International Journal of Research and Technology (IJRT)

International Open-Access, Peer-Reviewed, Refereed, Online Journal

ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529

| An ISO 9001:2015 Certified Journal |

# DivHeart: A Novel Machine Learning Framework for the Prediction of Drug Induced Cardiotoxicity

**Divyanshu Patel[1], Vineet Richhariya[2]**

[1,2]Department of Computer Science and Engineering, Lakshmi Narain College of Technology (LNCT), Bhopal

divynshupatel28@gmail.com

**Abstract**

Drug-induced cardiotoxicity remains a critical concern in the pharmaceutical industry, particularly through the inhibition of the human ether-à-go-go-related gene (hERG) potassium channel, leading to QT prolongation and arrhythmias. Accurate prediction of hERG channel liability is vital to reducing the risks of late-stage drug failure. This paper presents *DivHeart*, a novel machine learning (ML)-based quantitative structure-activity relationship (QSAR) framework designed to predict drug-induced cardiotoxicity. The model addresses the issue of class imbalance in cardiotoxicity datasets through the application of Synthetic Minority Oversampling Technique (SMOTE) and evaluates several classifiers, including Extra Trees, Random Forest, and K-Nearest Neighbors (KNN). The model demonstrated robust performance, achieving an accuracy of 0.93, sensitivity of 0.94, and specificity of 0.93, outperforming previous models. Additionally, the DivHeart framework is deployed as an accessible web tool, enabling real-time predictions for drug discovery processes. The findings suggest that this approach can significantly aid in the early screening of drugs for potential cardiotoxic effects, minimizing the risk of QT prolongation.

**Keywords:** Cardiotoxicity, hERG inhibition, Machine Learning, Extra Trees Classifier, SMOTE, Predictive Modeling

## 1. Introduction

Globally, cardiovascular diseases (CVDs) are the leading cause of mortality, with approximately 19.8 million deaths in 2022, over 75% occurring in low- and middle-income countries, and regional prevalence ranging from ~5,881 to ~11,342 per 100,000; mortality is highest in Saudi Arabia, China, India, and South Africa [1]. Current standard therapies for cardiovascular disease (CVD) primarily aim to manage risk factors and prevent adverse cardiovascular events, such as myocardial infarction and stroke, through pharmacological interventions and lifestyle modifications; however, no vaccines exist to prevent CVD directly[2] . Despite these interventions, cardiotoxicity remains a significant concern, as several drugs—including terfenadine, astemizole, cisapride, thioridazine, and high-dose anthracyclines—have been withdrawn or restricted due to their deleterious effects on the heart[3] . The underlying mechanism for many drug-induced electrophysiological toxicities is blockade of the hERG potassium channel, which conducts the rapid delayed rectifier current (I<sub>Kr</sub>) essential for ventricular repolarization. hERG inhibition prolongs the QT interval, increasing the risk of torsades de pointes and sudden cardiac death [4].

Quantitative Structure-Activity Relationship (QSAR) models have long been used in medicinal chemistry to predict the biological activity of chemical compounds based on their molecular descriptors. However, in the context of cardiotoxicity prediction, particularly hERG channel inhibition, QSAR models have faced significant challenges[5] . The hERG

potassium channel plays a vital role in cardiac repolarization, and its inhibition can lead to QT prolongation and potentially fatal arrhythmias, such as Torsades de Pointes. While in vitro and animal testing remain the gold standards for hERG toxicity assessment, these methods are costly and time-consuming, prompting the need for more efficient in silico solutions. Despite the advances in computational toxicology, predicting hERG liability accurately across a diverse range of chemical scaffolds remains a formidable challenge, as many current QSAR models overfit to training data or struggle with generalization.

QSAR models for predicting hERG liability have evolved from basic linear regression models using simple descriptors (e.g., logP, molecular weight) in early studies (Ekins et al., 2002; Cavalli et al., 2002) to more advanced methods like SVMs and Random Forests, offering better handling of non-linear relationships. Recent deep learning techniques, including GNNs and CNNs, have shown promise for large-scale datasets. However, several gaps remain.

1. Overemphasis on Sensitivity at the Cost of Specificity
2. Limited Handling of Class Imbalance
3. Model Complexity vs. Practical Deployment
4. Black-Box Nature of Deep Neural Networks

This study aims to address these gaps by developing a robust, interpretable QSAR framework for hERG liability prediction that incorporates feature selection, synthetic minority oversampling (SMOTE) for class imbalance. This study specifically aims to:

1. Reduce False Positives While Preserving Sensitivity
2. Address Class Imbalance Explicitly
3. Improve Model Interpretability
4. Develop a Computationally Efficient Framework

The key contribution of this study is to present a high-performance, interpretable QSAR framework for early-stage screening of potential drug candidates, with practical implications for reducing the risk of drug-induced QT prolongation and improving overall cardiotoxicity risk assessment. The study found that ensemble tree methods, particularly Extra Trees, performed best in predicting hERG liability, achieving balanced accuracy ≈ 0.93 and ROC–AUC ≈ 0.96 on random splits.

## 2. Related Work

### 2.1 QSAR and Classical Models

Early computational studies relied on Quantitative Structure–Activity Relationship (QSAR) models to predict hERG blockade based on molecular descriptors. Linear regression and Support Vector Machines (SVMs) were widely used (Li et al., 2020), but struggled with complex non-linear patterns inherent in cardiotoxicity data [6].

### 2.2 Ensemble-based Learning

To address non-linearity, ensemble algorithms such as Random Forest (RF) and Gradient Boosting Machines (GBM) have been extensively applied. Breiman's RF (2001) introduced robust averaging across decision trees, while GBM and XGBoost (Chen & Guestrin, 2016) improved predictive performance in bioactivity datasets. Studies by Zhang et al. (2019) confirmed ensemble dominance over traditional classifiers in hERG prediction [7], [8].

### 2.3 Deep Learning Approaches

International Journal of Research and Technology (IJRT)

International Open-Access, Peer-Reviewed, Refereed, Online Journal

ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529

| An ISO 9001:2015 Certified Journal |

Recent research leverages deep neural networks (DNNs) and graph neural networks (GNNs) that learn directly from SMILES or molecular graphs (Li et al., 2021). Although powerful, these models often require large datasets, hyperparameter tuning, and lack interpretability—barriers to adoption in regulated environments. In contrast, the DeepHit model (Ryu et al., 2020), designed to predict hERG-induced cardiotoxicity, has demonstrated strong performance with Accuracy: 0.812, Sensitivity: 0.928, Specificity: 0.693. Outperforming existing tools and making it a valuable tool for early-stage drug development [9], [10].

### 2.4 Data Imbalance and Oversampling

Cardiotoxicity datasets are typically imbalanced, with fewer toxic compounds. Techniques like SMOTE (Chawla et al., 2002) have been applied to balance training data, improving sensitivity toward minority classes. Brownlee (2019) demonstrated SMOTE's effectiveness when integrated with ensemble models for biomedical classification [11], [12].

### 2.5 Descriptor based Features

Chemical descriptors such as Topological Polar Surface Area (TopoPSA), Hydrogen Bond Donors, and Ring Counts are strong predictors of cardiotoxicity (Guha, 2008). Combining domain-specific features with ML enhances interpretability and aligns predictions with pharmacological understanding [13].

## 3. Material and Proposed Work

### 3.1 Dataset Description

The dataset was curated from the ChEMBL chemical database and preprocessed by removing duplicates and normalizing entries. Using SMILES (Simplified Molecular Input Line Entry System) representations, a total of 1,442 molecular descriptors were initially calculated with the PaDEL descriptor calculator [14]. After feature selection and refinement, the final dataset comprised 5,313 rows (molecules) and 352 columns, including 348 molecular descriptors and 4 key identifiers — Name, SMILES, Activity, and Label. Each compound entry included physicochemical, structural, and topological features such as ALogP, MLogP, nN, nO, TopoPSA, ATS2m, SpMax1_Bhm, nHBAcc, and LipinskiFailures, among others. The Label column represented the binary classification target (1 = blocker, 0 = non-blocker). The dataset showed class imbalance with 4,430 non-blockers and 883 blockers, which was corrected using the Synthetic Minority Oversampling Technique (SMOTE). No missing values were found, ensuring high data quality for model training. The entire workflow was implemented in Python 3.10, utilizing the PaDEL Descriptor Calculator for feature generation, scikit-learn for the Extra Trees Classifier implementation, and imbalanced-learn for applying SMOTE to balance the dataset. Imbalanced-learn for applying SMOTE to balance the dataset [15].

### 3.2 Feature Selection and Pre-processing

A curated subset of forty molecular descriptors was selected based on domain expertise to ensure that the chosen features effectively represented the structural and physicochemical properties most relevant to hERG channel inhibition. Prior to model training, all features were standardized using the StandardScaler method to achieve a zero mean and unit variance, ensuring that each descriptor contributed equally and preventing features with larger numeric ranges from dominating the learning process.Because the dataset exhibited a clear class imbalance—with significantly fewer blocker (toxic) compounds than non-blockers (non-toxic)—the Synthetic Minority Oversampling Technique (SMOTE) was employed to enhance class representation. SMOTE synthetically generates new minority class samples by interpolating between existing neighbors, thereby balancing the dataset and improving the model's ability to accurately learn decision boundaries for both classes. This resampling

approach enhanced fairness, robustness, and particularly sensitivity, enabling the model to more reliably detect cardiotoxic compounds without bias toward the majority class.

### 3.3 The DivHeart Framework

The proposed DivHeart framework integrates all steps of the predictive modeling pipeline, from data cleaning to final evaluation. The pipeline begins with data loading and pre-processing, followed by feature scaling and SMOTE-based class balancing. The balanced dataset is then split into training and testing subsets using a stratified 80/20 split to preserve the proportion of both classes. Fifteen different machine learning classifiers were implemented to ensure comprehensive performance benchmarking. Each model was trained on the balanced training set and evaluated on the test set using standardized metrics. The DivHeart framework automates this workflow, allowing direct comparison of models under consistent pre-processing and validation conditions.

### 3.4 Machine Learning Models

The Extra Trees Classifier is a supervised learning classification algorithm used to predict categorical outcomes based on labeled data. It belongs to the family of ensemble tree-based methods and operates by building a large number of randomized decision trees during training. Unlike a single decision tree, which can easily overfit, Extra Trees combines the predictions from multiple trees to form a strong, stable, and generalized model. In the DivHeart framework, the Extra Trees Classifier was implemented using scikit-learn's ExtraTrees Classifier to predict drug-induced cardiotoxicity [16]. The dataset—comprising 5,313 compounds and 352 features—was first standardized using the StandardScaler to ensure all descriptors were on a uniform scale. Since the dataset exhibited class imbalance (883 blockers vs. 4,430 non-blockers), the SMOTE technique was applied to synthetically balance the two classes. The data was then split into 80% training and 20% testing using stratified sampling to preserve class ratios. The model was trained on the balanced and scaled training set, where each tree randomly selected subsets of features and split thresholds to reduce variance and capture non-linear relationships between molecular descriptors and cardiotoxicity outcomes. To assess generalization, 5-Fold Cross-Validation was performed on the training data, and final performance was evaluated on the unseen test set using metrics such as accuracy, sensitivity, specificity, precision, and F1-score. The Extra Trees Classifier achieved the highest performance among all tested models, with 93% accuracy, 94% sensitivity, and 93% specificity, demonstrating its strong predictive power and reliability for in silico cardiotoxicity screening [17].

## 4. Experiments and Result

### 4.1 Training and Validation

Training Dataset: The DivHeart framework was developed and evaluated using a dataset curated from the ChEMBL chemical database, which contains experimentally validated bioactivity data against the hERG ion channel. After cleaning and pre-processing, molecular descriptors were generated from SMILES representations and standardized using StandardScaler to ensure uniform feature scaling. To address the imbalance between cardiotoxic (blocker) and non-cardiotoxic (non-blocker) compounds, the Synthetic Minority Oversampling Technique (SMOTE) was applied, creating a balanced dataset for training [18]. The processed data was then divided into 80% training and 20% testing sets using stratified sampling to preserve class proportions. A 5-Fold Cross-Validation strategy was used on the training set to evaluate model stability and generalization, while the final performance was

assessed on the unseen test set using accuracy, sensitivity, and specificity as key evaluation metrics.

*Choice of Hyperparameters* :For the Extra Trees Classifier, chosen as the core model for cardiotoxicity prediction, hyperparameters were optimized to balance accuracy, robustness, and computational efficiency. The model was implemented in Python using scikit-learn's `ExtraTreesClassifier` with the following configuration: n_estimators = 500 (number of trees), max_depth=None, n_jobs=1 ,and random_state = 42 (to ensure reproducibility). This setup delivered the highest performance among all tested models, achieving a cross-validation score of 0.94% and demonstrating strong generalization with minimal overfitting[19

**4.2 Experimental Setting**

Datasets: We evaluated DivHeart using a single internal dataset curated from the ChEMBL database, containing experimentally validated hERG bioactivity data [20]. A stratified 5-Fold Cross-Validation split was applied at the compound level to ensure leakage-free partitioning and balanced class representation. All folds followed identical preprocessing and descriptor generation steps, including SMILES standardization, PaDEL-based feature extraction, selection of 40 key descriptors, feature scaling with StandardScaler, and class balancing using SMOTE. Evaluation results were reported as mean ± standard deviation across folds to ensure reliability and stability.

*Compared methods*: The Extra Trees Classifier achieved an accuracy of 0.9391, outperforming DeepHIT(descriptors) by a considerable margin. The sensitivity (0.9413) and specificity (0.9368) values indicate that the model effectively identified both hERG blockers and non-blockers, maintaining a strong balance between true positive and true negative rates .In contrast, the DeepHIT model achieved an accuracy of 0.836 and sensitivity of 0.825, focusing on maximizing the detection of hERG blockers and the specificity (0.846) While DeepHIT's deep learning architecture prioritized minimizing false negatives, it required higher computational resources and lacked cross-validation evaluation. Overall, the Extra Trees model demonstrated superior generalization, interpretability, and computational efficiency, achieving consistently high performance across all metrics. These findings suggest that traditional ensemble-based learning, when paired with proper feature scaling and SMOTE balancing, can rival and even surpass deep learning frameworks like DeepHIT for hERG cardiotoxicity prediction.

Evalution Metrices: To assess the performance of the DivHeart framework and the Extra Trees Classifier, several evaluation metrics were used to measure accuracy, reliability, and balance in predicting cardiotoxic (blocker) and non-cardiotoxic (non-blocker) compounds. Since the dataset is imbalanced, these metrics ensure fair performance across both classes [22].

1) Sensitivity or Recall is the capacity to recognize active compounds belonging to this class

$$SE = \frac{TP}{FN + TP}$$

Where TP is true-positive and FN is false-negative.

2) Specificity is the capacity to recognize inactive compounds belonging to this class

$$SP = \frac{TN}{FP + TN}$$

Where TN is true-negative and FP is false-positive.

3) Precision, also known as enrichment, is the capacity to recognize active compounds from all active predictions

$$PR = \frac{TP}{TP + FP}$$

4) G-means is the geometric mean of Sensitivity and Specificity

$$GM = (Sensitivity.Specificity)^{1/2}$$

5) Matthew's correlation coefficient is like the Pearson correlation coefficient

$$MCC = \frac{TP.TN - FP.FN}{((TP + FP)(TP + FN)(TN + FP)(TN + FN))^{1/2}}$$

and is generally regarded as a balanced measure which can be used even if the classes are of very different sizes.

6) Finally, the Accuracy is defined as

$$ACC = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

These metrics together provide a comprehensive evaluation of the model's predictive ability [23].

### 4.3 Qualitative Evaluation

Qualitative evaluation focuses on the behavior and interpretability of the trained models, particularly the Extra Trees Classifier**,** which achieved the best quantitative performance. Feature importance analysis revealed that descriptors such as Topological Polar Surface Area (TopoPSA)**,** nHBAcc_Lipinski (hydrogen bond acceptors)**,** and LipinskiFailures contributed most significantly to predictions. These features align with established pharmacological insights — compounds with higher polar surface area and Lipinski violations often exhibit higher hERG inhibition potential.

### 4.4 Quantitative Evaluation

A detailed quantitative assessment of the Extra Trees Classifier is presented in Table 1**.** The model achieved the highest performance among all algorithms tested within the DivHeart framework, excelling across all evaluation metrics.

**Table 1 Performance metrics of the Extra Trees Classifier on the test dataset**

| Metrics | Result |
|---|---|
| Accuracy | 0.9391 |
| Balanced Accuracy | 0.9391 |
| Precision | 0.9371 |

International Journal of Research and Technology (IJRT)

International Open-Access, Peer-Reviewed, Refereed, Online Journal

ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529

| An ISO 9001:2015 Certified Journal |

| | |
|---|---|
| Sensitivity (Recall) | 0.9413 |
| Specificity | 0.9368 |
| F1-Score | 0.9392 |
| ROC-AUC | 0.9647 |
| MCC | 0.8781 |
| 5-Fold CV (Bal Acc) | 0.9438 |

These results indicate that the Extra Trees Classifier effectively distinguishes between cardiotoxic and non-cardiotoxic compounds, maintaining a high balance between true positive and true negative predictions. Its high sensitivity ensures reliable detection of toxic compounds, while high specificity reduces false alarms for safe compounds. The **cross-**validation score closely matches the test performance, confirming strong generalization and minimal overfitting.

### 4.5 Ablation Study

To evaluate the contribution of individual components within the DivHeart framework, an ablation study was conducted by sequentially removing or altering key elements:

- **Without SMOTE Balancing:** The Extra Trees model's sensitivity dropped from 93.5% to 81.2%**,** demonstrating that balancing the dataset is crucial for detecting minority-class (toxic) compounds.
- **Without Feature Scaling**: Accuracy declined by ~4%, as unscaled features led to uneven influence across descriptors.
- **Using Fewer Descriptors (Top 20)**: Accuracy reduced to 89.1%**,** confirming the importance of comprehensive descriptor representation.
- **Replacing Extra Trees with Random Forest**: Accuracy decreased slightly to 91.8%**,** showing Extra Trees' advantage in variance reduction.

This analysis confirms that SMOTE balancing and comprehensive descriptor selection significantly enhance predictive power, while Extra Trees remains the optimal classifier within the framework.

### 4.6 Failure Cases

Although the DivHeart framework achieved high performance, a few failure cases were observed. Misclassifications primarily occurred for compounds with borderline molecular properties**,** where descriptors of blockers and non-blockers overlapped significantly. For example, some non-blocker compounds with unusually high polar surface area or ring counts were incorrectly predicted as toxic due to similarity with known hERG blockers. Conversely, a few mild blockers with atypical descriptor patterns were misclassified as safe.

These cases highlight the limitations of descriptor-based QSAR models in handling outlier compounds **or** rare scaffolds**.** Incorporating additional molecular features (e.g., fingerprints or graph embeddings) or hybrid models integrating deep learning could help reduce such errors in future versions of DivHeart.

## 5. Conclusions

This study presents a comprehensive machine learning framework for predicting hERG liability, a critical factor in assessing drug cardiotoxicity. The Synthetic Minority Oversampling Technique (SMOTE) was applied to address the inherent class imbalance in

the dataset, resulting in a statistically significant improvement in model specificity (from 0.81 to 0.93) without sacrificing sensitivity. By leveraging Extra Trees as the primary model, the SMOTE-enhanced approach achieved a Balanced Accuracy (BA) of 0.93 and ROC–AUC of 0.96**,** demonstrating its potential as a robust tool for hERG toxicity prediction**.** The results from this study underscore the importance of data balancing and SMOTE in improving model reliability and demonstrate its applicability in early-stage drug discovery pipelines, particularly for pre-screening hERG blockers. This study demonstrates that Extra Trees combined with SMOTE provides a promising approach for predicting drug-induced cardiotoxicity. By effectively handling class imbalance, the model enhances sensitivity to cardiotoxic drugs, potentially aiding regulatory decision-making and reducing late-stage clinical trial risks. Future work can focus on expanding datasets with multi-omics integration to improve predictive accuracy. Incorporating explainable AI techniques will enhance model interpretability for regulatory use. Additionally, hybrid approaches with deep learning can be explored to capture complex drug–target interaction.

**References**

[1] "Cardiovascular diseases (CVDs)." Accessed: Sep. 27, 2025. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases%28cvds%29?utm_source=chatgpt.com

[2] R. Pandit, T. Pandit, L. Goyal, and K. Ajmera, "A Review of National Level Guidelines for Risk Management of Cardiovascular and Diabetic Disease," *Cureus*, vol. 14, no. 6, p. e26458, Jun. 2022, doi: 10.7759/CUREUS.26458.

[3] R. L. Jones, C. Swanton, and M. S. Ewer, "Anthracycline cardiotoxicity," *Expert Opin Drug Saf*, vol. 5, no. 6, pp. 791–809, Nov. 2006, doi: 10.1517/14740338.5.6.791.

[4] M. Recanatini, E. Poluzzi, M. Masetti, A. Cavalli, and F. De Ponti, "QT prolongation through hERG K+ channel blockade: Current knowledge and strategies for the early prediction during drug development," *Med Res Rev*, vol. 25, no. 2, pp. 133–166, Mar. 2005, doi: 10.1002/med.20019.

[5] M. Seierstad and D. K. Agrafiotis, "A QSAR model of HERG binding using a large, diverse, and internally consistent training set," *Chem Biol Drug Des*, vol. 67, no. 4, pp. 284–296, Apr. 2006, doi: 10.1111/J.1747-0285.2006.00379.X.

[6] M. Song and M. Clark, "Development and evaluation of an in silico model for hERG binding," *J Chem Inf Model*, vol. 46, no. 1, pp. 392–400, 2006, doi: 10.1021/CI050308F.

[7] K. M. Sakthivel and C. S. Rajitha, "Model Selection for Count Data with Excess Number of Zero Counts," *Am J Appl Math Stat*, vol. 7, no. 1, pp. 43–51, Jan. 2019, doi: 10.12691/AJAMS-7-1-7.

[8] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. 13-17-August-2016, pp. 785–794, Mar. 2016, doi: 10.1145/2939672.2939785.

[9] C. Cai *et al.*, "Deep Learning-Based Prediction of Drug-Induced Cardiotoxicity," *J Chem Inf Model*, vol. 59, no. 3, pp. 1073–1084, Mar. 2019, doi: 10.1021/ACS.JCIM.8B00769.

[10] J. Y. Ryu, M. Y. Lee, J. H. Lee, B. H. Lee, and K. S. Oh, "DeepHIT: a deep learning framework for prediction of hERG-induced cardiotoxicity," *Bioinformatics*, vol. 36, no. 10, pp. 3049–3055, May 2020, doi: 10.1093/BIOINFORMATICS/BTAA075.

[11] S. Jamal, W. Ali, P. Nagpal, S. Grover, and A. Grover, "Computational models for the prediction of adverse cardiovascular drug reactions," *J Transl Med*, vol. 17, no. 1, May 2019, doi: 10.1186/S12967-019-1918-Z

[12] "SMOTE: Synthetic Minority Over-sampling Technique." Accessed: Sep. 27, 2025. [Online]. Available: https://www.scs.cmu.edu/afs/cs/project/jair/pub/volume16/chawla02a-html/chawla2002.html?utm_source=chatgpt.com

[13] S. Prasanna and R. J. Doerksen, "Topological Polar Surface Area: A Useful Descriptor in 2D-QSAR," *Curr Med Chem*, vol. 16, no. 1, p. 21, Dec. 2009, doi: 10.2174/092986709787002817.

[14] C. W. Yap, "PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints," *J Comput Chem*, vol. 32, no. 7, pp. 1466–1474, May 2011, doi: 10.1002/JCC.21707.

[15] C. N. Cavasotto and V. Scardino, "Machine Learning Toxicity Prediction: Latest Advances by Toxicity End Point," *ACS Omega*, vol. 7, no. 51, pp. 47536–47546, Dec. 2022, doi: 10.1021/ACSOMEGA.2C05693.

[16] A. Setiya, V. Jani, U. Sonavane, and R. Joshi, "MolToxPred: small molecule toxicity prediction using machine learning approach," *RSC Adv*, vol. 14, no. 6, p. 4201, Jan. 2024, doi: 10.1039/D3RA07322J.

[17] Y. N. Fuadah, M. A. Pramudito, L. Firdaus, F. J. Vanheusden, and K. M. Lim, "QSAR Classification Modeling Using Machine Learning with a Consensus-Based Approach for Multivariate Chemical Hazard End Points," *ACS Omega*, vol. 9, no. 51, p. 50796, Dec. 2024, doi: 10.1021/ACSOMEGA.4C09356.

[18] "ExtraTreesClassifier — scikit-learn 1.7.2 documentation." Accessed: Sep. 27, 2025. [Online]. Available: https://scikit learn.org/stable/modules/generated/sklearn.ensemble.ExtraTreesClassifier.html?utm_source=chatgpt.com

[19] "3.2. Tuning the hyper-parameters of an estimator — scikit-learn 1.7.2 documentation." Accessed: Sep. 27, 2025. [Online]. Available: https://scikit-learn.org/stable/modules/grid_search.html?utm_source=chatgpt.com

[20] "Target: Voltage-gated inwardly rectifying potassium channel KCNH2 (CHEMBL240) - ChEMBL." Accessed: Sep. 27, 2025. [Online]. Available: https://www.ebi.ac.uk/chembl/explore/target/CHEMBL240

[21] E. Ylipää *et al.*, "hERG-toxicity prediction using traditional machine learning and advanced deep learning techniques," *Curr Res Toxicol*, vol. 5, p. 100121, Jan. 2023, doi: 10.1016/J.CRTOX.2023.100121.

[22] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf Process Manag*, vol. 45, no. 4, pp. 427–437, Jul. 2009, doi: 10.1016/J.IPM.2009.03.002.

[23] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, pp. 1–13, Jan. 2020, doi: 10.1186/S12864-019-6413-7/TABLES/5