RESEARCH

Open Access



AttenhERG: a reliable and interpretable graph neural network framework for predicting hERG channel blockers

Tianbiao Yang^{1†}, Xiaoyu Ding^{1†}, Elizabeth McMichael², Frank W. Pun², Alex Aliper³, Feng Ren^{1*}, Alex Zhavoronkov^{1,2,3*} and Xiao Ding^{1*}

Abstract

Cardiotoxicity, particularly drug-induced arrhythmias, poses a significant challenge in drug development, highlighting the importance of early-stage prediction of human ether-a-go-go-related gene (hERG) toxicity. hERG encodes the pore-forming subunit of the cardiac potassium channel. Traditional methods are both costly and time-intensive, necessitating the development of computational approaches. In this study, we introduce AttenhERG, a novel graph neural network framework designed to predict hERG channel blockers reliably and interpretably. AttenhERG demonstrates improved performance compared to existing methods with an AUROC of 0.835, showcasing its efficacy in accurately predicting hERG activity across diverse datasets. Additionally, uncertainty evaluation analysis reveals the model's reliability, enhancing its utility in drug discovery and safety assessment. Case studies illustrate the practical application of AttenhERG in optimizing compounds for hERG toxicity, highlighting its potential in rational drug design.

Scientific contribution

AttenhERG is a breakthrough framework that significantly improves the interpretability and accuracy of predicting hERG channel blockers. By integrating uncertainty estimation, AttenhERG demonstrates superior reliability compared to benchmark models. Two case studies, involving APH1A and NMT1 inhibitors, further emphasize AttenhERG's practical application in compound optimization.

Keywords Cardiotoxicity, hERG prediction, Graph neural network, Attention mechanism

[†]Tianbiao Yang and Xiaoyu Ding have contributed equally to this work.

*Correspondence:
Feng Ren
feng.ren@insilico.ai
Alex Zhavoronkov
alex@insilico.ai
Xiao Ding
xiao.ding@insilico.ai
¹ Insilico Medicine Shanghai Ltd, Suite 901, Tower C, Changtai Plaza, 2889
Jinke Road, Pudong New District, Shanghai 201203, China
² Insilico Medicine Hong Kong Ltd, Unit 310, 3/F, Building 8W, Phase 2,
Hong Kong Science Park, Pak Shek Kok, Hong Kong, China
³ Insilico Medicine AI Limited, Level 6, Unit 08, Block A, IRENA HQ Building,
Masdar City, Abu Dhabi, UAE

Introduction

The adverse effects of pharmaceutical agents on the heart represent a significant challenge in drug development. Drug-induced arrhythmias are of particular concern, with severe consequences resulting in mortality, as observed in treatments such as dofetilide, haloperidol, and trovafloxacin [1, 2]. Among the numerous ion channels involved in cardiac repolarization, the human ether-a-go-go-related gene (hERG) potassium channel is pivotal in regulating cardiac action potential [3]. Accurate prediction of hERG toxicity in the early stages of drug development is critical for mitigating risks and ensuring the safety profile of emerging therapeutics [4].



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Recognizing the growing concerns regarding cardiac safety, regulatory bodies such as the International Conference on Harmonization of Technical Requirements for the Registration of Pharmaceuticals for Human Use (ICH) now mandate the evaluation of drug candidates' hERG channel blockage properties in preclinical stages [4, 5]. Traditional hERG inhibition detection methods like patch-clamp electrophysiology and in vivo QTc assays are hindered by their cost and time-intensive nature [6]. Therefore, computational methods have emerged as a promising avenue to enhance the efficiency of hERG channel blocker screening [7, 8]. Quantitative structure-activity relationship (QSAR) models were initially developed, offering interpretability by dissecting hERG channel blocker pharmacophore patterns [9-12]. However, these models often rely on small-sized training datasets, limiting their robustness for diverse hERG channel-blocking compounds [13]. Subsequently, various data-driven machine learning (ML)-based models emerged [14-20]. However, ML methods based on expert-defined molecular fingerprints and feature engineering approaches might be constrained by predefined rules. Predictive performance may decrease for novel compound scaffolds. Encoding the atoms of the compound and performing end-to-end prediction may help alleviate the limitations imposed by predefined rules.

Deep learning-based (DL) models have emerged as a novel method for predicting hERG channel inhibition [21], propelled by the remarkable success of deep neural networks (DNNs). Recent advancements have focused on integrating diverse molecular features and model consensuses to enhance prediction reliability and explanation. Certain standouts emphasize reliability, applicability, or interpretability among these models, including CardPred [22], DeepHIT [23], CardioTox [24], hERG-att [25], ADMETLAB 2.0 [26], BayeshERG [27], DMFGAM [28], Pred-hERG 5.0 [29], CToxPred [30], and CardioDPi [31]. Specifically, BayeshERG was developed via a graphbased Bayesian deep learning model and a directed message-passing neural network (D-MPNN). However, these models are frequently perceived as black boxes, yielding prediction outcomes that are difficult to fully interpret, along with a need to improve uncertainty estimation. Several research teams have utilized substructure-based methods to analyze chemistry-intuitive explanations, as observed in SME and OptADMET [32, 33]. Still, these models lack robustness due to the limited availability of hERG channel-blocking experimental data. We summarize the performance, interpretability, and availability of ML- / DL-based models in Table S1. We also summarize the rationale for uncertainty quantification implementation as a strategy to enhance model robustness (Table S1). Among these models, hERG-att, BayeshERG,

and the recently updated models, such as CardioDPi and Pre-hERG 5.0, provide interpretability. While BayeshERG also incorporates uncertainty quantification, both its uncertainty estimation and overall accuracy require considerable improvement. Therefore, we are developing a new approach aimed at addressing the three aspects of reliability, interpretability, and uncertainty quantification.

We developed AttenhERG, a novel graph neural network framework designed to reliably predict compound hERG channel blocking risk to address this unmet clinical need. Notably, we enhanced predictive performance and improved interpretability through structure optimization. Next, we thoroughly evaluated the model's predictive performance using internal and external test datasets, highlighting its efficacy in accurately predicting hERG activity. We demonstrate the model's robustness and reliability through meticulous optimization and uncertainty estimation. We also comprehensively evaluate our model analysis by comparing it with benchmark models, providing valuable insight into its technical advantage over existing tools. We conclude by presenting compelling case studies involving APH1A [34] and NMT1 inhibitors [35] to illustrate the practical utility of our approach in real-world scenarios.

Results

Model architecture and optimization

We began by utilizing the Attentive FP algorithm [36] in conjunction with uncertainty evaluation analysis to construct an interpretable and reliable deep learning model named AttenhERG for predicting hERG channel blockers (Fig. 1). The methods section provides specific details regarding this methodology.

We employed a systematic approach combining grid search and early stopping techniques during the hyperparameter tuning phase, conducted exclusively on the validation set. This approach enabled us to efficiently explore a range of hyperparameter configurations while safeguarding against potential overfitting. The parameters subjected to optimization included dropout rate, hidden layer units, learning rate, and L2 regularization (Table S2, Fig. 2A). We initially fixed the dropout and L2 regularization parameters to determine the appropriate number of hidden units. Subsequently, we separately optimized the regularization L2 rate and dropout rate. To mitigate overfitting, we implemented an early stopping strategy (see Methods), emphasizing the learning rate due to its critical impact on model training steps.

After the optimization stage, an optimal model configuration with a learning rate of $10^{-3.5}$, 200 hidden layer units, a dropout rate of 0.1, and an L2 regularization rate of $10^{-4.5}$ was identified. Next, we evaluated its performance characteristics and training dynamics in greater



Fig. 1 Overall architecture of the model



Fig. 2 Heat maps and model predictive performance for hyperparameter search. A Impact of different hyperparameters on AUROC on the validation set; B Performance metrics of the optimal model; C Loss and AUROC during the training process

depth. This analysis provided insights into the model's training progression and illustrated the evolution of loss and area under the receiver operating characteristic curve (AUROC) metrics on the validation and test sets (Fig. 2B). Notably, the validation set's AUROC metrics plateaued after 83 epochs, indicating peak model performance achievement. AttenhERG also exhibited commendable performance metrics, including an AUROC of 0.835, accuracy of 0.777, an area under precision-recall curve (AUPRC) of 0.834, Matthew's correlation coefficient (MCC) of 0.543, balanced accuracy (BAC) of 0.767, and F1 score of 0.812 (Table S3, Fig. 2C). These metrics underscore the model's efficacy in accurately predicting hERG activity.

Internal evaluation

The predictive performance of all models was conducted to measure MCC, BAC, and AUROC on the test dataset (Table S4). Among the models evaluated, AttenhERG displayed improved performance across the metrics, with an MCC of 0.543, BAC of 0.767, and AUROC of 0.835 (Fig. 3A). This advantage could be attributed to its dual-level attention mechanism, which first captures local features at the atomic level and subsequently incorporates global molecular features. In contrast, BayeshERG and D-MPNN models encode local features, deriving atomic embeddings from molecular structures. However, the influence between atoms significantly diminishes with increasing distance, while



Fig. 3 The evaluation of predictive performance results with the internal test. A The internal test of the model; B The internal test-strict of the model

long-range interactions, such as intramolecular hydrogen bonding, can still be impactful. Incorporating an attention mechanism at the atomic level in AttenhERG likely optimizes these aspects, resulting in enhanced performance compared to BayeshERG and D-MPNN. Additionally, SVM and RF models exhibited comparatively lower performance metrics, potentially due to limitations in expert-defined molecular fingerprints.

Interestingly, we also observed that SVM and RF method performance was improved relative D-MPNN. One possible reason for this advantage may be that expert-defined molecular fingerprints have specific predefined rules for molecular scaffolds and list all fragments of the compound. To test this hypothesis, we constructed a test-strict dataset comprising compounds with low scaffold similarity to the training dataset (see Methods for details). The predictive performance on the test-strict dataset revealed distinct differences among the models (Table S5, Fig. 3B). AttenhERG exhibited the highest performance, achieving an MCC of 0.492, BAC of 0.744, and AUROC of 0.818. In contrast, SVM (FPS) and RF (FPS) demonstrate relatively poorer performance, with lower MCC and AUROC than the other models. We conclude that predefined rules constrain ML methods relying on expert-defined molecular fingerprints, resulting in significant performance dips when encountering novel scaffolds. These findings highlight the role of model design, feature engineering, and attention mechanisms in improving predictive performance.

External evaluation

Next, we delved into a comprehensive evaluation of the predictive performance of our model across diverse external test sets to shed light on its efficacy in realworld scenarios. The external review utilizes identical evaluation metrics employed in the internal assessment. We present a detailed overview of the external evaluation results that showcase the predictive performance of our model compared to baseline models across four distinct external test sets (Tables S6-S9). Overall, our model demonstrates comparable performance to the baseline models across these test sets, underscoring its robustness in various scenarios. Our study focuses on two critical aspects: model interpretability and reliability. We selected models that performed strongly in these areas for comparative analysis for the in-house dataset, such as BayeshERG, CardioDPi, CToxPred, Pred-hERG 5.0 and DMFGAM (Table S10). Our model consistently ranks highly among the models evaluated across various metrics and datasets, showing overall superiority over recently updated models (Fig. 4A). Upon analyzing the performance differences, we found that many of the updated baseline models primarily rely on expert-defined fingerprints and descriptors for molecular representation, which could constrain their performance due to predefined rules. Notably, BayeshERG's predictive performance on the external test sets was influenced by the presence of duplicate data arising from molecular stereochemistry and ionization, which introduced a bias in its predictions, leading to overestimated prediction results.



Fig. 4 The evaluation of predictive performance results with the External Test Sets using AttenhERG, Baseline and Updated baseline (models updated after 2022). A The four external test sets of the model performance; B The in-house test sets of the model performance

To provide a more rigorous comparison of novel compounds, we selected an in-house dataset that is structurally distinct from the four external test sets (Figure S1A). AttenhERG significantly outperformed both BayeshERG and recently updated models, including CardioDPi, CToxPred, Pred-hERG 5.0 and DMFGAM in this analysis (Fig. 4B). This improvement was attributed to AttenhERG's ability to autonomously learn the chemical environment of atoms, thereby effectively identifying substructures that significantly impact hERG inhibition rather than just the entire molecule. Overall, the external evaluation provides compelling evidence of our model's efficacy, highlighting its ability to generalize across diverse, novel datasets.

Uncertainty evaluation analysis

Within deep learning models, uncertainty estimation has become a crucial component for assessing the authenticity of prediction outcomes. Specifically, the source of this uncertainty can be affected by both algorithmic and data-availability constraints. We delved into the impact of uncertainty estimation on model performance, particularly in scenarios with insufficient hERG data. We employed two uncertainty estimation techniques, Entropy and MC-Dropout, that are known for their efficacy in similar attribute prediction tasks [37] within the framework of the AttenhERG model. These methods capture predictive uncertainty in classification models without altering the model framework and are contrasted with two uncertainty estimation methods employed in the BayeshERG model. Additionally, we explored the linear relationship between uncertainty levels and prediction accuracy to enhance the model's reliability.

Despite being a probabilistic model, the uncertainty analysis in the BayeshERG model revealed no significant improvement in model performance. This observation is supported by the AUROC curve, which indicates that models referencing uncertainty estimates perform equivalently to the random group, showing no discernible change (Fig. 5A). Furthermore, introducing Entropy and MC-Dropout uncertainty methods into the AttenhERG model resulted in enhanced model performance compared to the random group, as evidenced by the MCC, BAC, and AUROC metrics. This demonstrates the superior performance of the model over the random group (Fig. 5B), validating the reliability of the AttenhERG model equipped with uncertainty estimation.

Case study

We employed AttenhERG's predictive ability in multiple case studies to assess its real-world utility. This analysis provided crucial insight into how structural modifications influence hERG activity across various circumstances. For instance, CHEMBL2021101 was identified as a potent γ -secretase modulator (Fig. 6A), demonstrating single-digit nanomolar Aβ42 (APH1A) IC₅₀ in cell-based assays. Despite its therapeutic potential, this compound possesses significant hERG toxicity risks. Optimization studies originating from aryl triazole leads were initiated, culminating in developing novel amides and lactams within the series [34]. These modifications significantly enhanced activity and reduced the compound's affinity towards hERG channels. AttenhERG forecasted the directional shifts in hERG binding proportions for these molecules and pinpointed crucial atoms and substructure fragments that may contribute to this modification. Atom attention weight visualization revealed deeper hues of red in the phenyl group of the initial structure, indicating segments likely pivotal to hERG properties. The introduction of trifluoromethyl substitution to the phenyl group mitigated its impact on hERG properties, with predictive results indicating that the optimized compounds exhibited hERG inhibitory activity above 10uM, in agreement with prior results [35].

In another case study concerning the Pyrazole Sulfonamide Series of Trypanosoma brucei N-Myristoyltransferase (NMT1) inhibitors, we scrutinized the structure-activity relationships of a novel series of pyrazole sulfonamide compounds to identify fragment modifications that result in hERG inhibition [35]. AttenhERG analyzed two representative compounds, CHEMBL3358114 and CHEMBL1230468, inhibition risk (Fig. 6B). Predictive for hERG analysis demonstrated alignment between the



Fig. 5 The evaluation of uncertainty evaluation analysis. A The uncertainty evaluation analysis of BayeshERG model and AttenhERG model; B The uncertainty evaluation analysis of the AttenhERG model



Fig. 6 Case Study: Optimization of hERG Toxicity for Various Compounds. A An illustration of optimizing a compound targeting APH1A to reduce hERG inhibitory risk; B Optimizing the hERG inhibitory activity of a compound targeting NMT1; C Visualization of AttenhERG's capability to autonomously learn molecular features

magnitude and directionality of hERG changes and experimental observations. We used the NMT1 inhibitor CHEMBL3358114 as an illustrative example. First, we provided a visualization of the model-derived weights to illustrate how molecular features are captured in our model. The AttenhERG model autonomously learns the chemical environment of atoms, utilizing hERG prediction as a supervisory task. The model assesses the atomic vectors' correlation, with negatively correlated atoms highlighted in yellow and positively correlated atoms in blue (Fig. 6C). The analysis indicates that the molecule exhibits distinct structural patterns, which are more pronounced in the deeper hidden layers. In the case of CHEMBL3358114's structure, atomic correlations are predominantly clustered in the C20-C29 tail region, indicating a significant impact of this region on the molecule's hERG inhibitory activity. Structural modifications in this region (Fig. 6B) reduced its hERG risk, consistent with experimental validation in which we measured reduced hERG affinity from 0.6 μ M to 28 μ M.

Overall, these case studies underscore the rationale behind deriving fragments from the AttenhERG model that significantly impact hERG and provide instruction for structural optimization, giving valuable insights into model-learned knowledge associated with hERG properties.

Discussion and conclusion

Here, we developed AttenhERG, a novel graph neural network framework tailored for predicting hERG channel blockers with enhanced reliability and interpretability. Inspired by the Attentive FP algorithm, our work introduces key innovations that significantly advance the field. Notably, AttenhERG integrates graph-based molecular representations, attentive encoding mechanisms, and uncertainty evaluation analysis. These advancements distinguish it from previous hERG prediction models, enhancing both predictive performance and model transparency. Our model demonstrates notable improvements in classification accuracy within multiple datasets, highlighting its efficacy in accurately predicting hERG activity across diverse molecular structures. Furthermore, uncertainty analysis reveals that excluding predictions with higher uncertainty enhances the model's performance, thereby bolstering prediction reliability. Our case studies illustrate AttenhERG's utility in optimizing compounds for hERG toxicity, showcasing its capability to effectively identify and modify atomic fragments to optimize hERG properties. The development of AttenhERG represents a substantial advancement in drug discovery and safety assessment methodology, offering a robust and interpretable model for early-stage prediction of hERG toxicity.

In early-stage drug discovery, uncertainty analysis results prioritize structures with low uncertainty scores, as these predictions are more reliable and associated with a reduced risk of failure. However, exploring effective strategies for addressing high-uncertainty predictions presents a valuable direction for future research. Compounds and their analogs with elevated uncertainty scores merit further investigation, as they may reveal underexplored regions of chemical space or complex features overlooked by existing models. We recommend conducting additional in vitro testing on these compounds to validate their biological activity. Compounds that consistently exhibit high uncertainty should be deprioritized in the development pipeline to optimize resource allocation toward more reliable candidates. Expanding validated compounds into the training dataset's chemical space and fine-tuning the model with this augmented dataset may enhance predictive performance, enabling uncertainty estimates to serve as critical insights that effectively guide earlystage drug discovery efforts.

Future investigations could explore to improve model generalization and the model wider scope of applicability. AttenhERG's model performance may also be limited by the dataset size. With limited datasets, the chemical space represented by the compounds is restricted, which can hinder predictive performance for novel structures. In addition to utilizing activity data from patch-clamp experiments, it would be beneficial to incorporate high-throughput data while remaining cautious of biases that such data may introduce. Up-sampling or down-sampling strategies could be applied when integrating large-scale high-throughput data to manage class imbalance. For tasks with limited sample sizes, meta-learning approaches could be considered in future work to enhance predictive performance. Moreover, this framework could be adapted for predicting drug metabolism, molecular carcinogenicity, aquatic toxicity, and drug-induced liver injury [38-41]. Furthermore, OCHEM offers a comprehensive collection of toxicity data, which could be combined with the AttenhERG framework to uncover potential relationships among different toxicity attributes [42]. Additionally, investigations into the transferability of AttenhERG to other ion channels or toxicity endpoints could expand its utility in drug safety assessment. Efforts to enhance the interpretability of the model's predictions by developing visualization techniques or feature attribution methods could provide deeper insights into the molecular mechanisms underlying hERG inhibition. Overall, AttenhERG represents a promising step towards more efficient and reliable prediction of hERG toxicity, with broad implications for drug discovery and safety evaluation.

Methods

Dataset construction

The datasets utilized for model construction encompass compound activity data from various public databases, including ChEMBL, PubChem, and BindingDB, along-side the compound activity data extracted from multiple scientific publications [27, 43–45]. The final datasets (Table S11) employed for training, validation, and testing the model comprised 14,322 molecules, including 8,488 positive and 5,834 negative instances, with a predefined activity threshold of 10 μ M. Within the testing dataset, we established a stringent subset called the test-strict dataset, wherein compounds with highly similar scaffolds (similarity>0.8) to those in the training set were excluded. Furthermore, the literature data can serve as external test sets, providing a more comprehensive evaluation of the model's performance.

The external test sets are entirely independent of the training, validation, and test datasets. The datasets for all mentioned experiments are summarized in Table S11. External Test Set 1 consisted of 30 positive and 14 negative instances [23]. External Test Set 2 comprised 157 FDA-approved drugs, with 41 positive and 116 negative instances [16]. External Test Set 3 contained 11 positive instances and 30 negative instances [24]. External Test Set 4 incorporated data from the thallium flux assay, featuring 34 positive and 706 negative instances [24]. Additionally, an in-house dataset included 143 positive instances and 90 negative instances. Integrating these diverse external datasets enables a comprehensive evaluation of the model's performance across various experimental conditions and assay methodologies, thereby enhancing the reliability and applicability of the model evaluation. The t-SNE distribution demonstrates the structural novelty of compounds in the in-house dataset, particularly in comparison to the external test sets (Figure S1A). Among the four external test sets, only External Test Set 2 and External Test Set 4 share 17 FDAapproved drugs, as illustrated in the Venn diagrams (Figure S1B). There is no other overlap among the external test sets. Importantly, we used identical, previously published dataset processing methods to ensure consistency and comparability [27]. To ensure the reliability of the inhouse dataset, we applied the StandardizeSmiles function from the RDKit library to standardize the SMILES representations of compounds. This rigorous standardization corrects for inconsistencies or irregularities present in the chemical structures of the compounds. To address bias introduced by stereochemistry and ionization [46], we employed RDKit to deduplicate molecules across the five external test sets using InChI Keys. In the training set, molecules sharing identical first-block InChI Keys were removed, and the training dataset was subsequently

updated. As summarized in Table S12, external test set 1 contained six duplicates, test set 2 had one, and test set 3 had seven. No duplicates were identified in external test set 4 or the inhouse set. After deduplication, the training set for external test sets 1–3 was revised, and three separate models were retrained and re-evaluated on their respective external test sets.

Model framework construction

The AttenhERG strategy encompasses three pivotal elements crucial for its efficacy: a well-trained interpretable deep learning model capable of delivering precise predictions and meaningful interpretations; a visualization protocol aimed at discerning atom-level features most pertinent to hERG inhibition activity; and an uncertainty scoring mechanism employed to gauge the robustness of the model.

We constructed an interpretable deep learning model leveraging the attention-based graph neural network (Attentive FP) algorithm [36]. This algorithm, constituting a specialized graph neural network architecture integrated with a graph attention mechanism, has been validated for its exceptional predictive performance across diverse datasets. Notably, the algorithm's capability extends to extracting non-local intra-molecular interactions and facilitating the visualization of knowledge acquired by the model. Initially, nine distinct atomic features and four bond feature types were computed to serve as node and edge features for each molecular graph. Specifically, the framework and molecular representations are summarized in Figure S2A. The input node features of the model include 39 descriptors, such as atom symbol, number of covalent bonds, electrical charge, radical electrons, hybridization, aromaticity, hydrogens, chirality, and chirality type. Furthermore, the input edge features account for bond type, conjugation, ring status, and stereochemistry (Figure S2B). Subsequently, employing a fully connected layer, initial vectors of uniform length were generated for each atom and its neighboring entities. Initial vectors underwent updates after aggregating additional neighborhood information in the subsequent two hidden layers embedding attention mechanisms. A novel state vector characterizing the entire molecule was derived by assembling state vectors for each atom, assigning attention weights to neighbors based on their respective contributions. Ultimately, a fully connected layer was harnessed for task training and prediction. The holistic network architecture underwent hyper-parameter tuning via grid search and gradient descent optimization employing the Adam optimizer. The continuous updating of attention weights throughout the model's iteration process is particularly significant. At each iteration, a novel model iteration was instantiated and subjected to performance evaluation using the validation dataset. To mitigate against overfitting and to select the definitive model exhibiting exemplary performance, an early stopping strategy predicated on evaluation outcomes from training and validation sets was instituted. Consequently, if no improvement in validation AUROC was observed after 20 epochs or validation loss remained stagnant after 30 epochs, the training regimen prematurely ceased.

Following the iterative refinement to yield the optimal model, attention weight visualization facilitated the identification of atom-level features most pertinent to hERG inhibition activity, thereby providing invaluable insights for medicinal chemists engaged in structural optimization endeavors. The dual-level attention mechanism in AttenhERG first captures local atomic features and then integrates global molecular characteristics. This allows AttenhERG to autonomously learn the chemical environments of atoms, effectively pinpointing substructures critical for hERG inhibition rather than merely evaluating the entire molecular structure. While preserving model integrity, introducing two uncertainty estimation methodologies, Entropy and MC-Dropout [37], facilitated a comprehensive assessment of prediction uncertainty, thereby bolstering prediction reliability. We were then able to sequentially discard the top 10% highest uncertainty predictions from the in-house dataset and compute the MCC, BAC, and AUROC of the prediction outcomes.

Baseline model

We employed four machine learning algorithms as baseline models for hERG prediction: Random Forest (RF), Support Vector Machine (SVM), Directed Message Passing Neural Network (D-MPNN), and Bayesian Neural Network (BayeshERG). These four baseline models have been reported in previous studies with the same parameter settings [27]. Additionally, to ensure a fair comparison of model performance, several models and their results selected from the literature were used to evaluate the generalization performance of the models on multiple external test sets. The chosen baseline models include CardPred, DeepHIT, hERG-att, ADMETlab 2.0, PredhERG 4.2, AdmetSAR 2.0, AMED, Ochem I/II, Cardio-Tox, BayeshERG, and the Siramshetty et al. model. The performance values for these baseline models were primarily sourced from previous studies [27]. The chosen updated models include Pred-hERG 5.0, CToxPred and CardioDPi. Among these models, predictions for PredhERG 5.0 were obtained directly from the web service, with the selected result being the consensus-weighted value, which integrates both binary and regression predictions. Similarly, CardioDPi predictions were also conducted using web services, with the CardioDPi-hERG value being selected. CToxPred was implemented using the provided open-source code, with default parameters applied. DMFGAM was implemented using the available open-source code with default parameters.

Evaluation metrics

To assess prediction performance, we utilized standard evaluation metrics, including accuracy (ACC), balanced accuracy (BAC), Matthews's correlation coefficient (MCC), F1 score, sensitivity (SEN), specificity (SPE), area under the receiver operating characteristic curve (AUROC), and area under the precision-recall curve (AUPRC). Additionally, in evaluating the test set, we focused on analyzing MCC, BAC, and AUROC, with their respective calculation formulas as follows:

$$BAC = \frac{1}{2} \times \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right) \tag{1}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$
(2)

$$SEN = \frac{TP}{TP + FN} \tag{3}$$

$$SPE = \frac{TN}{TP + FN} \tag{4}$$

where TP, TN, FP, and FN represent the number of true positives, true negatives, false positives, and false negatives, respectively. The AUROC represents the area under the ROC curve, incorporating the true positive rate (TPR) and false positive rate (FPR). These metrics are computed using the scikit-learn package in Python 3.6.15.

Abbreviations

hERG	Human ether-a-go-go-related gene
ICH	International Conference on Harmoniszation of Technical Require-
	ments for the Registration of Pharmaceuticals for Human Use
SVM	Support vector mechanisms
ML	Machine learning
DL	Deep learning
DNNs	Deep neural networks
D-MPNN	Directed message passing neural network
APH1A	Aph-1 homolog A, gamma-secretase subunit:
NMT1	N-Myristoyltransferase
AUPRC	Area under precision-recall curve
MCC	Matthew's correlation coefficient
BAC	Balanced accuracy
TPR	True positive rate
FPR	False positive rate

Supplementary Information

The online version contains supplementary material available at https://doi. org/10.1186/s13321-024-00940-y.

Supplementary material 1.

Acknowledgements

We sincerely appreciate Fadi E. Pulous for his valuable assistance in polishing the language of this manuscript.

Author contributions

T. Y., X. D. designed and performed the experiments, prepared the figures, and wrote the manuscript; E. M., F. W. P., A. A. reviewed and assisted in revising the manuscript; X. D., A. Z., F. R. conceived and drove the project.

Funding

Not applicable.

Data availability

Data is provided within the manuscript or supplementary information files.

Declarations

Ethics approval and consent to participate Not applicable.

Competing interests

The authors declare no competing interests.

Received: 12 September 2024 Accepted: 8 December 2024 Published online: 23 December 2024

References

- Alexandre J, Moslehi JJ, Bersell KR et al (2018) Anticancer drug-induced cardiac rhythm disorders: current knowledge and basic underlying mechanisms. Pharmacol Ther 189:89–103. https://doi.org/10.1016/j. pharmthera.2018.04.009
- Fermini B, Fossa AA (2003) The impact of drug-induced QT interval prolongation on drug discovery and development. Nat Rev Drug Discov 2:439–447. https://doi.org/10.1038/nrd1108
- Recanatini M, Poluzzi E, Masetti M et al (2005) QT prolongation through hERG K+ channel blockade: current knowledge and strategies for the early prediction during drug development. Med Res Rev 25:133–166. https://doi.org/10.1002/med.20019
- Villoutreix BO, Taboureau O (2015) Computational investigations of hERG channel blockers: new insights and current predictive models. Adv Drug Deliv Rev 86:72–82. https://doi.org/10.1016/j.addr.2015.03.003
- Darpo B, Nebout T, Sager PT (2006) Clinical Evaluation of QT/QTc Prolongation and proarrhythmic potential for nonantiarrhythmic drugs: the international conference on harmonization of technical requirements for registration of pharmaceuticals for human use E14 guideline. J Clin Pharmacol. 46:498–507. https://doi.org/10.1177/0091270006286436
- Fenichel RR, Malik M, Antzelevitch C et al (2004) Drug-induced torsades de pointes and implications for drug development. J Cardiovasc Electrophysiol 15:475–495. https://doi.org/10.1046/j.1540-8167.2004.03534.x
- Kalyaanamoorthy S, Barakat KH (2018) Development of safe drugs: the hERG challenge. Med Res Rev 38:525–555. https://doi.org/10.1002/med. 21445
- AlRawashdeh S, Chandrasekaran S, Barakat KH (2023) Structural analysis of hERG channel blockers and the implications for drug design. J Mol Graph Model 120:108405. https://doi.org/10.1016/j.jmgm.2023.108405

- Thai K-M, Ecker GF (2008) A binary QSAR model for classification of hERG potassium channel blockers. Bioorg Med Chem 16:4107–4119. https:// doi.org/10.1016/j.bmc.2008.01.017
- 10. Seierstad M, Agrafiotis DK (2006) A QSAR model of hERG binding using a large, diverse, and internally consistent training set. Chem Biol Drug Des 67:284–296. https://doi.org/10.1111/j.1747-0285.2006.00379.x
- 11. Braga CR, Alves MV, Silva FBM et al (2014) Tuning hERG Out: antitarget QSAR models for drug development. Curr Top Med Chem 14:1399–1415. https://doi.org/10.2174/1568026614666140506124442
- Delre P, Lavado GJ, Lamanna G et al (2022) Ligand-based prediction of hERG-mediated cardiotoxicity based on the integration of different machine learning techniques. Front Pharmacol 13:951083. https://doi. org/10.3389/fphar.2022.951083
- Diller JD (2009) In Silico hERG modeling: challenges and progress. Curr Comput Aided Drug Des 5:106–121. https://doi.org/10.2174/1573409097 88451928
- Rácz A, Bajusz D, Miranda-Quintana RA, Héberger K (2021) Machine learning models for classification tasks related to drug safety. Mol Divers 25:1409–1424. https://doi.org/10.1007/s11030-021-10239-x
- Cavasotto CN, Scardino V (2022) Machine learning toxicity prediction: latest advances by toxicity end point. ACS Omega 7:47536–47546. https:// doi.org/10.1021/acsomega.2c05693
- Siramshetty VB, Nguyen D-T, Martinez NJ et al (2020) Critical assessment of artificial intelligence methods for prediction of hERG channel inhibition in the "big data" era. J Chem Inf Model 60:6007–6019. https://doi.org/ 10.1021/acs.jcim.0c00884
- Braga RC, Alves VM, Silva MFB et al (2015) Pred-hERG: a novel webaccessible computational tool for predicting cardiac toxicity. Mol Inform 34:698–701. https://doi.org/10.1002/minf.201500040
- Ogura K, Sato T, Yuki H, Honma T (2019) Support Vector Machine model for hERG inhibitory activities based on the integrated hERG database using descriptor selection by NSGA-II. Sci Rep 9:12220. https://doi.org/10. 1038/s41598-019-47536-3
- Li X, Zhang Y, Li H, Zhao Y (2017) Modeling of the hERG K+ channel blockage using online chemical database and modeling environment (OCHEM). Mol Inform 36:1700074. https://doi.org/10.1002/minf.20170 0074
- Yang H, Lou C, Sun L et al (2019) admetSAR 2.0: web-service for prediction and optimization of chemical ADMET properties. Bioinformatics 35:1067–1069. https://doi.org/10.1093/bioinformatics/bty707
- Muller C, Rabal O, Diaz Gonzalez C (2022) Artificial intelligence, machine learning, and deep learning in real-life drug design cases. In: Heifetz A (ed) Artificial intelligence in drug design. Springer, US, New York, NY, pp 383–407
- Lee H-M, Yu M-S, Kazmi SR et al (2019) Computational determination of hERG-related cardiotoxicity of drug candidates. BMC Bioinf 20:250. https://doi.org/10.1186/s12859-019-2814-5
- Ryu JY, Lee MY, Lee JH et al (2020) DeepHIT: a deep learning framework for prediction of hERG-induced cardiotoxicity. Bioinformatics 36:3049– 3055. https://doi.org/10.1093/bioinformatics/btaa075
- Karim A, Lee M, Balle T, Sattar A (2021) CardioTox net: a robust predictor for hERG channel blockade based on deep learning meta-feature ensembles. J Cheminf 13:60. https://doi.org/10.1186/s13321-021-00541-z
- Kim H, Nam H (2020) hERG-Att: Self-attention-based deep neural network for predicting hERG blockers. Comput Biol Chem 87:107286. https://doi. org/10.1016/j.compbiolchem.2020.107286
- Xiong G, Wu Z, Yi J et al (2021) ADMETIab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. Nucleic Acids Res 49:W5–W14. https://doi.org/10.1093/nar/gkab255
- 27. Kim H, Park M, Lee I, Nam H (2022) BayeshERG: a robust, reliable and interpretable deep learning model for predicting hERG channel blockers. Brief Bioinf 23:bbca211. https://doi.org/10.1093/bib/bbac211
- Wang T, Sun J, Zhao Q (2023) Investigating cardiotoxicity related with hERG channel blockers using molecular fingerprints and graph attention mechanism. Comput Biol Med 153:106464. https://doi.org/10.1016/j. compbiomed.2022.106464

- 29. Sanches IH, Braga RC, Alves VM, Andrade CH (2024) Enhancing hERG risk assessment with interpretable classificatory and regression models. Chem Res Toxicol 37:910–922. https://doi.org/10.1021/acs.chemrestox. 3c00400
- Arab I, Egghe K, Laukens K et al (2024) Benchmarking of small molecule feature representations for hERG, Nav1.5, and Cav1.2 cardiotoxicity prediction. J Chem Inf Model 64:2515–2527. https://doi.org/10.1021/acs. jcim.3c01301
- Chen Z, Li N, Zhang P et al (2024) CardioDPi: an explainable deep-learning model for identifying cardiotoxic chemicals targeting hERG, Cav1.2, and Nav1.5 channels. J Hazard Mater 474:134724. https://doi.org/10. 1016/j.jhazmat.2024.134724
- Wu Z, Wang J, Du H et al (2023) Chemistry-intuitive explanation of graph neural networks for molecular property prediction with substructure masking. Nat Commun 14:2585. https://doi.org/10.1038/ s41467-023-38192-3
- Yi J, Shi S, Fu L et al (2024) OptADMET: a web-based tool for substructure modifications to improve ADMET properties of lead compounds. Nat Protoc. https://doi.org/10.1038/s41596-023-00942-4
- Fischer C, Zultanski SL, Zhou H et al (2012) Triazoloamides as potent γ-secretase modulators with reduced hERG liability. Bioorg Med Chem Lett 22:3140–3146. https://doi.org/10.1016/j.bmcl.2012.03.054
- 35. Brand S, Norcross NR, Thompson S et al (2014) Lead optimization of a pyrazole sulfonamide series of trypanosoma brucei N-myristoyltransferase inhibitors: identification and evaluation of CNS penetrant compounds as potential treatments for stage 2 human african trypanosomiasis. J Med Chem 57:9855–9869. https://doi.org/10.1021/jm500809c
- Xiong Z, Wang D, Liu X et al (2020) Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. J Med Chem 63:8749–8760. https://doi.org/10.1021/acs.jmedchem.9b009 59
- Tong X, Wang D, Ding X et al (2022) Blood–brain barrier penetration prediction enhanced by uncertainty estimation. J Cheminf 14:44. https:// doi.org/10.1186/s13321-022-00619-2
- Wang J, Zhang L, Sun J et al (2024) Predicting drug-induced liver injury using graph attention mechanism and molecular fingerprints. Methods 221:18–26. https://doi.org/10.1016/j.ymeth.2023.11.014
- Yang X, Sun J, Jin B et al (2024) Multi-task aquatic toxicity prediction model based on multi-level features fusion. J Adv Res. https://doi.org/10. 1016/j.jare.2024.06.002
- Sun F, Sun J, Zhao Q (2022) A deep learning method for predicting metabolite–disease associations via graph neural network. Brief Bioinform 23:266. https://doi.org/10.1093/bib/bbac266
- Chen Z, Zhang L, Sun J et al (2023) DCAMCP: A deep learning model based on capsule network and attention mechanism for molecular carcinogenicity prediction. J Cell Mol Med 27:3117–3126. https://doi.org/ 10.1111/jcmm.17889
- Sushko I et al (2011) Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information. J Cheminf 3:P20–P20. https://doi.org/10.1186/ 1758-2946-3-S1-P20
- Mendez D, Gaulton A, Bento AP et al (2019) ChEMBL: towards direct deposition of bioassay data. Nucleic Acids Res 47:D930–D940. https://doi. org/10.1093/nar/gky1075
- Kim S, Chen J, Cheng T et al (2021) PubChem in 2021: new data content and improved web interfaces. Nucleic Acids Res 49:D1388–D1395. https://doi.org/10.1093/nar/gkaa971
- 45. Gilson MK, Liu T, Baitaluk M et al (2016) BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. Nucleic Acids Res 44:D1045–D1053. https://doi.org/10. 1093/nar/gkv1072
- Tetko IV, van Deursen R, Godin G (2024). Be aware of overfitting by hyperparameter optimization!. arXiv preprint arXiv:2407.20786

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.