sys- tem. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794).

[24] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu,

[25] T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree.

[26] Advances in Neural Information Processing Systems.

[27] Zhang, Y., Liu, H., & Zhang, J. (2023). Discovery of novel NLRP3 in- hibitors based on machine learning models. BMC Chemistry, 17(1), 1–12.

[28] Ho, T. K., Hull, J. J., & Srihari, S. N. (2022). Compound vir- tual screening by learning-to-rank with gradient boosting decision trees. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 19(3), 1234–1244.

[29] Sheridan, R. P., Liaw, A., & Tudor, M. (2021). Light gradient boosting machine as a regression method for quantitative structure–activity relation- ships. Journal of Chemical Information and Modeling, 61(5), 2204–2210.

[30] Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., & Overington, J. P. (2012). ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Research, 40(D1), D1100-D1107.

[31] Copeland, R. A. (2005). Evaluation of enzyme inhibitors in drug discovery: a guide for medicinal chemists and pharmacologists. John Wiley & Sons.

[32] Yap, C. W. (2011). PaDEL-Descriptor: An open-source software to calcu- late molecular descriptors and fingerprints. Journal of Computational Chem- istry, 32(7), 1466-1474.

[33] Yap, C. W. (2011). PaDEL-Descriptor: An open-source software to calcu- late molecular descriptors and fingerprints. Journal of Computational Chem- istry, 32(7), 1466-1474.

[34] Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (2001). Ex- perimental and computational approaches to estimate solubility and perme- ability in drug discovery and development settings. Advanced Drug Delivery Reviews, 46(1-3), 3-26.