Silico Predictions of Acute Aquatic Toxicity for Organic Chemicals using QSAR Model

Misba Inamdar

School of Computer Science MIT World Peace University

Pune,India [1032211287@mitwpu.edu.in](mailto:1032211287@mitwpu.edu.in)

Drishti Sinha

School of Computer Science MIT World Peace University Pune,India [1032210911@mitwpu.edu.in](mailto:1032210911@mitwpu.edu.in)

Anmol Deshratna Saxena

School of Computer Science MIT World Peace University

Pune,India [1032211624@mitwpu.edu.in](mailto:1032211624@mitwpu.edu.in)

Prof. Jayshree Aher

School of Computer Science

MIT World Peace University Pune,India <jayshree.aher@mitwpu.edu.in>

Tirth Thesiya

School of Computer Science MIT World Peace University Pune,India [1032210879@mitwpu.edu.in](mailto:1032210879@mitwpu.edu.in)

Ayush Thakre

School of Computer Science MIT World Peace University

Pune,India [1032211517@mitwpu.edu.in](mailto:1032211517@mitwpu.edu.in)

***Abstract*—** Determining the aquatic toxicity of organic compounds is still a crucial requirement for environmental management and regulatory compliance. QSAR models are a suitable alternative for use in predictive toxicology since they can assist address the problem of high prices and time spent on experimental assays, which will ultimately lead to a reduction in the cost of resources. The main goal of this work is to evaluate and forecast acute aquatic toxicity using QSAR modeling methodologies. This gives decision-makers in chemical management and risk assessment a starting point. The accuracy of forecasting is astounding thanks to a thorough data fusion and model optimization, and it can be a useful tool for determining the levels of toxicity on the abundance of organic compounds. Because computational methods are based on the fundamental principles governing the intermolecular interactions of chemical structures with aquatic species, they not only expedite the assessment process but also enhance our understanding of the fundamentals of structure-toxicity relationships.

***Keywords— QSAR modeling, Acute aquatic toxicity, Predictive toxicology, Environmental risk assessment and Computational methods***

1. INTRODUCTION

To ensure environmental safety and comprehend the possible effects of organic compounds on aquatic ecosystems, it is critical to evaluate the acute aquatic toxicity of these compounds. Alternative methods that are dependable and efficient are needed since traditional experimental assays for toxicity assessment are frequently expensive and time-consuming. In predictive toxicology, Quantitative Structure-Activity Relationship (QSAR) models have become important instruments because they provide a computational framework for predicting toxicity based on chemical structure information.

In order to meet the demand for accurate and effective toxicity prediction in environmental management, we investigate the use of QSAR modeling in this study to predict the acute aquatic toxicity of organic chemicals. We improve the accuracy of toxicity predictions by combining data fusion with model optimization, which facilitates risk assessment and decision-making. QSAR-based systems provide an affordable substitute for conventional experimental techniques, enabling rapid screening of chemical libraries and ranking compounds for additional evaluation, in accordance with computational toxicology and animal welfare guidelines.

We create a thorough framework that covers every step of QSAR modeling using Jupyter Notebook. Applications for the suggested approach include scientific research, hazard classification, product design, regulatory compliance, and environmental risk assessment. Improved accuracy, model transparency, broad applicability, increased efficiency, validation and verification, and user accessibility are among the main goals.

Through the use of QSAR modeling to predictive toxicity, we are able to preserve aquatic habitats and practice sustainable chemical management.

LITERATURE REVIEW

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sr.No | Publication Title with authors [ mention whether Journal or Conference paper] | Publication Year | Positive points of the Publication | Gaps in publication work |
| 1 | Machine Learning Models for Identification and Prediction of Toxic Organic Compounds Using Daphnia 2,Magna Transcriptomic Profiles [M. Gustavsson, Patrik Svedberg, Juan S. Inda-Díaz, Sverker Molander, Jessica Coria, Thomas Backhaus, Erik Kristiansson] | 2023 | Fast and cost-efficient alternative to in vivo toxicity testing. | Slow adaptation in regulatory settings due to low accuracy and narrow applicability domains. |
| 2 | Machine Learning Models for Identification and Prediction of Toxic Organic Compounds Using Daphnia magna Transcriptomic Profiles [Tae-June Choi, Hyung-Eun An, Chang Bae Kim] | 2022 | The combination of feature selection and random forest classification algorithm showed high accuracy. | Misclassification of similar compounds can occur.  Validation using a novel dataset and larger datasets are needed. |
| 3 | Modeling of Aquatic Toxicity of a Set of Phenols in Silico [Khadidja Amirat, N. Ziani, Souhaila Meneceur, Fatiha Mebarki, Abderrhmane Bouafia] | 2023 | The model has a high coefficient of determination (R2) value of 79.24%, indicating a strong correlation between the variables. | The model relies on the use of multiple linear regression, that can’t get complex relationships |
| 4 | In silico prediction of chemical aquatic toxicity by multiple machine learning and deep learning approaches.  [Minjie, Xu., Hongbin, Yang., Guixia, Liu., Yun-Tao, Tang., Weihua, Li.] | 2022 | In silico models can prioritize compounds for testing, reducing costs and time. | Limitations in applicability domain and lower accuracy for some compounds. |
| 5 | Large-Scale Modeling of Multispecies Acute Toxicity Endpoints Using Consensus of Multitask Deep Learning Methods.[Sankalp Jain, Vishal B. Siramshetty, Vinicius M. Alves, Eugene N. Muratov,] | 2021 | Development of computational methods for reliable toxicity predictions  Creation of the largest publicly available dataset for acute systemic toxicity | Consensus models based on multitask learning approaches were developed.  These models showed significantly better performance in predicting acute toxicity endpoints. |
| 6 | In silico estimation of chemical aquatic toxicity on crustaceans using chemical category methods [Qianqian Cao, Lin Liu, Hongbin Yang, Yingchun Cai, Weihua Li, Guixia Liu, Philip W. Lee, Yun Tang] | 2018 | In silico models provide a priori evaluation of chemical aquatic toxicity.  Structural alerts identified for mechanistic interpretation of toxicity. | The study used machine learning methods and molecular fingerprints to develop models |
| 7 | Comparison of seven in silico tools for evaluating of daphnia and fish acute toxicity: case study on Chinese Priority Controlled Chemicals and new chemicals [Mainak Chatterjee, Kaushik Roy] | 2022 | High accuracy of VEGA in predicting daphnia and fish acute toxicity. | Read Across and Trent Analysis have lowest performance among tested tools. |
| 8 | Consensus QSAR models estimating acute toxicity to aquatic organisms from different trophic levels: algae, Daphnia and fish.[F. Lunghini, F. Lunghini, Gilles Marcou, P. Azam, Marie-Hélène Enrici,] | 2020 | Reasonable predictive performances achieved, models have good accuracy and data coverage. | Industrial data performances lower than public data, existing models fail to meet industrial needs. |
| 9 | Comparison of seven in silico tools for evaluating of daphnia and fish acute toxicity: case study on Chinese Priority Controlled Chemicals and new chemicals. [Zhou, Linjun, et al] | 2021 | Comprehensive comparison of seven in silico tools for evaluating acute toxicity in aquatic organisms. | Limited generalizability due to focus on Chinese Priority Controlled Chemicals and new chemicals. |
| 10 | Ecotoxicological assessment of pharmaceuticals and personal care products using predictive toxicology approaches. [Kar, Supratik, et al] | 2020 | Provides a comprehensive evaluation of pharmaceuticals and personal care products using predictive toxicology methods. | Potential limitations in the applicability and accuracy of predictive toxicology approaches. |
| 11 | Chronic aquatic toxicity assessment of diverse chemicals on Daphnia magna using QSAR and chemical read-across. [Kumar, Ankur, et al.] | 2024 | Utilizes QSAR and chemical read-across for chronic aquatic toxicity assessment on Daphnia magna. | Reliance on predictive models may lack real-world validation, potential limitations in applicability to all chemicals. |
| 12 | In Silico Prediction of the Toxicity of Nitroaromatic Compounds: Application of Ensemble Learning QSAR Approach[Amirreza Daghighi, Gerardo M. Casañola-Martin, Troy Timmerman,] | 2022 | Improved performance of toxicity prediction, high levels of goodness-of-fit, robustness, and predictivity. | The model uses multiple linear regression, so it can’t get complex relationships. |
| 13 | In Silico Predictability of Toxicity Parameters Using the OECD QSAR Toolbox of Some Components of Cannabis sativa[Victor H. Vázquez-Valadez, María Virginia Oliva‐Arellano, Pablo A. Martínez-Soriano, Manuel Alejandro Hernández‐Serda] | 2023 | Can calculate acute aquatic toxicity parameters for a wide range of organic compounds. | Not applicable for inorganic substances, organometallic compounds,polymers,chemicals containing metal ions. |
| 14 | Transformers enable accurate prediction of acute and chronic chemical toxicity in aquatic organisms. [Gustavsson, Mikael, et al] | 2024 | Transformers facilitate precise prediction of acute and chronic chemical toxicity in aquatic organisms. | Potential limitations in scalability or applicability to other domains may exist. |

According to the combined literature review of the articles that were submitted, there should be a strong emphasis on creating and evaluating predictive models to gauge the acute aquatic toxicity of organic compounds. A range of methodologies, such as QSAR modeling, in silico tools, computational techniques, and machine learning approaches, have been investigated for the purpose of forecasting toxicity values for various aquatic organisms, including fish,algae,Daphnia, and Vibrio fischeri. Considerable work has gone into assessing these models' performance, taking into account variables including regulatory acceptability, applicability domain, and forecast accuracy.

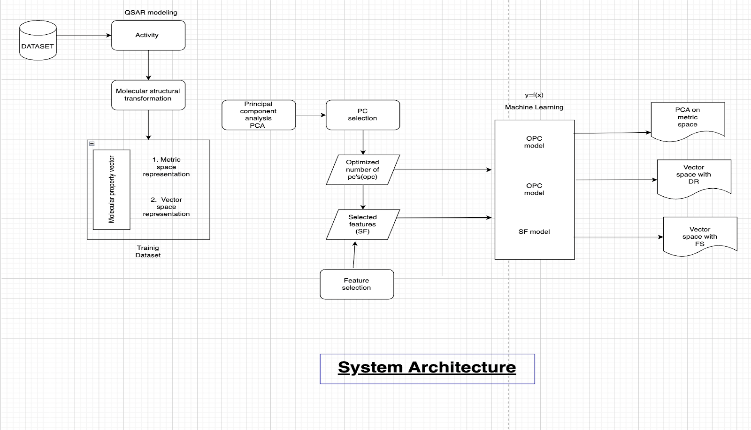


In order to guarantee accurate predictions, a number of researchers stress the significance of thorough validation and open reporting in predictive modeling techniques. Furthermore, using predictive toxicology techniques, attempts have been made,to evaluate the toxicity of particular chemical groups, including pharmaceuticals, personal care items, nitroaromatic compounds, polycyclic aromatic hydrocarbons (PAHs), and polychlorinated naphthalenes (PCNs). In order to prevent overestimating model performance, cross-validation techniques have been used, especially for predicting the aquatic toxicity of chemical combinations.

Additionally, improvements in modeling methodologies such as consensus modeling, ensemble learning, and transformer application have demonstrated potential to enhance the precision and resilience of toxicity forecasts. By offering trustworthy instruments for projecting the potentially detrimental impacts of chemicals on aquatic ecosystems, these advancements seek to solve issues in environmental risk assessment and regulatory decision.

1. PROPOSED SYSTEM

*System Architecture*



**Fig 1. System Architecture of the model**

* The focus of QSAR Modeling is centered primarily on the concept of Quantitative Structure-Activity Relationship (QSAR) modeling, which aims to establish a strong and insightful connection between the intricate molecular structures and the various biological properties they exhibit. This modeling approach plays a vital role in enhancing our comprehension of how alterations in molecular structure can directly influence and impact the biological activity observed in different compounds, paving the way for significant advancements in the field of molecular biology.
* When delving into Molecular Structure Transformation, the in-depth discussion revolves around the meticulous process of converting highly complex molecular structures into a series of numerical descriptors that can be easily interpreted and analyzed. This transformation process serves as a crucial facilitator in enabling the effective application of advanced machine learning techniques for comprehensive analysis and prediction purposes. By transforming molecules into numerical representations, researchers are empowered to harness the capabilities of computational algorithms to delve deeper into the realm of biological properties and make informed predictions with enhanced accuracy.

* The utilization of Machine Learning Techniques, with a particular emphasis on the prominent Principal Component Analysis (PCA), is underscored as a pivotal aspect of QSAR modeling. PCA plays a crucial role in feature selection, a fundamental step in the QSAR modeling process, by aiding in the reduction of dataset dimensionality while preserving the utmost

* relevant information. This reduction in dimensionality not only simplifies the intricate analysis procedures but also significantly boosts computational efficiency, thereby streamlining the entire modeling process for enhanced outcomes and insightful results.

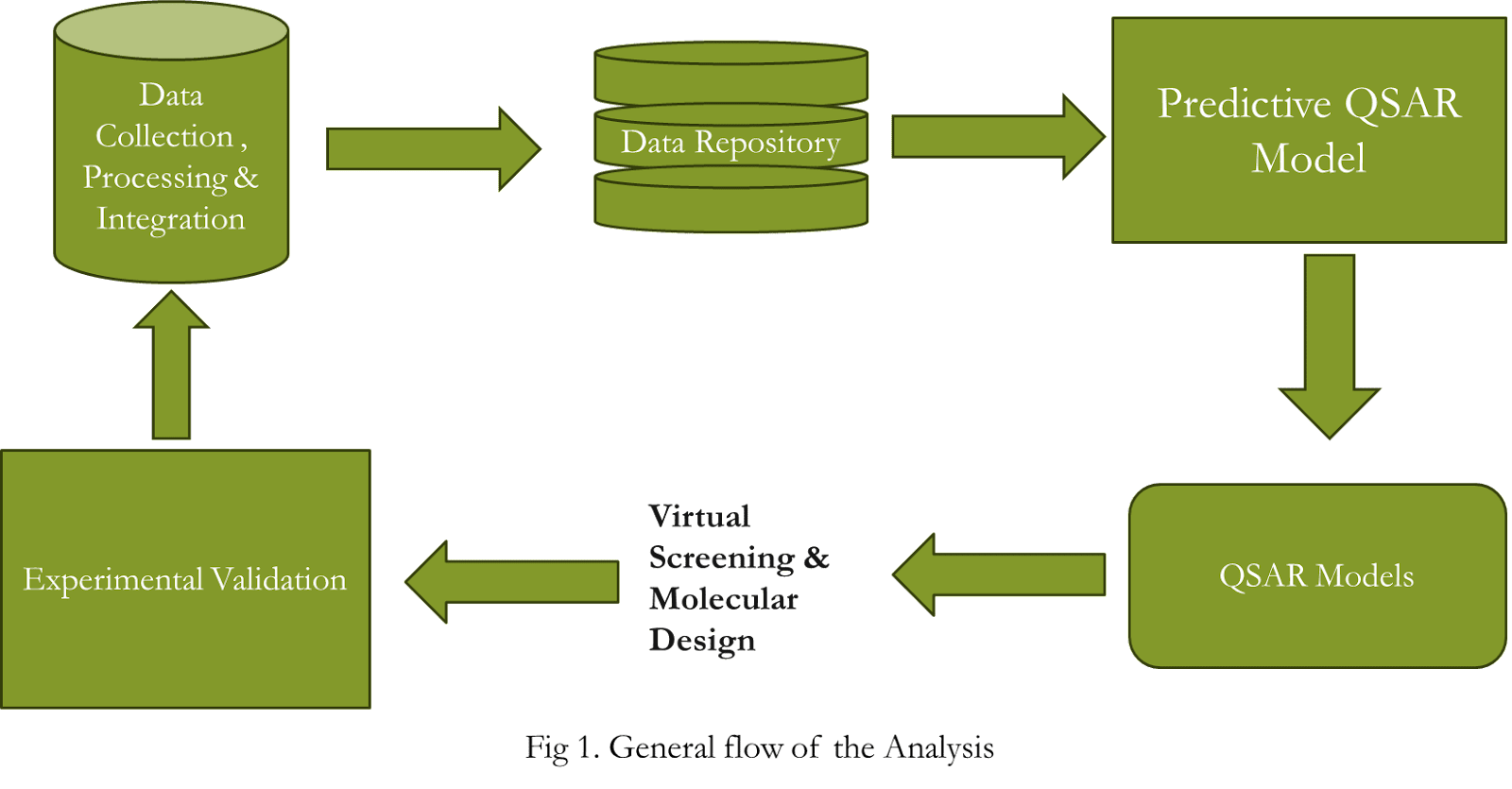
* A significant focus is placed on Optimized Spaces and Vectors, highlighting the importance of fine-tuning feature spaces and representing compounds in the form of vectors for enhanced clarity and efficiency. Through the optimization of feature spaces, researchers strive to craft a more streamlined and effective representation of molecular characteristics, ultimately leading to an exponential enhancement in computational efficiency and an elevation in the overall accuracy of QSAR models. Furthermore, representing compounds as vectors paves the way for the seamless application of mathematical operations and advanced machine learning algorithms to extract invaluable insights and predictions.
  + The indispensable concept of Molecular Property Vectors is deemed essential in the realm of QSAR modeling as it serves as the cornerstone for input data. These property vectors act as numerical representations of vital molecular properties such as size, shape, and chemical composition, enabling researchers to systematically quantify and analyze these properties for the accurate prediction of biological activity. This systematic approach proves to be instrumental in the realm of drug discovery and development, offering invaluable insights that fuel advancements in the field of molecular biology.

* + In essence, the content underscores the paramount significance of QSAR modeling in unraveling the complexities of molecular interactions and predicting biological activity with precision. By transforming intricate molecular structures into numerical descriptors and harnessing the power of advanced machine learning techniques, researchers are equipped to fine-tune feature spaces, represent compounds as vectors, and cultivate highly efficient QSAR models. The seamless integration of key system architecture components guarantees the seamless development and utilization of QSAR models in the realm of drug discovery and various

*Workflow Diagram*

1. *Quantitative Structure-Activity Relationship (QSAR) Modeling Concept:*

Our system architecture is based on Quantitative Structure-Activity Relationship (QSAR) modeling, which highlights the significance of developing a strong relationship between molecular structures and the biological traits they display. This method is essential for expanding our knowledge of how changes in molecular structure directly affect biological activity, which in turn propels tremendous advancements in molecular biology.



1. *Molecular Structure Transformation:*

Molecular Structure Transformation is a critical step in our system architecture, involving the conversion of complex molecular structures into numerical descriptors. This transformation enables the effective utilization of machine learning techniques for comprehensive analysis and prediction. By transforming molecules into numerical representations, researchers can delve deeper into the realm of biological properties and make informed predictions with enhanced accuracy.

It is a flowchart that illustrates the main steps involved in QSAR (Quantitative Structure-Activity Relationship)analysis forecasting. Its component elements include data collection, processing, integration, validation via virtual experiments, screening, molecular QSAR models, and design. The picture most likely depicts the gathering, processing, and use of data to build prediction models in the context of QSAR. It aims to provide an overview of the stages involved in the anticipated analytical process for QSAR modeling.

1.An illustration showing the steps involved in data analysis, such as gathering, cleaning, analyzing, and visualizing the data.

 2.a picture showing the several stages of data analysis, from gathering data to visualizing it.

3.An illustration of the data analysis procedure that shows the phases of data cleansing, analysis, and visualization.

Fig2. WorkFlow of Figure

****

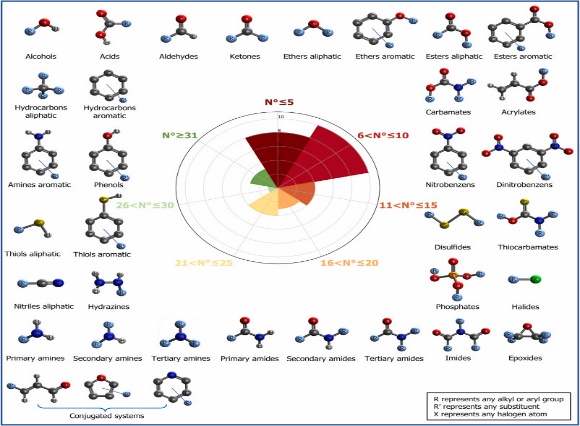


Fig3. Chemical Structures

*C. Utilization of Machine Learning Techniques:*

Our design uses Principal Component Analysis (PCA) as a fundamental component and integrates machine learning techniques. PCA reduces dataset dimensionality while maintaining pertinent information, which helps in feature selection. This decrease in dimensionality improves computational performance and streamlines analysis processes, improving modeling as a whole.

*D. Optimized Spaces and Vectors:*

Our architecture places a strong emphasis on optimized spaces and vectors to increase model accuracy and computational performance. Researchers can improve the efficiency and insight of their QSAR modeling procedures by describing chemicals as vectors and fine-tuning feature spaces.

*E. Molecular Property Vectors:*

Important molecule features including size, shape, and composition are represented by molecule Property Vectors, which are the basic input data used in our architecture. These numerical representations make it possible to analyze and quantify chemical properties in a systematic way, which makes it easier to anticipate biological activity. All things considered, our system architecture emphasizes how crucial QSAR modeling is to understanding molecular interactions and accurately forecasting biological activity.

1. DATASET

Aiming to characterize the chemical properties of organic substances and their associated quantitative response, LC50, which evaluates acute aquatic toxicity, the collection consists of molecular descriptors. The descriptors are as follows: TPSA(Tot) is the topological polar surface area, which indicates the polarity of the molecule; SAacc is the surface area corresponding to the acceptor atoms in hydrogen bonds; and H-050 is the number of hydrogen atoms bound to heteroatoms. The logP, a measure of lipophilicity, is calculated using MLOGP, whereas RDCHI offers information on molecule size and branching.

nN represents the number of nitrogen atoms, C-040 represents carbon atoms bound to electronegative atoms, and GATS1p represents molecular polarizability. When taken as a whole, these descriptors provide thorough insights into molecular structure and characteristics, which are essential for comprehending toxicity and chemical behavior. A measure of acute aquatic toxicity, LC50 is a quantitative response variable that shows the dose that is deadly to 50% of test organisms.

Predictive models can be created to estimate the toxicity of organic compounds by establishing a correlation between these characteristics and LC50 values. This will help with risk assessment and regulatory decision-making in environmental management. Thus, the dataset advances predictive toxicology and the sustainable management of chemicals in aquatic ecosystems by making it easier to apply quantitative structure-activity relationship (QSAR) modeling approaches to forecast acute aquatic toxicity.

|  |  |  |
| --- | --- | --- |
| **SNo.** | **Descriptor** | **Description of the molecular descriptor** |
| **1.** | **TPSA(Tot)** | is the topological polar surface area calculated using a contribution method that takes N, O, P, and S into account. |
| **2.** | **SAacc** | demonstrates the reason why hydrogen bond acceptor atoms have a larger Van der Waals surface area (VSA). |
| **3.** | **H-050** | represents the number of hydrogen atoms bonded to heteroatoms. |
| **4.** | **MLOGP** | is calculated to provide the octanol–water partition coefficient (LogP) using the Moriguchi model. |
| **5.** | **RDCHI** | is a topological index that, without taking heteroatoms into account, provides information on molecule size and branching. |
| **6.** | **GATS1p** | encodes information on molecular polarisability. |
| **7.** | **nN** | is the number of nitrogen atoms present in the molecule. |
| **8.** | **C-040** | Any electronegative atom, including halogens, O, N, S, P, and Se, is represented by the letter X. The formula for the number of carbon atoms is R–C(=X)–X / R–C#X / X=C=X. |
| **9.** | **Quantitative response[LC-50]** | LC50 measures concentration lethal to 50% test organisms; lower values indicate higher toxicity in toxicology. |

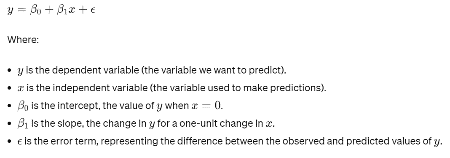
**Table 1. Molecular Descriptors in  the dataset**

1. METHODOLOGY
2. *Preprocessing Techniques:*

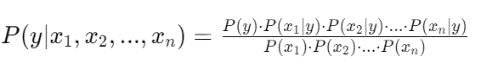
Prior to analysis and modeling jobs, data preprocessing is crucial for guaranteeing the quality and integrity of datasets. Preprocessing is important since it includes managing NaN (Not a Number) values, which are frequently seen in datasets. There are a number of methods for handling missing data, including as removing rows or columns that contain NaN values or using imputation techniques to replace missing values with statistical measures like mean, median, or mode. Time series data can be filled with the next available non-null value or the last known non-null value by using forward or backward fill techniques. Furthermore, one of the most important preprocessing steps for improving the quality and dependability of data is outlier removal. Data points that significantly vary from the majority of observations and have the potential to skew statistical studies and machine learning algorithms. To ensure the validity and dependability of data analysis and modeling findings, outliers must be identified and removed using statistical approaches, visual examination, domain expertise, and suitable outlier removal procedures.

1. *Machine Learning Techniques:*

One of the fundamental machine learning algorithms for predictive modeling is linear regression, which is particularly useful for determining the correlations between a dependent variable and one or more independent variables. To determine the best-fitting straight line, it minimizes the sum of squared discrepancies between observed and forecasted values. One independent variable in a simple linear regression might have several independent variables in a multiplex linear regression.

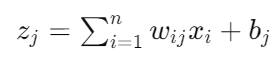
---------(1)

The Naive Bayes algorithm for classification is a straightforward yet efficient method that utilizes the Bayes theorem and feature independence. Naive Bayes is surprisingly effective despite its simplicity, especially when it comes to spam filtering and text classification.

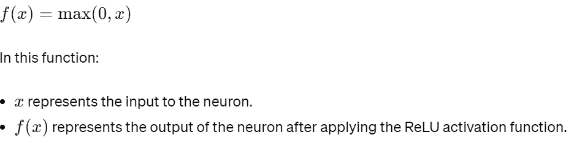
--------(2)

1. *Deep Learning Techniques:*

In order to learn intricate patterns and representations from data, deep learning entails training neural networks with numerous layers. Interconnected nodes, or neurons, arranged into input, hidden, and output layers make up neural networks. Via activation functions and weighted connections, every neuron process incoming information. In a number of domains, including speech recognition, computer vision, and natural language processing, deep learning has produced remarkable results. Backpropagation is one technique that minimizes the discrepancy between expected and actual outputs by adjusting connection weights during training.

-------(3)

A popular activation function in deep learning models is the Rectified Linear Unit (ReLU), which offers non-linearity to identify complex patterns in data. Because ReLU is piecewise-linear, gradient-based optimization strategies like gradient descent can be used effectively.

---------(4)

1. RESULT AND ANALYSIS
2. *Confusion Matrix:*

When assessing classification models, such as those used in Support Vector Machine (SVM) toxicity prediction, the confusion matrix is essential. The model predictions are classified as True Positives, which are cases that are correctly identified as toxic, False Positives, which are cases that are wrongly labeled as non-toxic, True Negatives, which are cases that are correctly identified as non-toxic, and False Negatives, which are cases that are misdiagnosed as toxic. When calculating overall prediction accuracy, accuracy takes into account true positives and true negatives in relation to the total number of instances. The F1 score illustrates the harmony of precision, recall, and specificity, which provide insights into prediction accuracy and balance.

1. *ROC Curve:*

The trade-off between sensitivity (true positive rate) and specificity (false positive rate) at various classification thresholds is represented by the ROC curve, which is an essential tool for assessing classification models, like the ones used in SVM for acute aquatic toxicity prediction. The total performance of the model is measured by the ROC AUC, where a number nearer 1 denotes superior discrimination and a value of 0.5 denotes performance comparable to random guessing. It's quite helpful when evaluating datasets with unequal distributions of classes.

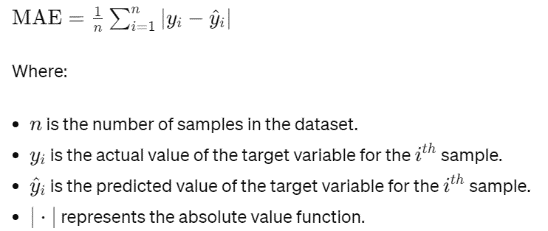
1. *R Squared:*

The percentage of the dependent variable's variation that can be predicted from the independent variable(s) is expressed as R Squared (R²). On a scale of 0 to 1, 1 denotes an ideal fit. Regression analysis frequently uses R2 to evaluate the performance of the model.

-----------(5)

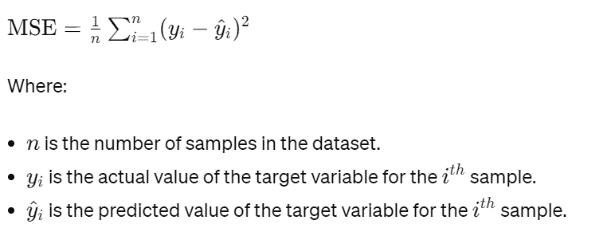
1. *Mean Absolute Error (MAE):*

The average absolute difference (MAE) between the actual and predicted values provides information on the correctness of the model in regression tasks. Better model performance is indicated by a lower MAE; this model is especially appropriate when outliers are present or when equal error weighting is needed.

**---------(6)**

1. *Mean Squared Error (MSE):*

Large errors are penalized more severely with the MSE, which calculates the average squared difference between actual and projected values. Even though MSE is frequently employed, it may not adequately represent model performance when there are outliers. In these situations, it is crucial to take into account alternate metrics as MAE or Root Mean Squared Error (RMSE).

--------(7)

1. CONCLUSION

Analyzing common compounds' possible distinctive influence requires a strong focus on their oceanic toxicity. The main goal of the suggested framework is to evaluate the effectiveness of similarity-based machine learning techniques in predicting the actually harmful nature of structured characteristic compounds at sea. The results of the study are expected to emphasize the importance of degree in selecting strongly harmfulness, lipophilicity, electrophilic reactivity, and atomic polarizability. The comprehensive outline support will ensure that the methodology is a basic tool for assessing harmful quality in novel or unproven compounds. If the offered models fall inside the fitting space, a comparative analysis of comparative factors reveals that they could replace creature testing in the assessment of the chemical toxicity of untested compounds.

REFERENCES

1. Melnikov, F., Kostal, J., Voutchkova-Kostal, A., Zimmerman, J. B., & Anastas, P. T. (2016). ‘Assessment of predictive models for estimating the acute aquatic toxicity of organic chemicals’. Green chemistry, 18(16), 4432-4445.
2. Netzeva, T. I., Pavan, M., & Worth, A. P. (2008). Review of (quantitative) structure–activity relationships for acute aquatic toxicity. QSAR & Combinatorial Science, 27(1), 77-90.
3. Zhou, L., Fan, D., Yin, W., Gu, W., Wang, Z., Liu, J., ... & Ji, G. (2021). Comparison of seven in silico tools for evaluating of daphnia and fish acute toxicity: case study on Chinese Priority Controlled Chemicals and new chemicals. BMC bioinformatics, 22, 1-31.
4. Voigt, M., & Jaeger, M. (2024). In silico and in vivo ecotoxicity—QSAR-based predictions and experimental assays for the aquatic environment. QSAR in Safety Evaluation and Risk Assessment, 495-509.
5. Li, F., Sun, G., Fan, T., Zhang, N., Zhao, L., Zhong, R., & Peng, Y. (2023). Ecotoxicological QSAR modeling of the acute toxicity of fused and non-fused polycyclic aromatic hydrocarbons (FNFPAHs) against two aquatic organisms: Consensus modeling and comparison with ECOSAR. Aquatic Toxicology, 255, 106393.
6. Nath, A., Ojha, P. K., & Roy, K. (2023). Computational modeling of aquatic toxicity of polychlorinated naphthalenes (PCNs) employing 2D-QSAR and chemical read-across. Aquatic Toxicology, 257, 106429.
7. Seth, A., & Roy, K. (2020). QSAR modeling of algal low level toxicity values of different phenol and aniline derivatives using 2D descriptors. Aquatic Toxicology, 228, 105627.
8. [Fei-xiong, C. H. E. N. G., et al. "In silico prediction of terrestrial and aquatic toxicities for organic chemicals." Chinese Journal of Pesticide Science 12.4 (2010): 477-488.
9. Zhou, Linjun, et al. "Comparison of seven in silico tools for evaluating of daphnia and fish acute toxicity: case study on Chinese Priority Controlled Chemicals and new chemicals." BMC bioinformatics 22 (2021): 1-31.
10. Amirreza, Daghighi., Gerardo, M., Casañola-Martin., Troy, Timmerman., Dejan, Milenković., Bono, Lučić., Bakhtiyor, Rasulev. (2022). In Silico Prediction of the Toxicity of Nitroaromatic Compounds: Application of Ensemble Learning QSAR Approach. Toxics, doi: 10.3390/toxics10120746
11. Khalidya, Khizbulaevna, Khamidulina., E., V., Tarasova., Mikhail, Leonidovich, Lastovetskiy. (2022). Application of the OECD QSAR Toolbox software for calculating the parameters of acute aquatic toxicity of chemicals. doi: 10.47470/0869-7922-2022-30-1-45-54
12. Kar, Supratik, et al. "Ecotoxicological assessment of pharmaceuticals and personal care products using predictive toxicology approaches." Green Chemistry 22.5 (2020): 1458-1516.
13. Kumar, Ankur, et al. "Chronic aquatic toxicity assessment of diverse chemicals on Daphnia magna using QSAR and chemical read-across." Regulatory Toxicology and Pharmacology (2024): 105572.
14. F., Lunghini., F., Lunghini., Gilles, Marcou., P., Azam., Marie-Hélène, Enrici., E., Van, Miert., Alexandre, Varnek. (2020). Consensus QSAR models estimating acute toxicity to aquatic organisms from different trophic levels: algae, Daphnia and fish.. Sar and Qsar in Environmental Research, doi: 10.1080/1062936X.2020.1797872
15. Mainak, Chatterjee., Kaushik, Roy. (2022). Application of cross-validation strategies to avoid overestimation of performance of 2D-QSAR models for the prediction of aquatic toxicity of chemical mixtures. doi: 10.1080/1062936X.2022.2081255
16. Shengnan, Zhang., Ning, Wang., Limin, Su., Xiaoyan, Xu., Chao, Li., Weichao, Qin., Yuanhui, Zhao. (2020). MOA-based linear and nonlinear QSAR models for predicting the toxicity of organic chemicals to Vibrio fischeri. Environmental Science and Pollution Research, doi: 10.1007/S11356-019-06681-Y
17. Gustavsson, Mikael, et al. "Transformers enable accurate prediction of acute and chronic chemical toxicity in aquatic organisms." Science Advances 10.10 (2024): eadk6669.
18. Noga, M., Michalska, A. & Jurowski, K. Application of toxicology in silico methods for prediction of acute toxicity (LD50) for Novichoks. Arch Toxicol 97, 1691–1700 (2023). https://doi.org/10.1007/s00204-023-03507-2