Original Article

Check for updates

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

Malacca Pharmaceutics

Vol. 2, No. 2, 2024

Application of Ensemble Machine Learning Methods for QSAR Classification of Leukotriene A4 Hydrolase Inhibitors in Drug Discovery

Teuku Rizky Noviandy [1], Ghifari Maulana Idroes [2], Fazlin Mohd Fauzi [3] and Rinaldi Idroes [4,*]

[1] Interdisciplinary Innovation Research Unit, Graha Primera Saintifika, Aceh Besar 23771, Indonesia; trizkynoviandy@gmail.com (T.R.N.)
[2] Department of Nuclear Engineering and Engineering Physics, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia; ghifarimaulana145@gmail.com (G.M.I.)
[3] Faculty of Pharmacy, Universiti Teknologi MARA Selangor, Puncak Alam Campus, 42300 Bandar Puncak Alam, Selangor, Malaysia; fazlin5465@uitm.edu.my (F.M.F.)
[4] Department of Pharmacy, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala Banda Aceh 23111, Indonesia; rinaldi.idroes@usk.ac.id (R.I.)

* Correspondence: rinaldi.idroes@usk.ac.id

Article History

Received 11 July 2024
Revised 2 September 2024
Accepted 11 September 2024
Available Online 20 September 2024

Keywords:
Inflammatory diseases
Inhibitor screening
Molecular descriptors
Pharmacoinformatics
Supervised learning

Abstract

Inflammatory diseases such as asthma, rheumatoid arthritis, and cardiovascular conditions are driven by the overproduction of leukotriene B4 (LTB_4), a potent inflammatory mediator. Leukotriene A_4 hydrolase (LTA_4H) plays a critical role in converting leukotriene A_4 into LTB_4, making it a prime target for drug discovery. Despite ongoing efforts, developing effective LTA_4H inhibitors has been challenging due to the enzyme's complex binding properties and potential inhibitors' structural diversity. Traditional drug discovery methods, like high-throughput screening (HTS), are often time-consuming and inefficient, prompting the need for more advanced approaches. Quantitative Structure-Activity Relationship (QSAR) modeling, enhanced by ensemble machine learning techniques, provides a promising solution by enabling accurate prediction of compound bioactivity based on molecular descriptors. This study employed six ensemble machine learning methods (AdaBoost, Extra Trees, Gradient Boosting, LightGBM, Random Forest, and XGBoost) to classify LTA_4H inhibitors. The results show that the LightGBM model achieved the highest classification accuracy (83.59%) and Area Under the Curve (AUC) value (0.901), outperforming other models. These findings suggest that ensemble machine learning models, particularly LightGBM, are highly effective in predicting bioactivity, offering valuable tools for early-stage drug discovery. The results indicate that ensemble methods significantly enhance QSAR model accuracy, making them viable for identifying promising LTA_4H inhibitors and potentially accelerating the development of anti-inflammatory therapies.

1. Introduction

Inflammatory diseases, such as asthma, rheumatoid arthritis, psoriasis, and certain cardiovascular conditions, present ongoing challenges in medicine due to their chronic nature and complex underlying mechanisms [1, 2]. At the molecular level, leukotriene B_4 (LTB_4) is a key mediator in these inflammatory processes. It attracts immune cells like neutrophils and macrophages to inflammation sites, amplifying the body's response [3]. While this response is crucial for normal immune defense, it can damage tissue when dysregulated [4].

The production of LTB_4 involves the enzyme leukotriene A_4 hydrolase (LTA_4H), which catalyzes the conversion of leukotriene A_4 (LTA_4) into LTB_4 [5]. This makes LTA_4H an attractive target for developing new therapies to control excessive inflammation [6, 7]. However, efforts to develop effective inhibitors of LTA_4H have been met with significant challenges. The enzyme's complex binding properties and the diverse structures of potential inhibitors have made the drug development process difficult, often resulting in candidate molecules that show initial promise but fail later due to issues like poor efficacy, unintended effects, or toxicity [8, 9].

Traditional methods used in drug discovery, such as high-throughput screening (HTS) and structure-based drug design (SBDD), have provided important leads [10, 11]. However, these approaches are often labor-intensive, time-consuming, and costly [12, 13]. While they can identify many promising compounds, many do not make it past preclinical or clinical trials because they lack sufficient bioactivity and stability or have poor pharmacokinetic profiles [14]. This has prompted interest in developing new strategies to enhance the efficiency of finding and refining LTA_4H inhibitors, potentially using advanced computational techniques to predict compound behavior and improve drug design.

Pharmacoinformatics, specifically through Quantitative Structure-Activity Relationship (QSAR) models, offers a valuable tool in early-stage drug discovery by providing predictive insights into the biological activity of new chemical entities (NCEs). QSAR modeling establishes a mathematical relationship between the chemical structure of compounds and their biological effects, enabling the prediction of pharmacological properties such as potency, selectivity, and toxicity. In the context of LTA_4H inhibitors, QSAR models can facilitate the identification of potential lead compounds by screening large chemical libraries and predicting which molecules are likely to exhibit inhibitory activity against the enzyme.

In recent years, machine learning has emerged as a powerful tool in pharmacoinformatics, enhancing the ability to model complex, nonlinear relationships between molecular descriptors and biological activity. Ensemble machine learning methods, in particular, have shown significant promise in improving the accuracy and robustness of QSAR models [15, 16]. Ensemble methods, which combine the predictions of multiple models, can reduce the risk of overfitting, handle diverse and noisy datasets more effectively, and improve predictive performance [17–19].

This study aims to develop a robust QSAR model for classifying LTA_4H inhibitors using ensemble machine learning techniques. By utilizing these advanced methods, the study seeks to enhance predictive accuracy and robustness, enabling the identification of lead compounds with high therapeutic potential for inhibiting LTA_4H. Through this approach, the study aims to contribute to discovering novel anti-inflammatory drugs. Ultimately, the study strives to streamline the drug discovery process and provide promising candidates for treating diseases characterized by excessive inflammation, including asthma, rheumatoid arthritis, and cardiovascular diseases.

2. Materials and Methods

This study employs a systematic approach to develop Quantitative Structure-Activity Relationship (QSAR) models for predicting the activity of LTA_4H inhibitors. The methodology consists of several key steps: data collection and preprocessing, exploratory data analysis, feature preprocessing, model development using various ensemble machine learning techniques, and model evaluation and validation (Figure 1).

2.1. Data Collection and Preprocessing

The biological activity data for LTA_4H inhibitors were retrieved from the ChEMBL database (ChEMBL target ID: CHEMBL4618) [20]. Initially, the dataset comprised 1,042 compounds with reported half-maximal inhibitory concentration (IC_{50}) values. Duplicates and entries with incomplete data were removed during preprocessing, resulting in a final dataset of 636 unique compounds.

The IC_{50} values were converted into pIC_{50}, a logarithmic transformation commonly used in drug discovery to standardize activity values [21]. Compounds with pIC_{50} values greater than or equal to 6 were labeled as active, while those below 6 were classified as inactive [22]. This resulted in 432 active and 204 inactive compounds, forming the basis for the subsequent classification modeling. Representative chemical structures from the dataset used in this study can be seen in Figure 2.

2.2. Exploratory Data Analysis

As part of the Exploratory Data Analysis (EDA), we first visualized the chemical space of the dataset using Principal Component Analysis (PCA). PCA was employed to reduce the dimensionality of the molecular descriptor data and project the compounds onto a two-dimensional space [23]. This visualization allowed for a clearer understanding of the overall structure of the dataset and helped reveal any natural separation between active and inactive compounds [24]. By reducing the high-dimensional descriptor data, PCA provided insights into the dataset's chemical diversity.

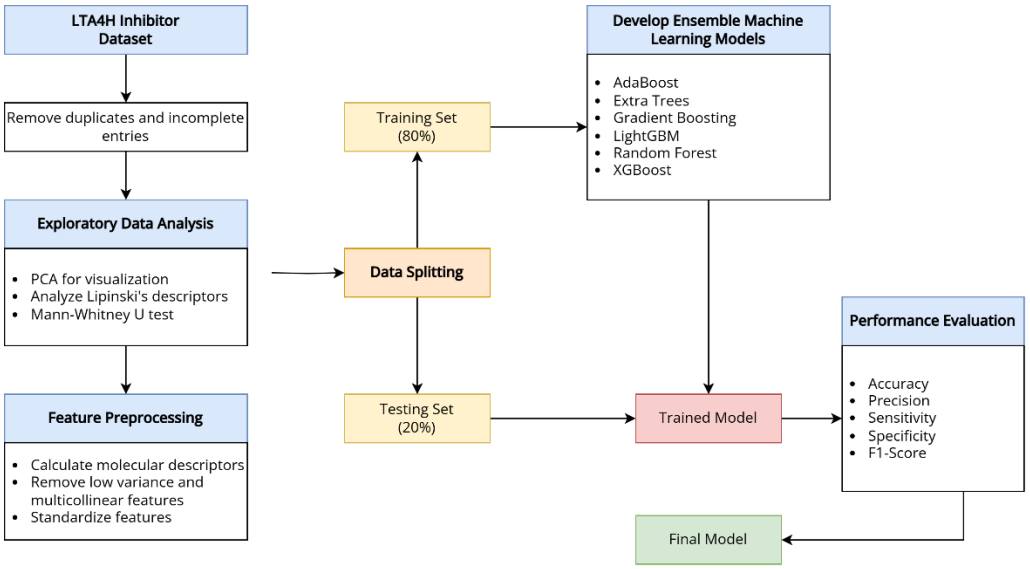

Figure 1. Flowchart summarizing the methodological approach used in the development of QSAR models for LTA_4H inhibitors.

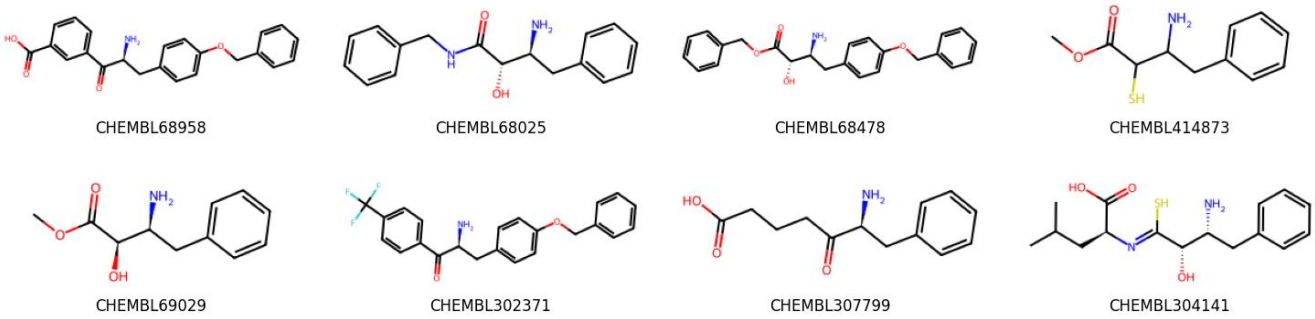

Figure 2. Representative chemical structures from the dataset used in this study.

In addition to PCA, EDA was performed on Lipinski's Rule of Five descriptors, including molecular weight (MW), logP, hydrogen bond donors (nHD), and hydrogen bond acceptors (nHA), to assess the compounds' drug-like properties [25]. Descriptive statistics such as min, max, median, mean, skew, and kurtosis were calculated for each descriptor, and box plots were generated to visualize their distributions and detect potential outliers.

The Mann-Whitney U test was performed to determine whether there were significant differences in the drug-like properties between the active and inactive compounds. This non-parametric test was selected for its robustness in comparing two independent groups without assuming a normal distribution [26]. The test results provided insights into whether Lipinski descriptors played a role in the activity of LTA_4H inhibitors.

2.3. Molecular Descriptors Calculation

In addition to the four Lipinski descriptors, a comprehensive set of two-dimensional molecular descriptors was calculated using the Mordred descriptor calculator to serve as input features for the machine learning models [27]. These descriptors included atomic counts, bond counts, molecular weight, topological indices, hydrophilic and hydrophobic properties, and other physicochemical properties relevant to the molecules under study. After descriptor calculation, features with zero variance were removed, as they did not provide meaningful information for model training [28]. Additionally, features exhibiting multicollinearity greater than 0.95 were dropped to prevent redundancy and improve model performance [29]. All remaining descriptors were standardized to ensure they were on the same scale, which is crucial for optimizing the performance of many machine learning algorithms sensitive to the input features' magnitude [30]. This process resulted in a final set of 450 molecular descriptors used as features for the QSAR classification models.

2.4. Development of Ensemble Machine Learning Models

This study employed six ensemble machine learning methods to develop QSAR classification models for predicting the activity of LTA_4H inhibitors. These models included AdaBoost, Extra Trees, Gradient Boosting

Machine, LightGBM, Random Forest, and XGBoost. Each model was trained on 80% of the dataset, while the remaining 20% was reserved for testing [31]. This splitting strategy ensures that models are trained on a significant portion of the data and evaluated on an independent test set for an unbiased performance assessment. Initially, each model was trained using default hyperparameters to provide a baseline for comparison.

AdaBoost is a boosting method that gives more weight to incorrectly predicted samples, helping the model focus on challenging cases [32]. Extra Trees is similar to Random Forest but introduces more randomness in the splitting criteria, leading to a more diverse set of trees and better generalization [33]. Gradient Boosting Machine builds models sequentially, with each new model focusing on correcting the errors of the previous one [34]. LightGBM is a faster and more memory-efficient version of gradient boosting that works well with large datasets due to its histogram-based approach to building trees [24, 35]. Random Forest constructs multiple decision trees during training and combines their predictions to improve accuracy and reduce overfitting [36]. XGBoost is a highly efficient and scalable implementation of gradient boosting suited for structured data, offering speed and performance improvements over traditional methods [37, 38].

Ensemble methods like these were chosen because they combine the strengths of multiple algorithms to improve overall model performance [39]. These models are well-suited to structured data like molecular descriptors by reducing the risk of overfitting and increasing robustness. Additionally, models like Random Forest and XGBoost provide useful insights into feature importance, which is valuable in drug discovery research for understanding which molecular features are most influential in predicting biological activity.

2.5. Model Evaluation and Validation

Several metrics were employed to evaluate and compare the performance of the ensemble models, including accuracy, precision, recall, and F1-score [40, 41]. These metrics provide insights into the models' ability to correctly classify active and inactive compounds while balancing false positives and negatives. The formulas for these metrics are presented in Equations 1-5:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (3)$$

$$Specificity = \frac{TN}{TN + FP} \quad (4)$$

$$F1 - Score = 2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity} \quad (5)$$

Where TP represents the number of true positives, TN represents the number of true negatives, FP represents the number of false positives, and FN represents the number of false negatives.

In addition, visualizations such as Receiver Operating Characteristic (ROC) curves, precision-recall plots, and feature importance rankings were generated. The ROC curve illustrates the trade-off between sensitivity and specificity [42], while the precision-recall plot highlights the balance between precision and recall [43]. Feature importance rankings, particularly for tree-based models, offer insights into the most influential molecular descriptors in classifying LTA_4H inhibitors [44].

2.6. Computational Tools and Software

All computational analyses were performed using Python 3.10.9. The scikit-learn, XGBoost, and LightGBM libraries implemented the machine learning models. Molecular descriptors were generated using the Mordred library. Data preprocessing, statistical analysis and visualizations were conducted using the pandas, NumPy, scipy, matplotlib, and seaborn libraries. The Mann-Whitney U test was performed using the scipy library.

3. Results and Discussion

3.1. Exploratory Data Analysis Results

The chemical space visualization using PCA (Figure 3) showed no clear linear separation between active and inactive compounds. While some clustering was observed within both groups, significant overlap occurred, indicating that active and inactive compounds share many structural features. This overlap suggests that simple linear models would struggle to effectively classify the compounds, reinforcing the need for more complex, non-linear models such as the ensemble machine learning methods used in this study.

The spread of active and inactive compounds across the principal component axes also highlights the diverse nature of the chemical space. The inability to distinctly separate the two groups in two-dimensional space further demonstrates the complexity of the molecular interactions involved in LTA_4H inhibition. This underscores the importance of leveraging advanced machine learning techniques to capture the subtle,

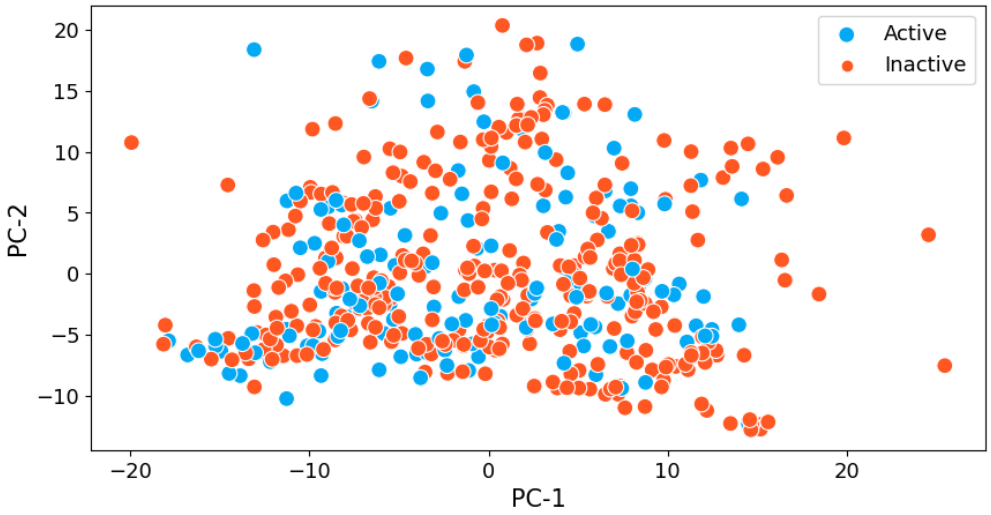

Figure 3. PCA plot of the chemical space showing the distribution of active and inactive compounds.

Table 1. Descriptive statistics and Mann-Whitney U test results for Lipinski's Rule of Five descriptors.

	MW		LogP		nHA		nHD	
	Inactive	**Active**	**Inactive**	**Active**	**Inactive**	**Active**	**Inactive**	**Active**
p-value	1.076045×10^{-16}		5.653747×10^{-9}		2.295832×10^{-3}		2.124595×10^{-11}	
Min	133.15	201.27	-0.49	0.33	0	0	0	1
Max	721.88	540.46	8.98	7.7	4	5	9	10
Median	315.61	362.98	3.24	3.93	1	1	4	5
Mean	320.09	372.43	3.2	3.89	1.51	1.19	3.76	4.63
Skew	0.77	0.26	0.2	-0.1	0.37	0.76	0.65	0.19
Kurtosis	2.76	-0.7	0.82	0.35	-0.91	0.28	0.98	-0.35

Note: The p-value represents the result of the Mann-Whitney U test.

nonlinear relationships between molecular descriptors and biological activity.

The analysis of Lipinski's Rule of Five descriptors indicated that both active and inactive compounds generally adhered to drug-like properties, with no significant outliers detected. However, the Mann-Whitney U test (Table 1) revealed statistically significant differences between active and inactive compounds in terms of molecular weight (MW), logP, hydrogen bond acceptors (nHA), and hydrogen bond donors (nHD), with p-values well below the 0.05 threshold, which suggests that these descriptors have different distributions across active and inactive compounds, which may play a role in defining LTA_4H inhibitor activity.

Regarding descriptive statistics, active compounds generally have slightly larger molecular weights, with a median of 362.98 compared to 315.61 for inactive compounds. Active compounds also show slightly higher logP values, indicating a higher tendency for lipophilicity. The skewness of the MW distribution for inactive compounds (0.77) is greater than that for active compounds (0.26), indicating that inactive compounds have a more right-skewed distribution, with a longer tail towards higher molecular weights. Similarly, the skewness for nHA in inactive compounds (0.37) is higher than in active compounds (0.76), showing a more uneven distribution among inactive compounds.

The kurtosis values provide additional insight: the negative kurtosis for MW in active compounds (-0.7) suggests a flatter distribution, while inactive compounds have a more peaked distribution (kurtosis of 2.76). For nHA, the kurtosis in inactive compounds is also negative (-0.91), further indicating a flatter distribution than active compounds (0.28). These differences in skewness and kurtosis highlight the greater variability in molecular weights and hydrogen bond acceptor distributions among inactive compounds. In contrast, active compounds tend to be more homogeneous in these properties.

Thus, the statistically significant differences in the distributions of MW, logP, nHA, and nHD between active and inactive compounds imply that these molecular descriptors play an important role in classifying LTA_4H inhibitors, contributing to their potential biological activity.

Further analysis of the Lipinski descriptors using box plots (Figure 4) visually represents the differences between active and inactive compounds. Regarding MW,

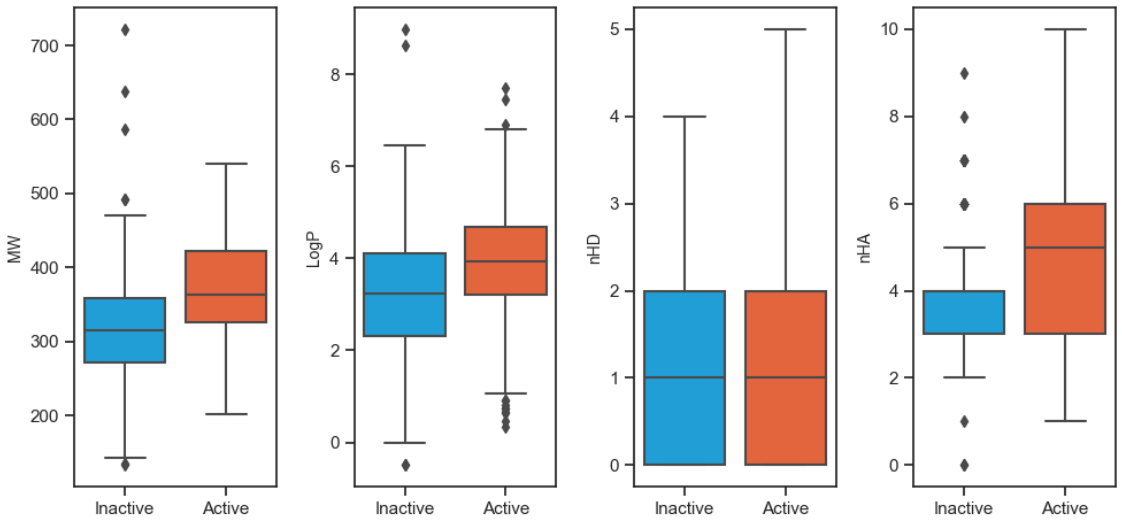

Figure 4. Box plots comparing the distributions of Lipinski's Rule of Five descriptors between active and inactive compounds.

Table 2. Performance of ensemble machine learning models on the classification of LTA_4H inhibitors.

Model	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F-1 Score (%)
AdaBoost	82.03	84.27	89.29	68.16	86.71
Extra Trees	83.59	84.62	91.67	68.18	88.00
Gradient Boosting	82.81	82.29	94.05	61.36	87.78
LightGBM	83.59	83.87	92.86	65.91	88.14
Random Forest	82.81	82.98	92.86	63.64	87.64
XGBoost	83.59	82.47	95.24	61.36	88.40

the box plot clearly shows that active compounds tend to have higher molecular weights, with a narrower interquartile range and fewer outliers than inactive compounds. This supports the observation that active compounds are more clustered around a higher MW, while inactive compounds exhibit greater variability.

The logP distribution for active compounds shows higher values overall, with the median logP for active compounds being higher than that of inactive compounds. Low outliers further emphasize this difference in the inactive group, which are less prevalent in the active group. The tighter distribution of logP for active compounds suggests that greater lipophilicity may contribute to their inhibitory activity, reinforcing the trends identified through the Mann-Whitney U test.

For hydrogen bond donors (nHD), the box plots indicate a slight difference in medians but largely similar distributions between active and inactive compounds. This confirms that this descriptor may not significantly differentiate the two groups. On the other hand, hydrogen bond acceptors (nHA) show a clearer distinction, with active compounds tending to have more acceptors, as evidenced by the higher median and a more compact distribution. Inactive compounds, in contrast, display a broader range of nHA values and more outliers, suggesting that active compounds may require a more consistent presence of hydrogen bond acceptors for effective LTA_4H inhibition.

3.2. Machine Learning Model Performance

The performance of six ensemble machine learning models was evaluated using key classification metrics: accuracy, precision, sensitivity, specificity, and F1-score (Table 2). All models demonstrated strong classification abilities for LTA_4H inhibitors, with accuracy ranging from 82.03% to 83.59%. Extra Trees, LightGBM, and XGBoost showed the highest accuracy (83.59%), while AdaBoost had the lowest accuracy at 82.03%.

While all models performed similarly in accuracy, their sensitivity and specificity varied, revealing important differences in model behavior. For example, XGBoost exhibited the highest sensitivity at 95.24%, indicating correctly identifying active compounds (true positives). However, this was accompanied by lower specificity (61.36%), meaning that XGBoost also had a higher rate of false positives, classifying inactive compounds as active. This trade-off suggests that XGBoost is highly sensitive to identifying potential drug candidates but at the cost of generating more false positives, which may increase

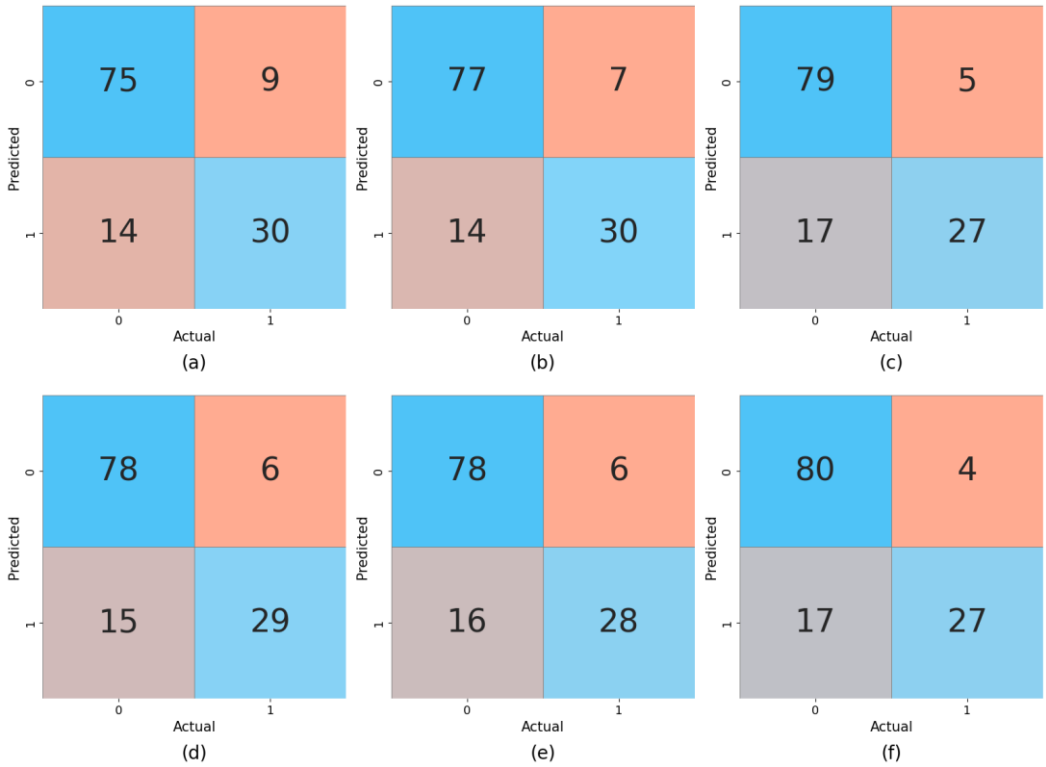

Figure 5. Confusion matrices of the ensemble machine learning models for the classification of LTA_4H inhibitors: (a) AdaBoost, (b) Extra Trees, (c) Gradient Boosting, (d) LightGBM, (e) Random Forest, and (f) XGBoost.

resource expenditure in the later stages of drug development.

In contrast, AdaBoost and Extra Trees had the highest specificity (68.16% and 68.18%, respectively), indicating they were better at correctly identifying inactive compounds (true negatives). This suggests that these models may be more conservative, minimizing false positives but potentially missing some active compounds, as reflected by their lower sensitivity scores.

The F1 Score, which balances precision and sensitivity, was highest for XGBoost at 88.40%, indicating its overall strong classification performance despite the trade-off in specificity. LightGBM and Extra Trees followed closely, with F1 Scores of 88.14% and 88.00%, respectively. These results indicate that LightGBM and Extra Trees maintained a good balance between correctly identifying active and inactive compounds.

The differences in performance across these models can be attributed to how they handle the inherent complexity and non-linearity in the molecular descriptor data. XGBoost and LightGBM, which are gradient-boosting-based methods, tend to capture subtle, non-linear relationships in the data more effectively, leading to higher sensitivity. However, this also makes them more prone to overfitting the active compounds, as seen in their lower specificity. In contrast, models like AdaBoost, which applies weight adjustments to handle misclassified data, and Extra Trees, which introduces randomness to improve generalization, excel at reducing false positives but may be less effective at capturing all potential actives.

Thus, while XGBoost and LightGBM show higher promise for identifying potential LTA_4H inhibitors, they may be more suited for early-stage drug discovery when the goal is to maximize the identification of possible candidates. Models like Extra Trees and AdaBoost may be more useful in later stages, where minimizing false positives becomes more critical to avoid pursuing ineffective compounds.

The confusion matrices in Figure 5 provide a detailed view of the classification performance of the six ensemble models—AdaBoost, Extra Trees, Gradient Boosting, LightGBM, Random Forest, and XGBoost—used to predict active and inactive LTA_4H inhibitors. Each matrix highlights the number of true positives, false positives, and false negatives, offering insights into how well the models handle the binary classification task.

Regarding true positives, Gradient Boosting, AdaBoost, Extra Trees, and LightGBM correctly classified 30 active compounds, while XGBoost and Random Forest slightly lagged with 27 and 28 true positives, respectively. The high number of true positives indicates that these models are adept at identifying compounds with inhibitory activity against LTA_4H. However, a slight difference in

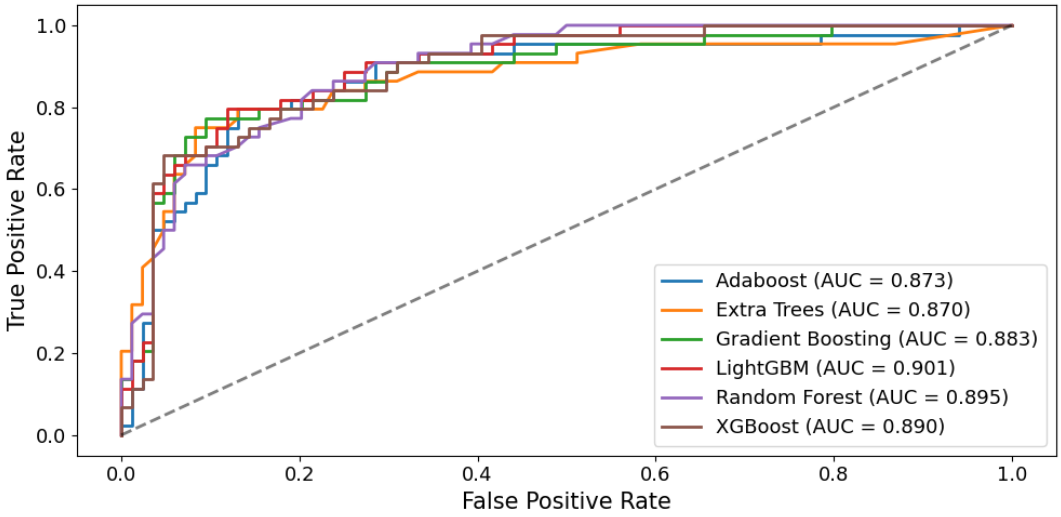

Figure 6. ROC curves with AUC values for ensemble models in predicting LTA_4H inhibitors.

their ability to identify active compounds is evident, with Gradient Boosting emerging as the best performer. The performance across the models is similar for true negatives, which represent correctly classified inactive compounds, with XGBoost and LightGBM leading the group by correctly classifying 80 inactive compounds each. This suggests that these models effectively avoid false positives and accurately identify inactive compounds.

False positives, or instances where inactive compounds were incorrectly classified as active, were lowest in Gradient Boosting (5), followed by XGBoost (4) and AdaBoost (9). This makes these models particularly valuable for drug discovery, where minimizing false positives is crucial to prevent investing in compounds unlikely to show biological activity. Extra Trees, LightGBM, and Random Forest also minimized false positives, with 6 to 7 errors reflecting their robustness. The low false positive rates across all models indicate their strong ability to distinguish inactive compounds from active ones accurately.

The false negative rate, where active compounds were misclassified as inactive, was lowest for AdaBoost and Gradient Boosting, with 14 false negatives each. These models exhibit a high sensitivity, meaning they are less likely to miss potential drug candidates. However, XGBoost had the highest false-negative count (17), indicating a trade-off between reducing false positives and missing some true active compounds. This slight increase in false negatives for XGBoost suggests a more conservative classification approach, focusing on ensuring that inactive compounds are not misclassified as active. LightGBM and Random Forest also exhibited a modest number of false negatives (15 and 16, respectively), indicating that while they are strong classifiers, they may miss some active compounds during the classification process.

Figure 6 shows the ROC curves for the six ensemble machine learning models that classify LTA_4H inhibitors. The Area Under the Curve (AUC) values quantitatively measure the model's ability to distinguish between active and inactive compounds. Among the models, LightGBM achieves the highest AUC value of 0.901, indicating the best performance balancing true positive and false positive rates. Random Forest and XGBoost follow closely with AUC values of 0.895 and 0.890, further confirming their strong classification abilities. Gradient Boosting also performs well with an AUC of 0.883, while AdaBoost and Extra Trees slightly lag with AUC values of 0.873 and 0.870, respectively. These ROC curves indicate that all models strongly discriminate between the two classes, with LightGBM being the most effective. The relatively close range of AUC values suggests that each model has comparable performance, although LightGBM, Random Forest, and XGBoost consistently demonstrate superior predictive capabilities.

Figure 7 illustrates the Precision-Recall curves for the six ensemble models that classify LTA_4H inhibitors. The PR curve is a valuable tool for evaluating models when dealing with imbalanced data, as it emphasizes the trade-off between precision and recall. We observe that Extra Trees performs best at low recall values, maintaining high precision, which indicates its strength in ensuring that they are highly likely to be true positives when it predicts active compounds. LightGBM, Random Forest, and XGBoost exhibit strong, balanced performances across a wide range of recall values, maintaining good precision even as recall increases. AdaBoost shows a sharper drop-off in precision at lower recall values, making it less

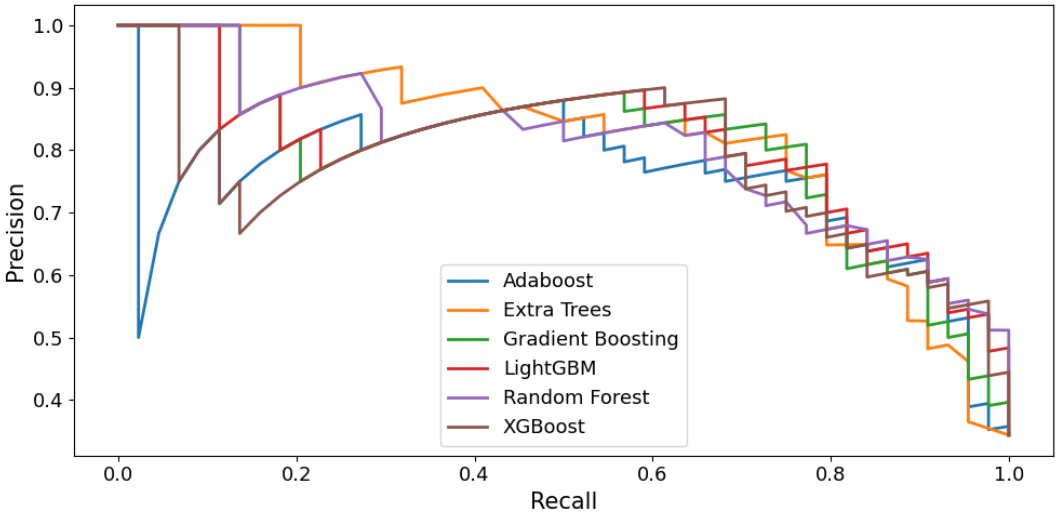

Figure 7. Precision-Recall curves for ensemble models in predicting LTA_4H inhibitors.

effective when the recall is low. Overall, the PR curves highlight that Extra Trees excels in precision for a narrower range of high-confidence predictions. At the same time, LightGBM, Random Forest, and XGBoost provide strong, consistent performance, balancing both recall and precision across the dataset.

The results of this study demonstrate that ensemble machine learning models provide powerful tools for classifying LTA_4H inhibitors, with each model offering unique strengths. LightGBM consistently emerged as the top performer among the models tested, achieving the highest AUC value (0.901) in the ROC analysis and strong performance in precision, recall, and balanced classification metrics. XGBoost and Random Forest also performed well, showing high AUC values (0.890 and 0.895, respectively) while maintaining good precision-recall balances across various thresholds. With their robust classification abilities, these models are well-suited for pharmaceutical applications where high accuracy in identifying active compounds is critical.

Identifying active inhibitors like LTA_4H inhibitors efficiently is essential for drug discovery. Models such as LightGBM, Random Forest, and XGBoost are well-suited for this task due to their ability to manage complex, non-linear relationships in molecular data. LightGBM's performance is notable for its speed and computational efficiency, making it ideal for processing large datasets commonly encountered in pharmaceutical research. The high sensitivity of models like Gradient Boosting and AdaBoost is also valuable, as it ensures that potential drug candidates are not overlooked, even if it results in slightly more false positives. This balance between sensitivity and specificity is critical in early-stage drug discovery, where prioritizing identifying as many promising candidates as possible often takes precedence over minimizing false positives.

However, while ensemble methods have clear advantages, there are limitations. XGBoost and Random Forest, while strong performers, showed a tendency to misclassify some active compounds as inactive (higher false negative rates), which could lead to missing potential therapeutic candidates. Similarly, models like AdaBoost showed a sharper drop in precision as recall increased, indicating that it may struggle in applications requiring high precision at high recall values. These limitations suggest that while ensemble models are powerful, their performance may vary depending on the dataset and the specific pharmacological context.

Another limitation is that the models are primarily data-driven, relying heavily on the molecular descriptors' quality as input. Although these models performed well with the calculated descriptors, future improvements could include using more advanced molecular representations, such as graph neural networks or deep learning-based molecular embeddings, which might capture more subtle structural features of the compounds. Additionally, the dataset is relatively small for machine learning applications, and expanding the dataset or using data augmentation techniques could improve model generalization.

Future directions for this research could focus on optimizing the models through hyperparameter tuning, ensemble stacking, or hybrid models that combine the strengths of multiple algorithms. Further investigation into explainability could provide deeper insights into the specific molecular features driving the classification decisions, which would be particularly valuable in drug discovery applications. Integrating these machine

learning models with in silico methods like molecular docking or virtual screening could enhance the drug discovery pipeline by providing a multi-step, high-throughput approach to identifying and optimizing potential drug candidates.

4. Conclusions

This study demonstrated the effectiveness of ensemble machine learning models in accurately classifying LTA_4H inhibitors, a key step in the drug discovery process for inflammatory diseases. LightGBM emerged as the top performer among the models tested, exhibiting the best balance between precision, recall, and computational efficiency, followed closely by Random Forest and XGBoost. These models proved highly capable of handling complex molecular data, providing valuable tools for predicting bioactive compounds. While limitations such as false negatives and dataset size were noted, these results highlight the potential of ensemble methods to enhance pharmaceutical research, particularly when integrated with more advanced molecular representations and further optimization. Future work could focus on model interpretability, larger datasets, and combining machine learning with traditional in silico approaches to streamline drug discovery pipelines.

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

Funding: This study does not receive external funding.

Ethical Clearance: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original dataset used in this study was retrieved from ChEMBL. The processed data used in the analysis are available upon request from the corresponding author.

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

References

1. Bennett, J. M., Reeves, G., Billman, G. E., and Sturmberg, J. P. (2018). Inflammation–Nature's Way to Efficiently Respond to All Types of Challenges: Implications for Understanding and Managing "the Epidemic" of Chronic Diseases, *Frontiers in Medicine*, Vol. 5. doi:10.3389/fmed.2018.00316.
2. Campanati, A., Marani, A., Martina, E., Diotallevi, F., Radi, G., and Offidani, A. (2021). Psoriasis as an Immune-Mediated and Inflammatory Systemic Disease: From Pathophysiology to Novel Therapeutic Approaches, *Biomedicines*, Vol. 9, No. 11, 1511. doi:10.3390/biomedicines9111511.
3. He, R., Chen, Y., and Cai, Q. (2020). The Role of the LTB4-BLT1 Axis in Health and Disease, *Pharmacological Research*, Vol. 158, 104857. doi:10.1016/j.phrs.2020.104857.
4. Brandt, S. L., and Serezani, C. H. (2017). Too Much of a Good Thing: How Modulating Ltb 4 Actions Restore Host Defense in Homeostasis or Disease, *Seminars in Immunology*, Vol. 33, 37–43. doi:10.1016/j.smim.2017.08.006.
5. Haeggström, J. Z. (2004). Leukotriene A4 Hydrolase/Aminopeptidase, the Gatekeeper of Chemotactic Leukotriene B4 Biosynthesis, *Journal of Biological Chemistry*, Vol. 279, No. 49, 50639–50642. doi:10.1074/jbc.R400027200.
6. Röhn, T. A., Numao, S., Otto, H., Loesche, C., and Thoma, G. (2021). Drug Discovery Strategies for Novel Leukotriene A4 Hydrolase Inhibitors, *Expert Opinion on Drug Discovery*, Vol. 16, No. 12, 1483–1495. doi:10.1080/17460441.2021.1948998.
7. Qin, R., Wang, H., and Yan, A. (2021). Classification and QSAR Models of Leukotriene A4 Hydrolase (LTA4H) Inhibitors by Machine Learning Methods, *SAR and QSAR in Environmental Research*, Vol. 32, No. 5, 411–431. doi:10.1080/1062936X.2021.1910862.
8. Li, X., Xie, M., Lu, C., Mao, J., Cao, Y., Yang, Y., Wei, Y., Liu, X., Cao, S., Song, Y., Peng, J., Zhou, Y., Jiang, Q., Lin, G., Qin, S., Qi, M., Hou, M., Liu, X., Zhou, H., Yang, G., and Yang, C. (2020). Design and Synthesis of Leukotriene A4 Hydrolase Inhibitors to Alleviate Idiopathic Pulmonary Fibrosis and Acute Lung Injury, *European Journal of Medicinal Chemistry*, Vol. 203, 112614. doi:10.1016/j.ejmech.2020.112614.
9. Wang, Z., and Yang, B. (2022). Polypharmacology in Clinical Applications—Anti-inflammation Polypharmacology, *Polypharmacology*, Springer International Publishing, Cham, 375–396. doi:10.1007/978-3-031-04998-9_11.
10. Berdigaliyev, N., and Aljofan, M. (2020). An Overview of Drug Discovery and Development, *Future Medicinal Chemistry*, Vol. 12, No. 10, 939–947. doi:10.4155/fmc-2019-0307.
11. Batool, M., Ahmad, B., and Choi, S. (2019). A Structure-Based Drug Discovery Paradigm, *International Journal of Molecular Sciences*, Vol. 20, No. 11, 2783. doi:10.3390/ijms20112783.
12. Bano, I., Butt, U. D., and Mohsan, S. A. H. (2023). New Challenges in Drug Discovery, *Novel Platforms for Drug Delivery Applications*, Elsevier, 619–643. doi:10.1016/B978-0-323-91376-8.00021-5.
13. Satpathy, R. (2024). Artificial Intelligence Techniques in the Classification and Screening of Compounds in Computer-Aided Drug Design (CADD) Process, *Artificial Intelligence and Machine Learning in Drug Design and Development*, Wiley, 473–497. doi:10.1002/9781394234196.ch15.
14. Lanne, A., Usselmann, L. E. J., Llowarch, P., Michaelides, I. N., Fillmore, M., and Holdgate, G. A. (2023). A Perspective on the Changing Landscape of Hts, *Drug Discovery Today*, Vol. 28, No. 8, 103670. doi:10.1016/j.drudis.2023.103670.
15. 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.
16. Khan, M. B., Shahrior, R., Asha, R. T., and Saha, P. S. (2021). Predicting AXL Inhibition of Chemicals using Molecular Descriptors and Machine Learning Methods, *2021 5th International Conference on Electrical Information and Communication Technology (EICT)*, IEEE, 1–6. doi:10.1109/EICT54103.2021.9733504.
17. Noviandy, T. R., Maulana, A., Idroes, G. M., Emran, T. B., Tallei, T. E., Helwani, Z., and Idroes, R. (2023). Ensemble Machine Learning Approach for Quantitative Structure Activity

Relationship Based Drug Discovery: A Review, *Infolitika Journal of Data Science*, Vol. 1, No. 1, 32–41. doi:10.60084/ijds.v1i1.91.

18. Supriatna, D. J. I., Saputra, H., and Hasan, K. (2023). Enhancing the Red Wine Quality Classification Using Ensemble Voting Classifiers, *Infolitika Journal of Data Science*, Vol. 1, No. 2, 42–47. doi:10.60084/ijds.v1i2.95.

19. Noviandy, T. R., Nainggolan, S. I., Raihan, R., Firmansyah, I., and Idroes, R. (2023). Maternal Health Risk Detection Using Light Gradient Boosting Machine Approach, *Infolitika Journal of Data Science*, Vol. 1, No. 2, 48–55. doi:10.60084/ijds.v1i2.123.

20. 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.

21. Thakur, A., Kumar, A., Sharma, V., and Mehta, V. (2022). PIC50: An open source tool for interconversion of PIC50 values and IC50 for efficient data representation and analysis, *BioRxiv*, 2022.10.15.512366. doi:10.1101/2022.10.15.512366.

22. 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.

23. Gaspar, H. A., Baskin, I. I., and Varnek, A. (2016). Visualization of a Multidimensional Descriptor Space, 243–267. doi:10.1021/bk-2016-1222.ch012.

24. Noviandy, T. R., Idroes, G. M., and Hardi, I. (2024). An Interpretable Machine Learning Strategy for Antimalarial Drug Discovery with LightGBM and SHAP, *Journal of Future Artificial Intelligence and Technologies*, Vol. 1, No. 2, 84–95. doi:10.62411/faith.2024-16.

25. Chen, X., Li, H., Tian, L., Li, Q., Luo, J., and Zhang, Y. (2020). Analysis of the Physicochemical Properties of Acaricides Based on Lipinski's Rule of Five, *Journal of Computational Biology*, Vol. 27, No. 9, 1397–1406. doi:10.1089/cmb.2019.0323.

26. Aqeel, I., Bilal, M., Majid, A., and Majid, T. (2022). Hybrid Approach to Identifying Druglikeness Leading Compounds against COVID-19 3CL Protease, *Pharmaceuticals*, Vol. 15, No. 11, 1333. doi:10.3390/ph15111333.

27. Moriwaki, H., Tian, Y. S., Kawashita, N., and Takagi, T. (2018). Mordred: A Molecular Descriptor Calculator, *Journal of Cheminformatics*, Vol. 10, No. 1, 1–14. doi:10.1186/s13321-018-0258-y.

28. Noviandy, T. R., Maulana, A., Idroes, G. M., Irvanizam, I., Subianto, M., and Idroes, R. (2023). QSAR-Based Stacked Ensemble Classifier for Hepatitis C NS5B Inhibitor Prediction, *2023 2nd International Conference on Computer System, Information Technology, and Electrical Engineering (COSITE)*, IEEE, 220–225. doi:10.1109/COSITE60233.2023.10250039.

29. Noviandy, T. R., Idroes, G. M., and Hardi, I. (2024). Machine Learning Approach to Predict AXL Kinase Inhibitor Activity for Cancer Drug Discovery Using XGBoost and Bayesian Optimization, *Journal of Soft Computing and Data Mining*, Vol. 5, No. 1, 46–56.

30. Noviandy, T. R., Maulana, A., Idroes, G. M., Maulydia, N. B., Patwekar, M., Suhendra, R., and Idroes, R. (2023). Integrating Genetic Algorithm and LightGBM for QSAR Modeling of Acetylcholinesterase Inhibitors in Alzheimer's Disease Drug Discovery, *Malacca Pharmaceutics*, Vol. 1, No. 2, 48–54. doi:10.60084/mp.v1i2.60.

31. Idroes, R., Noviandy, T. R., Maulana, A., Suhendra, R., and Sasmita, N. R. (2023). ANFIS-Based QSRR Modelling for Kovats Retention Index Prediction in Gas Chromatography, *Infolitika Journal of Data Science*, Vol. 1, No. 1, 1–14. doi:10.60084/ijds.v1i1.73.

32. Noviandy, T. R., Idroes, G. M., Hardi, I., Afjal, M., and Ray, S. (2024). A Model-Agnostic Interpretability Approach to Predicting Customer Churn in the Telecommunications Industry, *Infolitika Journal of Data Science*, Vol. 2, No. 1, 34–44. doi:10.60084/ijds.v2i1.199.

33. Sari, L., Romadloni, A., Lityaningrum, R., and Hastuti, H. D. (2023). Implementation of LightGBM and Random Forest in Potential Customer Classification, *TIERS Information Technology Journal*, Vol. 4, No. 1, 43–55. doi:10.38043/tiers.v4i1.4355.

34. Suhendra, R., Husdayanti, N., Suryadi, S., Juliwardi, I., Sanusi, S., Ridho, A., Ardiansyah, M., Murhaban, M., and Ikhsan, I. (2023). Cardiovascular Disease Prediction Using Gradient Boosting Classifier, *Infolitika Journal of Data Science*, Vol. 1, No. 2, 56–62. doi:10.60084/ijds.v1i2.131.

35. Noviandy, T. R., Idroes, G. M., and Hardi, I. (2024). Enhancing Loan Approval Decision-Making: An Interpretable Machine Learning Approach Using LightGBM for Digital Economy Development, *Malaysian Journal of Computing (MJOC)*, Vol. 9, No. 1, 1734–1745. doi:10.24191/mjoc.v9i1.25691.

36. Gupta, N. S., Mohta, Y., Heda, K., Armaan, R., Valarmathi, B., and Arulkumaran, G. (2023). Prediction of Air Quality Index Using Machine Learning Techniques: A Comparative Analysis, *Journal of Environmental and Public Health*, Vol. 2023, 1–26. doi:10.1155/2023/4916267.

37. Srisongkram, T., and Weerapreeyakul, N. (2022). Drug Repurposing against KRAS Mutant G12C: A Machine Learning, Molecular Docking, and Molecular Dynamics Study, *International Journal of Molecular Sciences*, Vol. 24, No. 1, 669. doi:10.3390/ijms24010669.

38. Safriandono, A. N., Setiadi, D. R. I. M., Dahlan, A., Rahmanti, F. Z., Wibisono, I. S., and Ojugo, A. A. (2024). Analyzing Quantum Feature Engineering and Balancing Strategies Effect on Liver Disease Classification, *Journal of Future Artificial Intelligence and Technologies*, Vol. 1, No. 1, 51–63. doi:10.62411/faith.2024-12.

39. Mienye, I. D., and Sun, Y. (2022). A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects, *IEEE Access*, Vol. 10, 99129–99149. doi:10.1109/ACCESS.2022.3207287.

40. Idroes, G. M., Noviandy, T. R., Maulana, A., Zahriah, Z., Suhendrayatna, S., Suhartono, E., Khairan, K., Kusumo, F., Helwani, Z., and Abd Rahman, S. (2023). Urban Air Quality Classification Using Machine Learning Approach to Enhance Environmental Monitoring, *Leuser Journal of Environmental Studies*, Vol. 1, No. 2, 62–68. doi:10.60084/ljes.v1i2.99.

41. Noviandy, T. R., Nisa, K., Idroes, G. M., Hardi, I., and Sasmita, N. R. (2024). Classifying Beta-Secretase 1 Inhibitor Activity for Alzheimer's Drug Discovery with LightGBM, *Journal of Computing Theories and Applications*, Vol. 2, No. 2, 138–147. doi:10.62411/jcta.10129.

42. Tharwat, A. (2021). Classification Assessment Methods, *Applied Computing and Informatics*, Vol. 17, No. 1, 168–192. doi:10.1016/j.aci.2018.08.003.

43. Cook, J., and Ramadas, V. (2020). When to Consult Precision-Recall Curves, *The Stata Journal: Promoting Communications on Statistics and Stata*, Vol. 20, No. 1, 131–148. doi:10.1177/1536867X20909693.

44. Zhou, Z., and Hooker, G. (2021). Unbiased Measurement of Feature Importance in Tree-Based Methods, *ACM Transactions on Knowledge Discovery from Data*, Vol. 15, No. 2, 1–21. doi:10.1145/3429445.