



OPEN ACCESS

EDITED BY

Yiqun Deng,
South China Agricultural University,
China

REVIEWED BY

Rodolpho C. Braga,
InsilicAll, Brazil
Ruili Huang,
National Center for Advancing
Translational Sciences (NIH),
United States
M. Natalia D. S. Cordeiro,
University of Porto, Portugal

*CORRESPONDENCE

Giuseppe Felice Mangiatordi,
giuseppe.mangiatordi@ic.cnr.it
Domenico Gadaleta,
domenico.gadaleta@marionegri.it

SPECIALTY SECTION

This article was submitted to Predictive
Toxicology,
a section of the journal
Frontiers in Pharmacology

RECEIVED 23 May 2022

ACCEPTED 20 July 2022

PUBLISHED 05 September 2022

CITATION

Delre P, Lavado GJ, Lamanna G,
Saviano M, Roncaglioni A, Benfenati E,
Mangiatordi GF and Gadaleta D (2022),
Ligand-based prediction of hERG-
mediated cardiotoxicity based on the
integration of different machine
learning techniques.
Front. Pharmacol. 13:951083.
doi: 10.3389/fphar.2022.951083

COPYRIGHT

© 2022 Delre, Lavado, Lamanna,
Saviano, Roncaglioni, Benfenati,
Mangiatordi and Gadaleta. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

Ligand-based prediction of hERG-mediated cardiotoxicity based on the integration of different machine learning techniques

Pietro Delre^{1,2}, Giovanna J. Lavado³, Giuseppe Lamanna^{1,2},
Michele Saviano⁴, Alessandra Roncaglioni³, Emilio Benfenati³,
Giuseppe Felice Mangiatordi^{1*} and Domenico Gadaleta^{3*}

¹CNR—Institute of Crystallography, Bari, Italy, ²Chemistry Department, University of Bari “Aldo Moro”, Bari, Italy, ³Laboratory of Environmental Chemistry and Toxicology, Department of Environmental Health Sciences, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milan, Italy, ⁴CNR-Institute of Crystallography, Caserta, Italy

Drug-induced cardiotoxicity is a common side effect of drugs in clinical use or under postmarket surveillance and is commonly due to off-target interactions with the cardiac human-ether-a-go-go-related (hERG) potassium channel. Therefore, prioritizing drug candidates based on their hERG blocking potential is a mandatory step in the early preclinical stage of a drug discovery program. Herein, we trained and properly validated 30 ligand-based classifiers of hERG-related cardiotoxicity based on 7,963 curated compounds extracted by the freely accessible repository ChEMBL (version 25). Different machine learning algorithms were tested, namely, random forest, K-nearest neighbors, gradient boosting, extreme gradient boosting, multilayer perceptron, and support vector machine. The application of 1) the best practices for data curation, 2) the feature selection method VSURF, and 3) the synthetic minority oversampling technique (SMOTE) to properly handle the unbalanced data, allowed for the development of highly predictive models ($BA_{MAX} = 0.91$, $AUC_{MAX} = 0.95$). Remarkably, the undertaken temporal validation approach not only supported the predictivity of the herein presented classifiers but also suggested their ability to outperform those models commonly used in the literature. From a more methodological point of view, the study put forward a new computational workflow, freely available in the GitHub repository (<https://github.com/PDelre93/hERG-QSAR>), as valuable for building highly predictive models of hERG-mediated cardiotoxicity.

KEYWORDS

hERG, cardiotoxicity, QSAR, ligand-based, consensus modeling

Introduction

Background

Cardiotoxicity is a common side effect of drugs, and one of the causes for it is the off-target interaction with different voltage-gated ion channels expressed in the heart (Ferdinandy et al., 2019). Among the others, the human ether-a-go-go related (hERG) channel has received increasing attention over the past few decades as several drugs have been restricted in their use or withdrawn because of their ability to block this channel by interacting with a hydrophobic pocket called central cavity (Kalyaanamoorthy and Barakat, 2018; Butler et al., 2020). Remarkably, a drug-induced hERG blockade can be responsible for potentially lethal cardiac arrhythmias in the form of the so-called long-QT syndrome (LQTS) (Priest et al., 2008; Danker and Möller, 2014). Since drugs belonging to very different chemical classes were proved to cause this severe side effect, an early evaluation of hERG blockade has become a necessary step during the development of drug discovery (DD) programs (Kalyaanamoorthy and Barakat, 2018; Ferdinandy et al., 2019; Cavalluzzi et al., 2020). Meaningful examples are represented by terfenadine (Kamiya et al., 2008; Tanaka et al., 2014), astemizole (Zhou et al., 1999), cisapride (Walker et al., 1999; Kamiya et al., 2008), and ziprasidone (Su et al., 2006). As a matter of fact, submission to regulatory reviews requires a preclinical assessment of hERG blockade activities, as clearly indicated by the guidelines defined at the International Conference on Harmonization of Technical Requirements for the Registration of Pharmaceuticals for Human Use (ICH) (EMA, 2005; FDA, 2005).

In silico evaluation of hERG blockade

In this context, to avoid hERG liability during the DD process and, therefore, prioritize safe drug candidates at the early preclinical stage, the employment of *in silico* tools is highly desirable since *in vitro* (e.g., fluorescence-based assays, electrophysiology measurements, rubidium-flux assays, radioligand binding assays) and *in vivo* experiments are much more laborious, time-consuming, and expensive (Priest et al., 2008; Jing et al., 2015). Accordingly, several *in silico* tools have been developed in the last few years using both ligand- and structure-based approaches (Jing et al., 2015; Villoutreix and Taboureau, 2015; Kalyaanamoorthy and Barakat, 2018; Creanza et al., 2021). In the absence of an atomic-resolution hERG structure, the attention of both academia and industry has been mainly focused on the development of ligand-based classifiers (e.g., pharmacophore models, quantitative structure–activity relationship (QSAR) approaches) (Jing et al., 2015; Villoutreix and Taboureau, 2015; Slavov et al., 2017; Sun et al., 2017; Kalyaanamoorthy and Barakat, 2018). The interested

reader is referred to references (Jing et al., 2015; Villoutreix and Taboureau, 2015; Kalyaanamoorthy and Barakat, 2018) for comprehensive reviews on this topic. However, despite providing good performances, many ligand-based models developed so far suffer from critical limitations. Most of them, in fact, were built from a limited number of congeneric analogs (Chavan et al., 2016; Gobbi et al., 2016; Wang et al., 2016; Zhang et al., 2016; Munawar et al., 2018; Liu et al., 2020), and for this reason, their applicability domain (AD) (Gadaleta et al., 2016) is too restricted for a real-life application, since hERG blockers are characterized by high structural diversity. Recent articles published by Cai et al. (2019), Ryu et al. (2020), and Karim et al. (2021) reported classifiers trained on more than 7,000 compounds, hence encompassing a broad AD. The authors used deep learning techniques and $IC_{50} = 10 \mu\text{M}$ as the toxicity threshold to discern hERG blockers from nonblockers. As a result, these models ensured performances better than those achieved using more traditional machine learning (ML) approaches ($AUC_{MAX} = 0.88, 0.90, \text{ and } 0.91$, respectively). However, the real-life application of such classifiers might be questionable as substances considered of critical concern in DD programs are responsible for IC_{50} lower than $1 \mu\text{M}$ (rather than $10 \mu\text{M}$) (Zolotoy et al., 2003; Katchman et al., 2006; Kim et al., 2008; Hong et al., 2013). In this regard, Krishna et al. (Krishna et al., 2022) recently developed QSAR models based on Tox21 quantitative high throughput screening (qHTS) thallium flux assay and ChEMBL (v27) data using $IC_{50} = 10 \mu\text{M}$ as the toxicity threshold to discern hERG blockers from nonblockers. Different ML methods and consensus modeling were evaluated. The best models were ultimately integrated into a consensus that improved the performance of the single best ones (BA = 0.791 on the external set).

Objectives

Building on these pieces of evidence and background, in this study we report several ligand-based models developed starting from 7,963 curated compounds (hERG-DB) and extracted from the ChEMBL (Gaulton et al., 2012) version (v) 25, employing six classification algorithms, random forest (RF) (Breiman, 2001), K-nearest neighbors (KNN) (Altman, 1992), gradient boosting (GB) (Friedman, 2001), extreme gradient boosting (XGB) (Chen and Guestrin, 2016), multilayer perceptron (MLP) (Haykin, 1994), and support vector machine (SVM) (Vapnik, 1963), and as toxicity thresholds both $IC_{50} = 1$ and $IC_{50} = 10 \mu\text{M}$. Notably, the models returning the best performances were challenged on a supplementary external set (ES) consisting of molecules recovered from the new ChEMBL (Gaulton et al., 2012) update (v28) and therefore not included in hERG-DB (see section “Materials and methods” for methodological details). This procedure, also known as time-split cross-validation or

temporal validation and involving data published a posteriori, is today considered of utmost importance to test the real-life applicability of the developed classifiers (Sheridan, 2013; Bosc et al., 2019). Last but not least, six ligand-based hERG models available in the literature and named DeepHIT (Ryu et al., 2020), CardioTox (Karim et al., 2021), Cardprep (Lee et al., 2019), ADMETlab (Xiong et al., 2021), and OCHEM consensus models I and II (Li et al., 2017) were applied on the ES to compare their performance with that returned by our top-performing model. The obtained results put forward the herein presented computational workflow as valuable for a robust ligand-based prediction of hERG-related cardiotoxicity. This study can be useful also considering that cardiovascular diseases are ranked high in the top causes of sudden death (Onakpoya et al., 2016).

Materials and methods

Dataset preparation

A total of 17,952 activity entries were extracted from ChEMBL (Gaulton et al., 2012) v25 according to the Target ID (ChEMBL240) assigned to the hERG channel. To ensure data validity, the database was mined retaining only those entries matching the following criteria already suggested in the literature: 1) annotated exclusively with IC_{50} (11,144 entries) measures, 2) referring to assays conducted on human targets ("target_organism" = "*Homo sapiens*"), 3) marked as direct binding ("assay_type" = "B"), and 4) free of warnings in the "data_validity_comment" field (Alberga et al., 2019; Bosc et al., 2019; Creanza et al., 2021). SMILES were curated using a semiautomated in-house procedure described by Gadaleta et al. (2018a). Such a process allows for removing organometallic and inorganic compounds, chemicals characterized by unusual elements and mixtures, neutralizing salts, and removing stereochemistry. The neutralized SMILES were converted to a standardized QSAR-ready format using OpenBabel (O'Boyle et al., 2011) implemented in the KNIME Analytics Platform (Berthold et al., 2008) to generate canonical SMILES. Each IC_{50} value was converted from molar concentration (M) to pIC_{50} ($-\log IC_{50}$), and compounds devoid of any pIC_{50} value but already marked as not active in the ChEMBL repository were also considered. In the last step, duplicates were aggregated in unique entries and the standard deviation (σ) related to the pIC_{50} values was computed. Also, 15 compounds were considered outliers ($\sigma > 2$) while for all of the others, the average pIC_{50} value was considered. In such a way, the curated dataset consists of 7,963 chemicals (hereinafter referred to as hERG-DB) and the corresponding experimental value. Consistent with the literature (Jing et al., 2015; Zolotoy et al., 2003), hERG-DB includes hERG blockers (ACT) having an $IC_{50} \leq 1 \mu M$ ($pIC_{50} \leq 6$), compounds showing a moderate hERG blocker potential with an IC_{50} ranging from 1 to 10 μM

($6 < pIC_{50} \leq 5$), and, finally, hERG nonblockers (INA) having IC_{50} values $>10 \mu M$ ($pIC_{50} > 5$). For this reason, in this work, we have developed two sets of binary models differing for the considered toxicity threshold ($pIC_{50} = 6$ or $pIC_{50} = 5$). Finally, we downloaded the recent version of ChEMBL (Gaulton et al., 2012) (version 28) to extract possible compounds not present in our hERG-DB, following the same data curation process described above. Thus, an external (ES) dataset of 792 chemicals was assembled and curated with the same procedure described above and then used to challenge the real-life predictivity of the top-performing classifiers.

Dataset division

We split hERG-DB into a training set (TS) and a validation set (VS) following a rational approach. Notably, the RDkit Picker Diversity node (Landrum et al., 2022) was employed separately on the two classes (i.e., ACT and INA) resulting from the considered toxicity thresholds to generate the Morgan fingerprints (Rogers and Hahn, 2010) for each SMILES, and 80% of the most diverse molecules was then picked based on the Tanimoto distance (Willett et al., 1998). In doing so, the resulting TS included 80% of the total molecules (6,371); the remaining 20% (1592) constituted the VS. Table 1 summarizes the final composition of TS, VS, and ES, indicating the number of ACT and INA for each of the two selected toxicity thresholds. Note that such a procedure allowed us to keep the INA/ACT ratio fixed in each subdivision. Interestingly, although not prepared by a rational strategy, the ES presented an INA/ACT ratio in line with those of TS and VS.

The splitting procedure was challenged by performing a principal component analysis (PCA) (Jolliffe and Cadima, 2016) based on the physicochemical properties of the molecules calculated by the molecular properties KNIME node based on the CDK toolkit (Steinbeck et al., 2003; CDK, 2022. Available at: <https://cdk.github.io/>). The score plot of the first two principal components, able to capture more than 90% of the data variance, confirms that the described procedure ensured a uniform distribution of the compounds in the TS and VS throughout the model space (Figure 1). In addition, Figure 1 includes the PCA of the ES. It is worth noting that, although not derived by a splitting approach, this covers a chemical space similar to those covered by both TS and VS.

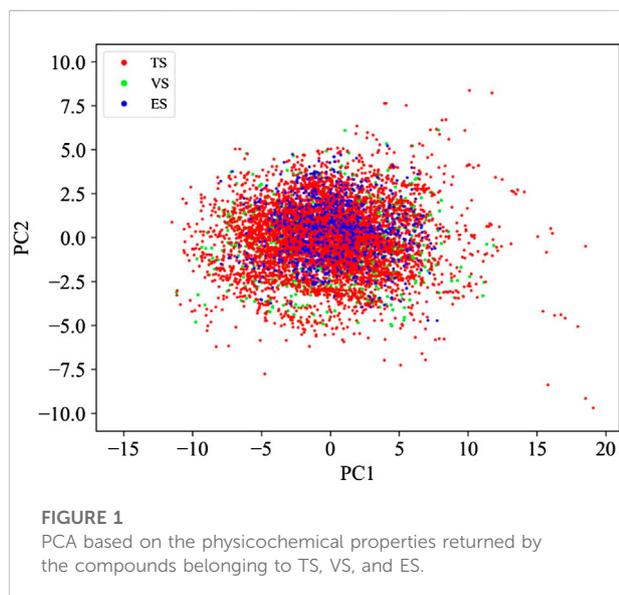
Development of statistically based models

Descriptors calculation

We used DRAGON v7.0.4 (Kode, 2017) as a software program to compute the 2D-descriptors of each chemical belonging to the datasets. Descriptors having missing values or constant/near-constant variables (i.e., standard deviation $<$

TABLE 1 Partitioning schemes before (top) and after the application of the AD at each considered toxicity threshold (bottom). For *hERG-DB*, the number of active and inactive chemicals and the related class distribution is reported for the training set (TS), validation set (VS), and external set (ES) and at each considered toxicity threshold. Notably, the total number of chemicals (#), the number of *hERG* blockers (ACT) and *hERG* non-blocker (INA) chemicals, as well as the ratio between nonblockers and blockers are shown.

Dataset	Toxicity threshold (pIC ₅₀)							
	6				5			
	#	INA	ACT	INA:ACT	#	INA	ACT	INA:ACT
Starting composition								
TS	6371	5388	983	05:01	6371	3295	3076	01:01
VS	1592	1346	246	05:01	1592	821	771	01:01
ES	792	676	116	05:01	792	365	427	01:01
Applicability domain (AD)								
VS	1583	1338	245	05:01	1579	810	769	01:01
ES	754	642	112	05:01	754	350	404	01:01



0.01) were removed along with those having an absolute pair correlation higher than 95% with other variables. Thus, we finally considered 1,070 descriptors. The obtained values were scaled with a standard normalization (i.e., mean equal to zero and standard deviation equal to 1). Models were built using both the entire pool of descriptors and a reduced set selected by the R package VSURF, an RF algorithm working in three steps to detect variables related to the activity and eliminate those redundant or irrelevant (Genuer et al., 2010). Following this protocol, we

finally selected 79 (pIC₅₀ threshold = 6) and 86 (pIC₅₀ threshold = 5) descriptors (Supplementary Table S1). The feature selection was based on the TS only to remove any putative artifact in the model selection.

Model development and validation

For each partitioning scheme reported in Table 1, we used six classification algorithms: RF (Breiman, 2001), KNN (Altman, 1992), GB (Friedman, 2001), XGB (Chen and Guestrin, 2016), MLP (Haykin, 1994), and SVM (Vapnik, 1963). The TS is characterized by an INA: ACT ratio equal to 5:1 when the pIC₅₀ toxicity threshold is 6; hence, it is strongly unbalanced. This could favor the convergence of algorithms trained on the majority class, neglecting classes with fewer samples (Zhang et al., 2018). For this reason, in addition to the models developed using this TS, additional models were developed, artificially altering the original TS using the Synthetic minority oversampling technique (SMOTE) to balance the number of blockers and nonblocker samples (Chawla et al., 2002). Such an approach, based on the KNN algorithm and operating in the “feature space”, oversampled the minority class by creating and introducing new synthetic samples until a ratio of INA: ACT of 1:1 is reached. As for the VSURF procedure, the SMOTE was applied only to the TS. VS, indeed, was kept unbalanced to properly evaluate the capability of the classifiers to predict the real distribution of data. In all of the cases, to find the optimal algorithm setting for training the final model, the parameter selection was based on hyperparameter tuning and 5-fold cross-validation (CV) performance (Refaeilzadeh et al.,

2009). To do this, we performed a grid search (LaValle et al., 2004), except for SVM and XGB, where we used Bayesian optimization to reduce the computational cost (Snoek et al., 2012). The optimal parameters for each algorithm, selected based on the best metrics in 5-fold CV, are shown in Supplementary Table S2. It is worth noting that RF models, trained on the original unbalanced TS, combine an equal size data sampling for both thresholds $pIC_{50} = 5$ and $pIC_{50} = 6$. This technique is also known as balanced random forest (BRF). After parameter setup and model training, top-performing models were selected based on external performance on the VS and then applied to the ES that represented the ultimate proof for real-life validation of the models. Finally, to possibly improve predictions provided by single best models, consensus modeling was applied. In particular, a compound was assigned to a category based on a straightforward majority voting approach, i.e., only when the top-performing models, selected based on the computed balanced accuracy (BA) and area under the curve (AUC) values, generated concordant predictions.

Applicability domain

To increase the confidence in the model's prediction, we defined the applicability domain (AD), namely, the chemical space from which the classifiers are derived and, therefore, where a prediction can be considered trustworthy (Roy et al., 2015; Gadaleta et al., 2016; Kar et al., 2018). To define the AD, we used the Enalos Domain—Leverages node for KNIME (Afantitis et al., 2008; Melagraki et al., 2009). This approach allows for the calculation of the leverage (h) for each chemical and defines a threshold that works as an upper bound limit. Compounds with leverage values of $h > 3p/n$, where p is the number of descriptors and n is the number of molecules, are considered chemically different from the TS compounds (Tropsha et al., 2003; Afantitis et al., 2008; Melagraki et al., 2009). Thus, 9 (threshold $pIC_{50} = 6$) and 13 (threshold $pIC_{50} = 5$) were discarded from VS, whereas 38 compounds were excluded from ES for both the considered activity thresholds. Table 1 reports the composition of both VS and ES after applying the AD-based filter.

Performance evaluation

The performance of the classification models was evaluated using Coopers statistics, i.e., balanced accuracy (BA), sensitivity (SE), and specificity (SP), computed as follows:

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

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

$$BA = \frac{SE + SP}{2}$$

where true positives (TPs), and true negatives (TNs) are, respectively, the positive and negative samples correctly classified by the models and false negatives (FNs) and false positives (FPs) are the misclassified positive and negative samples, respectively (Ting, 2017). Another used metric was the Matthews correlation coefficient (MCC). MCC indicates the quality of binary classification and is generally recognized as a reliable metric, although it deteriorates seriously when the TS is imbalanced (Zhu, 2020). MCC ranges between -1 and $+1$. A value of $+1$ means a perfect classification, 0 indicates a random classification, and -1 is a complete misclassification (Zhu, 2020).

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

Finally, the AUC, namely, the area under the receiver operating characteristic (ROC) curve, was also computed to estimate the predictive accuracy of the models. Notice that the AUC, ranging from 0 (miss-classifiers) to 1 (ideal-classifiers), reflects the probability of positive compounds being ranked earlier than decoy compounds (Fawcett, 2006). This quality metric was computed for each developed model based on the output scores associated with each prediction returned during the validation procedure and estimating the probability that a given compound is an hERG blocker.

Results and discussion

In the present work, we developed QSAR models employing different ML algorithms, RF (Breiman, 2001), KNN (Altman, 1992), GB (Friedman, 2001), XGB (Chen and Guestrin, 2016), MLP (Haykin, 1994), and SVM (Vapnik, 1963) available in the KNIME Analytics Platform (v. 4.1.4) (Berthold et al., 2008). The dataset used to build the model consists of highly curated pIC_{50} values for 7,963 organic compounds. This dataset allows for the covering of a wide range of structural characteristics of hERG blockers and nonblockers and also a broad chemical space, as evident in Figure 1. Based on the literature, threshold values for the blocker/nonblocker classification vary from $IC_{50} = 1 \mu M$ ($pIC_{50} = 6$) to $IC_{50} = 10 \mu M$ ($pIC_{50} = 5$) (Jing et al., 2015; Li et al., 2017; Siramshetty et al., 2018, 2020; Choi et al., 2020). For this reason, we used these two thresholds to develop binary classification models. In addition, one set of models accounted for all of the descriptors generated by DRAGON v7.0.4 (Kode, 2017), while the other, only the pool of descriptors selected by VSURF (Genuer et al., 2010) (Supplementary Table S1). Notably, the performances returned by the two sets of the models are comparable, remarking the effectiveness of the selection variable strategy, as already experienced in previous works (Gadaleta et al., 2018b; Baderna et al., 2020; Lavado et al., 2020). Therefore, we have focused our attention on the results returned by the VSURF models, being less complex and easier

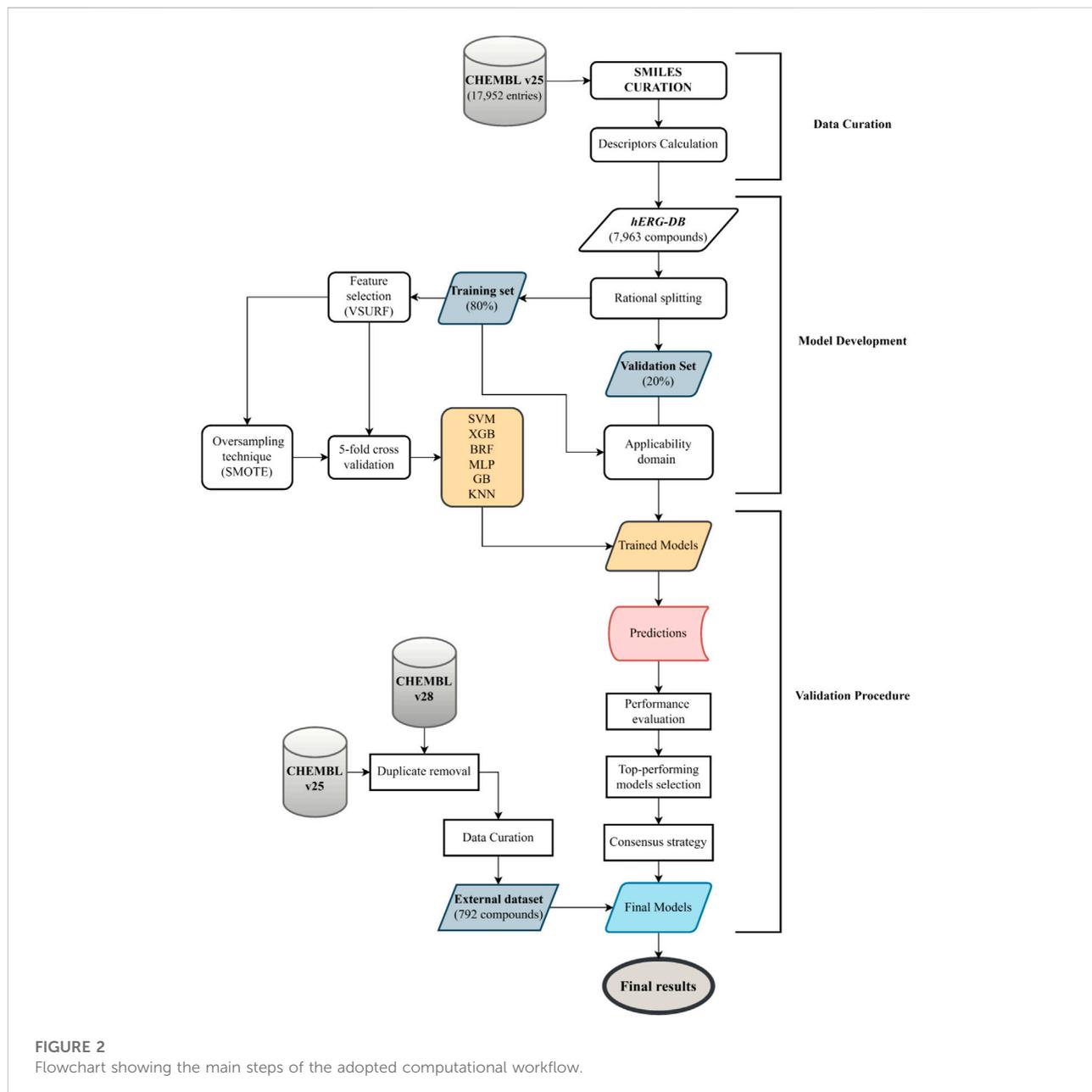


FIGURE 2
Flowchart showing the main steps of the adopted computational workflow.

to be implemented. However, the interested reader is referred to the supporting information for the performances in validation and 5-fold CV returned by all of the models trained with the entire set of descriptors (Supplementary Tables S3, S4). The discussion will focus on the most important metrics to determine the top-performing models, SE, SP, BA, and AUC, given the imbalance of TS. In addition, we used an ES as a temporal validation to assess the predictivity of the models in a real-life case study. Finally, for the sake of comparison, the performance of our top-selected models has been compared with that obtained on the ES with commonly employed and freely accessible models: DeepHIT (Ryu et al., 2020), CardioTox (Karim et al., 2021),

Cardprep (Lee et al., 2019), ADMETlab (Xiong et al., 2021), and OCHEM consensus models I and II (Li et al., 2017). For the sake of clarity, Figure 2 shows the workflow that summarizes the main steps of the adopted computational protocol. Notably, the use of a rational approach for data split allowed us to minimize the risk of variation in performance due to different TS-VS divisions. Indeed, the iteration of the rational data-split procedure performed on models based on the entire set of descriptors does not show relevant differences in terms of statistical performance (Supplementary Table S5) with respect to models based on the TS-VS split presented here (Supplementary Table S4).

TABLE 2 Performances on the VS of the models developed using $pIC_{50} = 6$ (top) and 5 (bottom). For each model, the following statistics are reported: balanced accuracy (BA), sensitivity (SE), specificity (SP), Matthews correlation coefficient (MCC), area under the ROC (AUC), number of true negatives (TNs), false positives (FPs), true positives (TPs), and false negatives (FNs). The top-performing model selected for additional validation is indicated in bold.

Toxicity threshold $pIC_{50} = 6$

Balancing	Method	BA	SE	SP	MCC	AUC	TP	FP	TN	FN
-	BRF	0.87	0.92	0.81	0.58	0.95	226	250	1088	19
	GB	0.82	0.68	0.96	0.68	0.94	170	52	1286	75
	KNN	0.84	0.72	0.96	0.70	0.91	176	51	1287	69
	MLP	0.76	0.56	0.96	0.57	0.89	136	54	1284	109
	XGB	0.84	0.73	0.95	0.69	0.94	179	60	1278	66
	SVM	0.84	0.72	0.96	0.69	0.93	176	58	1280	69
SMOTE	(S)RF	0.85	0.73	0.96	0.72	0.95	178	46	1292	67
	(S)GB	0.83	0.72	0.95	0.66	0.82	177	72	1266	68
	(S)KNN	0.86	0.86	0.85	0.59	0.92	214	209	1129	31
	(S)MLP	0.81	0.86	0.76	0.48	0.91	211	322	1016	34
	(S)XGB	0.78	0.74	0.82	0.46	0.87	181	239	1099	64
	(S)SVM	0.88	0.91	0.85	0.62	0.89	223	198	1140	22

Toxicity threshold $pIC_{50} = 5$

-	BRF	0.83	0.85	0.82	0.67	0.92	655	147	663	114
	GB	0.83	0.84	0.82	0.66	0.92	646	147	663	123
	KNN	0.83	0.84	0.82	0.66	0.91	647	151	659	122
	MLP	0.80	0.83	0.77	0.60	0.88	640	192	618	129
	XGB	0.83	0.84	0.82	0.65	0.91	643	150	660	126
	SVM	0.83	0.86	0.80	0.66	0.91	659	163	647	110
SMOTE	(S)RF	0.83	0.84	0.82	0.67	0.92	645	143	673	118
	(S)GB	0.83	0.84	0.82	0.64	0.92	637	147	672	123
	(S)KNN	0.82	0.86	0.78	0.64	0.90	652	178	641	108
	(S)MLP	0.77	0.84	0.69	0.65	0.85	645	143	665	126
	(S)XGB	0.84	0.85	0.83	0.68	0.90	681	136	647	115
	(S)SVM	0.84	0.84	0.84	0.68	0.91	650	130	678	121

Models developed using $pIC_{50} = 6$ as the toxicity threshold

Table 2 reports the computed performances on the VS returned by each model based on the activity threshold $pIC_{50} = 6$ without and with the application of the SMOTE (S) (Chawla et al., 2002). Performance refers only to chemicals included in the AD (see section Applicability Domain). Among the models trained on the unbalanced TS (nonblockers/blockers ratio equal to 5), the RF model, combined with a uniform size sampling strategy (BRF) to reduce the bias toward the majority class, returns the best performance. In particular, BRF is responsible for the most balanced statistics when predicting both positive and negative samples with a difference of SE (0.92) and SP (0.81) of only

0.11 and for the highest BA (0.87). All of the other classifiers are characterized by a high FN rate, despite returning acceptable values of BA. The gap between SP and SE ranges from 0.24 (SVM) to 0.40 in (MLP) and might be the result of the absence, in these models, of any procedure to properly consider the TS unbalance. The importance of having a balanced TS is supported by the statistics returned by the models taking advantage of the SMOTE. Indeed, some of them returned a significant performance improvement in predicting both positive and negative samples. (S)KNN and (S)SVM ensured the best performances among all of the developed models, with (S)SVM associated with the best BA (0.88), accounting for a SE of 0.91 and, as a consequence, for a low rate of false negatives. Building on these data, we can reasonably claim that the top-performing models to be selected for additional validation and

TABLE 3 Performance of the consensus models on the VS and on the ES (temporal validation) developed using $pIC_{50} = 6$ (top) and 5 (bottom). For each model, the following statistics are reported: balanced accuracy (BA), sensitivity (SE), specificity (SP), Matthews correlation coefficient (MCC), area under the ROC (AUC), number of true negatives (TNs), false positives (FPs), true positives (TPs), false negatives (FNs), and the total number of molecules (#). The top-performing models selected for temporal validation are indicated in bold.

Toxicity threshold $pIC_{50} = 6$

Method	Dataset	BA	SE	SP	MCC	AUC	TP	FP	TN	FN	#
(S)SVM+(S)KNN	VS	0.91	0.93	0.90	0.72	0.93	207	117	1048	15	1387
	ES	0.72	0.66	0.77	0.34	0.73	55	107	365	29	556
BRF+(S)SVM	VS	0.91	0.95	0.87	0.69	0.95	215	153	1043	11	1422
	ES	0.71	0.60	0.81	0.33	0.72	53	102	454	35	644
BRF+(S)KNN	VS	0.91	0.94	0.88	0.68	0.95	208	152	1031	13	1404
	ES	0.72	0.68	0.76	0.33	0.73	53	112	348	25	538

Toxicity threshold $pIC_{50} = 5$

BRF + SVM	VS	0.87	0.89	0.86	0.75	0.93	619	108	608	74	1409
	ES	0.72	0.67	0.76	0.43	0.75	220	67	212	107	606
BRF + GB	VS	0.87	0.88	0.86	0.74	0.93	618	104	622	86	1430
	ES	0.70	0.66	0.74	0.41	0.73	216	71	212	120	619
SVM + GB	VS	0.87	0.88	0.86	0.74	0.93	614	103	603	78	1398
	ES	0.71	0.66	0.75	0.41	0.74	223	69	213	116	621

consensus strategies are BRF, (S)KNN, and (S)SVM. This is supported also by the corresponding AUC values, being 0.95, 0.92, and 0.89, respectively. Remarkably, the 5-fold CV ensures the internal robustness of the three models with BRF, (S)KNN, and (S)SVM reaching BA as high as 0.79, 0.78, and 0.76, respectively (Supplementary Table S6).

Table 3 shows the performances on the VS of the three consensus models developed by integrating each possible pair of the three top-performing classifiers indicated above. Notice that all of the consensus models return very high BAs (0.91), thus outperforming all of the single classifiers. Importantly, irrespective of the single models involved in the consensus strategy, only a small fraction of compounds (<11%) has been excluded by the prediction because of a discordant classification. These results put forward the considered consensus strategies as being extremely powerful to maximize the predictive performance of the models developed when $pIC_{50} = 6$ is considered the toxicity threshold. These consensus models were further challenged via a successive temporal validation using the ES. Although the performances were worse than those obtained on the VS, as already experienced in previous works (Sheridan, 2013; Bosc et al., 2019) and similar to the performance observed with the 5-fold CV procedure, the consensus models are responsible for satisfactory values of BA and AUC. It is noteworthy that (S)KNN+(S)SVM (BA = 0.72; AUC = 0.73) and BRF+(S)KNN (BA = 0.72; AUC = 0.73) returned a more balanced statistic, thus outperforming BRF+(S)SVM (BA = 0.71; AUC = 0.72) (Table 3).

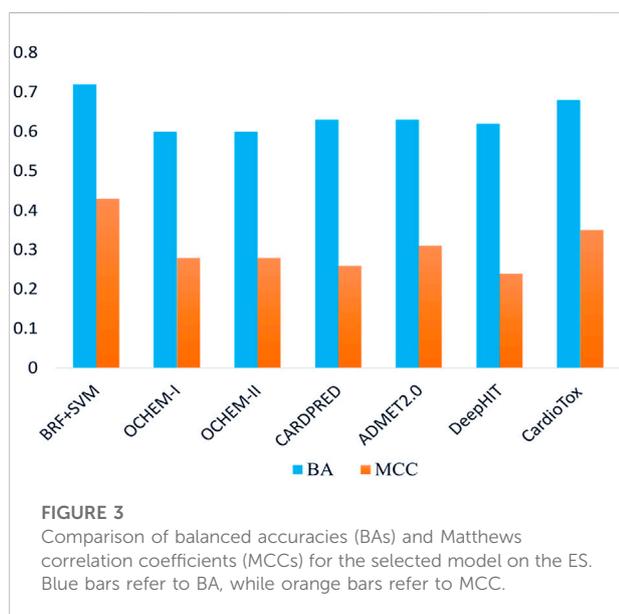
Models developed using $pIC_{50} = 5$ as toxicity threshold

Table 2 reports the performance returned by the validation performed on the VS for each model trained without using the SMOTE. Performance refers only to chemicals included in the AD (see section Applicability Domain). Similar BAs, ranging from 0.82 to 0.83, were obtained for all of the models, except for MLP responsible for the worst performance (BA = 0.80). Furthermore, all of the models return well-balanced statistics, with a difference between SP and SE ranging from 0.02 (XGB and GB) to 0.06 (MLP), consistent with the balanced composition of the TS. As expected, herein the SMOTE application did not lead to a significant performance improvement. Taken as a whole, these results suggest that, among all, the models developed using BRF, GB, and SVM as algorithms are the top-performing ones. This is evident looking at the computed BA and AUC values reported in Table 2. Importantly, the 5-fold CV confirms the internal robustness of these models (Supplementary Table S6) that were thus selected for further consensus and temporal validation procedures.

Again, the application of a consensus strategy led to a performance improvement in terms of both BA and AUC. In particular, Table 3 shows that all of the consensus models reach a BA equal to 0.87 and an AUC of 0.93, hence outperforming the BA_{MAX} (0.83) and AUC_{MAX} (0.92) obtained from the single models. Again, a small fraction of compounds (<11%) was

TABLE 4 Comparison in terms of performance on the ES (temporal validation) of the best performing model presented in this study (BRF + SVM) and different classifiers available in the literature. The following statistics are reported: balanced accuracy (BA), sensitivity (SE), specificity (SPE), Matthews correlation coefficient (MCC), and the total number of molecules (#).

	BRF + SVM	OCHEM-I	OCHEM-II	Cardprep	ADMET2.0	DeepHIT	CardioTox
BA	0.72	0.60	0.60	0.63	0.63	0.62	0.68
SE	0.67	0.24	0.24	0.66	0.89	0.80	0.70
SP	0.76	0.95	0.95	0.59	0.36	0.44	0.65
MCC	0.43	0.28	0.28	0.26	0.31	0.24	0.35
#	606	682	670	792	792	792	792



unpredicted as a consequence of a discordant classification between the involved models. Ultimately, the performed temporal validation (Table 3) confirms the overall good predictivity of the consensus classifiers, with the BRF + SVM ensuring the best BA and AUC values (BA = 0.72, AUC = 0.75).

Comparison with other classifiers available in the literature

The performances returned by the consensus models confirm that the integration of multiple strategies, applying a weight-of-evidence approach, leads to the detection and exclusion of erroneous predictions generated by the individual models, reinforcing, at the same time, those concordant. To the best of our knowledge, our consensus classifiers outperform on this VS (e.g., $AUC_{MAX} = 0.95$ and 0.93 when the toxicity threshold is $pIC_{50} = 6$ and 5 , respectively) all of the binary classifiers of hERG-related cardiotoxicity available in the literature and built using the

same toxicity thresholds. Encouraged by these pieces of evidence and aimed at performing a more detailed comparative analysis between our computational workflow and other ligand-based hERG-blockage predicting models available in the literature, the real-life application of six tools freely available and widely used for predicting the hERG blocking potential of chemicals, namely, Cardprep (Lee et al., 2019), ADMETlab (Xiong et al., 2021), OCHEM consensus models I and II (Li et al., 2017), DeepHIT (Ryu et al., 2020), and CardioTox (Karim et al., 2021) were challenged in temporal validation using the same ES herein employed. All of these tools are able to discern hERG blockers from nonblockers using as toxicity threshold $pIC_{50} = 5$ and, for this reason, their performances were compared with those returned by the best model (BRF + SVM) developed using the same threshold. Notice that our BRF + SVM model excluded ~23% of compounds (~5% being outside the AD and ~18% as a consequence of discordant predictions) while OCHEM I and II discarded ~14% compounds based on AD. Noteworthy, all of the other tools do not provide any AD-based filter to be applied. Furthermore, as the computation of AUC was not possible for all of the considered tools, being not able to provide a probability-based ranking, we used BA and MCC for the performance comparison. Remarkably, as shown in Table 4, the herein developed RF + SVM classifier showed more balanced statistics in predicting positive and negative samples than OCHEM I/II, ADMET 2.0, and DeepHIT models. In particular, OCHEM I/II presents a higher FN rate (SE = 0.24 and 0.38), whereas a higher FP rate characterizes ADMET 2.0 (SP = 0.36) and DeepHIT (SP = 0.44). As evident in Figure 3, BRF + SVM exhibits the highest BA and MCC (0.72 and 0.43, respectively) compared to all of the tested models, albeit excluding a fraction of compounds from the prediction, as previously mentioned. Hence, it can be reasonably considered as the best performing one in terms of real-life applicability.

QSAR-hERG: A freely accessible KNIME workflow

The top-performing QSAR models are available as a KNIME workflow at <https://github.com/PDelre93/hERG-QSAR>. The

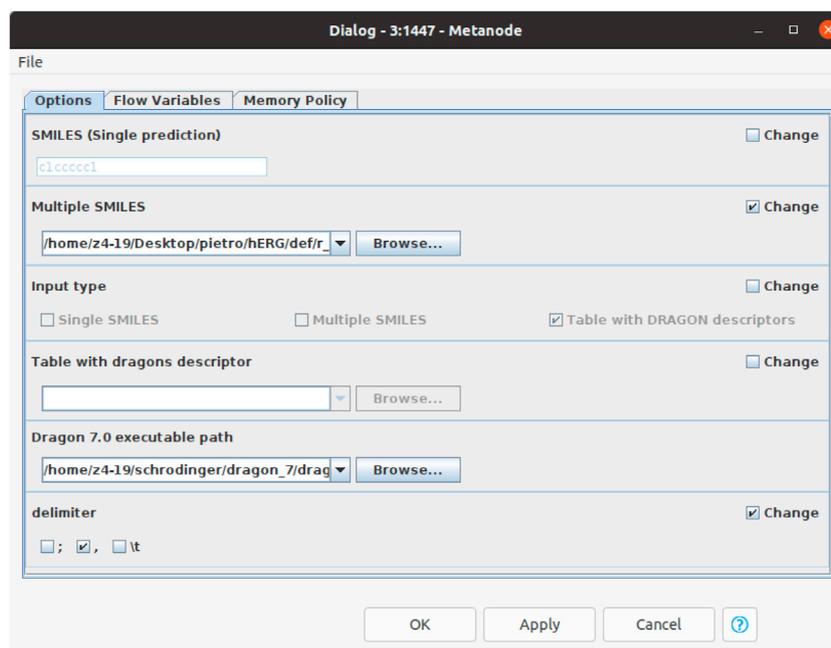


FIGURE 4
Dialog box to set up the calculation using the KNIME workflow.

S ID	S BRF...	S (S)SVM...	S (S)KNN...	S Applicability domain	I treshold (pIC50)
Mibefradil	ACT	ACT	ACT	TRUE	6
Terfenadine	ACT	ACT	ACT	TRUE	6
Sertindole	ACT	ACT	ACT	TRUE	6
S ID	S BRF...	S GB_Pre...	S SVM_p...	S Applicability domain	I treshold (pIC50)
Mibefradil	ACT	ACT	ACT	TRUE	5
Terfenadine	ACT	ACT	ACT	TRUE	5
Sertindole	ACT	ACT	ACT	TRUE	5

FIGURE 5
Output tables returned by the KNIME workflow for the three compounds examined: mibefradil, sertindole, and terfenadine.

implementation offers easy-to-use and intuitive options to use our predictive models. In the supporting information, a detailed guide explains how to install and perform fast hERG-related cardiotoxicity predictions. The graphical user interface (Figure 4) allows the users to choose the preferred way to proceed: 1) predict the activity of a single compound by manually entering the SMILES or 2) predict a batch of compounds from a SMILES list included in a .csv or .xlsx file. The workflow can automatically compute the required DRAGON descriptors (license is required). Alternatively, the user can include precalculated descriptors within the input file. The affinity toward the hERG channel is predicted using the top-performing models described above: BRF, (S)KNN, and (S)SVM for the activity thresholds $pIC_{50} =$

6; BRF, SVM, and GB for the activity threshold $pIC_{50} = 5$. At the end of the calculation, users can inspect the predictions generated by each model at $pIC_{50} = 6$ or 5, evaluating their reliability. Figure 5 shows the predictions returned by three compounds previously withdrawn from the market due to demonstrated hERG-related cardiotoxicity: mibefradil (Bezençon et al., 2017), sertindole (Sinha and Sen, 2011), and terfenadine (Sinha and Sen, 2011). Remarkably, all of the models predict the selected compounds as ACT (i.e., hERG-blockers). The column “applicability domain” expresses the reliability of the final prediction, which in our case is trustworthy (TRUE). It is worth noting that these findings agree with experimental data indicating pIC_{50} values of 6.24 for mibefradil (Bezençon et al.,

2017), 7.83 for sertindole (Sinha and Sen, 2011), and 6.67 for terfenafide (Sinha and Sen, 2011).

Conclusion

In this study, we developed 30 QSAR models based on 7,963 highly curated bioactivity data reported in ChEMBL (version 25) (Gaulton et al., 2012) and 1D and 2D descriptors computed by DRAGON 7.0.4 (Kode, 2017). By employing six machine learning algorithms, namely RF (Breiman, 2001), KNN (Altman, 1992), GB (Friedman, 2001), XGB (Chen and Guestrin, 2016), MLP (Haykin, 1994), and SVM (Vapnik, 1963), we implemented two sets of binary models differing for the considered toxicity threshold ($pIC_{50} = 6$ or $pIC_{50} = 5$). To maximize the performances, we followed three strategies for building ligand-based classifiers, namely: 1) VSURF (Genuer et al., 2010), to select relevant features to use in model construction, 2) the oversampling technique SMOTE (Chawla et al., 2002) to handle the unbalanced data; 3) a consensus approach to overcome single model limitations. Remarkably, the obtained results highlight the usefulness of these strategies, as testified by the high performances returned in the validation procedure. Importantly the performed temporal validation confirms the reliability of our models in real-life cases, given their ability to properly classify as hERG blockers or nonblocker compounds belonging to a repository (ChEMBL (Gaulton et al., 2012) v28) published after the data used for building TS and VS (ChEMBL (Gaulton et al., 2012) v25). Noteworthy, the models can be efficiently used in combination with structure-based strategies (Creanza et al., 2021) as testified by recent literature (Mansouri et al., 2016; Kamel et al., 2017). Finally, the performed comparative analysis indicates that the top-performing consensus model herein developed outperforms several commonly employed classifiers available in the literature. Our computational workflow is available to the cheminformatics community in the GitHub repository (<https://github.com/PDelre93/hERG-QSAR>), as valuable for a robust ligand-based prediction tool of hERG-related cardiotoxicity.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, and further inquiries can be directed to the corresponding authors.

References

Afantitis, A., Melagraki, G., Sarimveis, H., Koutentis, P. A., Markopoulos, J., and Igglessi-Markopoulou, O. (2008). Development and evaluation of a QSPR model for the prediction of diamagnetic susceptibility. *QSAR Comb. Sci.* 27, 432–436. doi:10.1002/qsar.200730083

Author contributions

Conception and design of the work: PD, EB, GM, and DG. Acquisition of data: PD and GM. Analysis and interpretation of data: PD, GJL, and DG. Creation of new software used in the work: PD and DG. Drafting of the work: PD and GM. Substantial revision of the work: GJL, GL, MS, AR, EB, and DG. All authors have approved the submitted version and agreed both to be personally accountable for the author's contributions and to ensure that questions related to the accuracy or integrity of any part of the work are appropriately investigated, resolved, and the resolution documented in the literature.

Funding

This work was supported by the European Union's Horizon 2020 research and innovation program (grant # 101037090). The content of this manuscript reflects only the author's view, and the Commission is not responsible for any use that may be made of the information it contains. The PhD fellowship of Dr Giuseppe Lamanna was co-funded by Chiesi Farmaceutici S. p. a. under the program "Dottorato Industriale CNR—XXXVI ciclo".

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2022.951083/full#supplementary-material>

Alberga, D., Trisciuzzi, D., Montaruli, M., Leonetti, F., Mangiatordi, G. F., and Nicolotti, O. (2019). A new approach for drug target and bioactivity prediction: The multifingerprint similarity search algorithm (MuSSeL). *J. Chem. Inf. Model.* 59, 586–596. doi:10.1021/acs.jcim.8b00698

- Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Statistician* 46, 175–185. doi:10.1080/00031305.1992.10475879
- Baderna, D., Gadaleta, D., Lostaglio, E., Selvestrel, G., Raitano, G., Golbamaki, A., et al. (2020). New *in silico* models to predict *in vitro* micronucleus induction as marker of genotoxicity. *J. Hazard. Mat.* 385, 121638. doi:10.1016/j.jhazmat.2019.121638
- Berthold, M. R., Cebon, N., Dill, F., Gabriel, T. R., Kötter, T., Meinel, T., et al. (2008). “Knime: The konstanz information miner,” in *in data analysis, machine Learning and applications studies in classification, data analysis, and knowledge organization*. Editors C. Preisach, H. Burkhardt, L. Schmidt-Thieme, and R. Decker (Berlin, Heidelberg: Springer), 319–326. doi:10.1007/978-3-540-78246-9_38
- Bezençon, O., Heidmann, B., Siegrist, R., Stamm, S., Richard, S., Pozzi, D., et al. (2017). Discovery of a potent, selective T-type calcium channel blocker as a drug candidate for the treatment of generalized epilepsies. *J. Med. Chem.* 60, 9769–9789. doi:10.1021/acs.jmedchem.7b01236
- Bosc, N., Atkinson, F., Felix, E., Gaulton, A., Hersey, A., and Leach, A. R. (2019). Large scale comparison of QSAR and conformal prediction methods and their applications in drug discovery. *J. Cheminform.* 11, 4. doi:10.1186/s13321-018-0325-4
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi:10.1023/A:1010933404324
- Butler, A., Helliwell, M. V., Zhang, Y., Hancox, J. C., and Dempsey, C. E. (2020). An update on the structure of hERG. *Front. Pharmacol.* 10, 1572. doi:10.3389/fphar.2019.01572
- Cai, C., Guo, P., Zhou, Y., Zhou, J., Wang, Q., Zhang, F., et al. (2019). Deep learning-based prediction of drug-induced cardiotoxicity. *J. Chem. Inf. Model.* 59, 1073–1084. doi:10.1021/acs.jcim.8b00769
- Cavalluzzi, M. M., Imbrici, P., Gualdani, R., Stefanachi, A., Mangiatordi, G. F., Lentini, G., et al. (2020). Human ether-à-go-go-related potassium channel: Exploring SAR to improve drug design. *Drug Discov. Today* 25, 344–366. doi:10.1016/j.drudis.2019.11.005
- CDK (2022). *Chemistry development kit*. Available at: <https://cdk.github.io/>
- Chavan, S., Abdelaziz, A., Wiklander, J. G., and Nicholls, I. A. (2016). A k-nearest neighbor classification of hERG K(+) channel blockers. *J. Comput. Aided. Mol. Des.* 30, 229–236. doi:10.1007/s10822-016-9898-z
- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* 16, 321–357. doi:10.1613/jair.953
- Chen, T., and Guestrin, C. (2016). “XGBoost: A scalable tree boosting system,” in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. doi:10.1145/2939672.2939785
- Choi, K.-E., Balupuri, A., and Kang, N. S. (2020). The study on the hERG blocker prediction using chemical fingerprint analysis. *Molecules* 25, 2615. doi:10.3390/molecules25112615
- Creanza, T. M., Delre, P., Ancona, N., Lentini, G., Saviano, M., and Mangiatordi, G. F. (2021). Structure-based prediction of hERG-related cardiotoxicity: A benchmark study. *J. Chem. Inf. Model.* 61, 4758–4770. doi:10.1021/acs.jcim.1c00744
- Danker, T., and Möller, C. (2014). Early identification of hERG liability in drug discovery programs by automated patch clamp. *Front. Pharmacol.* 5, 203. doi:10.3389/fphar.2014.00203
- Ema (2005). *ICH topic S7B the nonclinical evaluation of the potential for delayed ventricular repolarization (QT interval prolongation) by human Pharmaceuticals*.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognit. Lett.* 27, 861–874. doi:10.1016/j.patrec.2005.10.010
- FDA (2005). *S7B nonclinical evaluation of the potential for delayed ventricular repolarization (QT interval prolongation) by human Pharmaceuticals*.
- Ferdinandy, P., Baczkó, I., Bencsik, P., Giricz, Z., Görbe, A., Pacher, P., et al. (2019). Definition of hidden drug cardiotoxicity: Paradigm change in cardiac safety testing and its clinical implications. *Eur. Heart J.* 40, 1771–1777. doi:10.1093/eurheartj/ehy365
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Ann. Stat.* 29, 1189–1232. doi:10.1214/aos/1013203451
- Gadaleta, D., Lombardo, A., Toma, C., and Benfenati, E. (2018a). A new semi-automated workflow for chemical data retrieval and quality checking for modeling applications. *J. Cheminform.* 10, 60. doi:10.1186/s13321-018-0315-6
- Gadaleta, D., Manganelli, S., Roncaglioni, A., Toma, C., Benfenati, E., and Mombelli, E. (2018b). QSAR modeling of ToxCast assays relevant to the molecular initiating events of AOPs leading to hepatic steatosis. *J. Chem. Inf. Model.* 58, 1501–1517. doi:10.1021/acs.jcim.8b00297
- Gadaleta, D., Mangiatordi, G. F., Catto, M., Carotti, A., and Nicolotti, O. (2016). Applicability domain for QSAR models: Where theory meets reality. *IJQSPR* 1, 45–63. doi:10.4018/IJQSPR.2016010102
- Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., et al. (2012). ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* 40, D1100–D1107. doi:10.1093/nar/gkr777
- Genuer, R., Poggi, J.-M., and Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognit. Lett.* 31, 2225–2236. doi:10.1016/j.patrec.2010.03.014
- Gobbi, M., Beeg, M., Toropova, M. A., Toropov, A. A., and Salmona, M. (2016). Monte Carlo method for predicting of cardiac toxicity: hERG blocker compounds. *Toxicol. Lett.* 251, 42–46. doi:10.1016/j.toxlet.2016.04.010
- Haykin, S. (1994). *Neural networks: A comprehensive foundation*. 1st ed. USA: Prentice Hall PTR.
- Hong, H.-K., Hoon Lee, B., Park, M.-H., Ho Lee, S., Chu, D., Jin Kim, W., et al. (2013). Block of hERG K+ channel and prolongation of action potential duration by fluphenazine at submicromolar concentration. *Eur. J. Pharmacol.* 702, 165–173. doi:10.1016/j.ejphar.2013.01.039
- Jing, Y., Easter, A., Peters, D., Kim, N., and Enyedy, I. J. (2015). *In silico* prediction of hERG inhibition. *Future Med. Chem.* 7, 571–586. doi:10.4155/fmc.15.18
- Jolliffe, I. T., and Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philos. Trans. A Math. Phys. Eng. Sci.* 374, 20150202. doi:10.1098/rsta.2015.0202
- Kalyanamoorthy, S., and Barakat, K. H. (2018). Development of safe drugs: The hERG challenge. *Med. Res. Rev.* 38, 525–555. doi:10.1002/med.21445
- Kamel, M., Kleinstreuer, N., Watt, E., Harris, J., and Judson, R. (2017). CoMPARA: Collaborative modeling project for androgen receptor activity. *Environ. Health Perspect.* 128 (2), 27002. doi:10.13140/RG.2.2.16791.78241
- Kamiya, K., Niwa, R., Morishima, M., Honjo, H., and Sanguinetti, M. C. (2008). Molecular determinants of hERG channel block by terfenadine and cisapride. *J. Pharmacol. Sci.* 108, 301–307. doi:10.1254/jphs.08102fp
- Kar, S., Roy, K., and Leszczynski, J. (2018). “Applicability domain: A step toward confident predictions and decidability for QSAR modeling,” in *Computational Toxicology: Methods and protocols molecular biology*. Editor O. Nicolotti (New York, NY: Springer), 141–169. doi:10.1007/978-1-4939-7899-1_6
- Karim, A., Lee, M., Balle, T., and Sattar, A. (2021). CardioTox net: A robust predictor for hERG channel blockade based on deep learning meta-feature ensembles. *J. Cheminform.* 13, 60. doi:10.1186/s13321-021-00541-z
- Katchman, A. N., Koerner, J., Tosaka, T., Woosley, R. L., and Ebert, S. N. (2006). Comparative evaluation of hERG currents and QT intervals following challenge with suspected torsadogenic and nontorsadogenic drugs. *J. Pharmacol. Exp. Ther.* 316, 1098–1106. doi:10.1124/jpet.105.093393
- Kim, Y. J., Hong, H. K., Lee, H. S., Moh, S. H., Park, J. C., Jo, S. H., et al. (2008). Papaverine, a vasodilator, blocks the pore of the hERG channel at submicromolar concentration. *J. Cardiovasc. Pharmacol.* 52, 485–493. doi:10.1097/FJC.0b013e31818e65c2
- Kode (2017). *Dragon 7.0, 8*. Available at: https://chm.kode-solutions.net/products_dragon.php.
- Krishna, S., Borrel, A., Huang, R., Zhao, J., Xia, M., and Kleinstreuer, N. (2022). High-throughput chemical screening and structure-based models to predict hERG inhibition. *Biol. (Basel)* 11, 209. doi:10.3390/biology11020209
- Landrum, G., Tosco, P., Kelley, B., Vianello, R., Schneider, N., Kawashima, E., et al. (2022). *rdkit/rdkit: 2021_09_5 (Q3 2021) Release*. Zenodo. doi:10.5281/zenodo.6330241
- Lavado, G. J., Gadaleta, D., Toma, C., Golbamaki, A., Toropov, A. A., Toropova, A. P., et al. (2020). Zebrafish AC50 modelling: (Q)SAR models to predict developmental toxicity in zebrafish embryo. *Ecotoxicol. Environ. Saf.* 202, 110936. doi:10.1016/j.ecoenv.2020.110936
- LaValle, S. M., Branicky, M. S., and Lindemann, S. R. (2004). On the relationship between classical grid search and probabilistic roadmaps. *Int. J. Robotics Res.* 23, 673–692. doi:10.1177/0278364904045481
- Lee, H.-M., Yu, M.-S., Kazmi, S. R., Oh, S. Y., Rhee, K.-H., Bae, M.-A., et al. (2019). Computational determination of hERG-related cardiotoxicity of drug candidates. *BMC Bioinforma.* 20, 250. doi:10.1186/s12859-019-2814-5
- Li, X., Zhang, Y., Li, H., and Zhao, Y. (2017). Modeling of the hERG K+ channel blockage using online chemical database and modeling environment (OCHEM). *Mol. Inf.* 36, 1700074. doi:10.1002/minf.201700074
- Liu, M., Zhang, L., Li, S., Yang, T., Liu, L., Zhao, J., et al. (2020). Prediction of hERG potassium channel blockage using ensemble learning methods and molecular fingerprints. *Toxicol. Lett.* 332, 88–96. doi:10.1016/j.toxlet.2020.07.003

- Mansouri, K., Abdelaziz, A., Rybacka, A., Roncaglioni, A., Tropsha, A., Varnek, A., et al. (2016). Cerapp: Collaborative estrogen receptor activity prediction project. *Environ. Health Perspect.* 124, 1023–1033. doi:10.1289/ehp.1510267
- Melagraki, G., Afantitis, A., Sarimveis, H., Koutentis, P. A., Kollias, G., and Igglessi-Markopoulou, O. (2009). Predictive QSAR workflow for the *in silico* identification and screening of novel HDAC inhibitors. *Mol. Divers.* 13, 301–311. doi:10.1007/s11030-009-9115-2
- Munawar, S., Windley, M. J., Tse, E. G., Todd, M. H., Hill, A. P., Vandenberg, J. I., et al. (2018). Experimentally validated pharmacoinformatics approach to predict hERG inhibition potential of new chemical entities. *Front. Pharmacol.* 9, 1035. doi:10.3389/fphar.2018.01035
- O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., and Hutchison, G. R. (2011). Open Babel: An open chemical toolbox. *J. Cheminform.* 3, 33. doi:10.1186/1758-2946-3-33
- Onakpoya, I. J., Heneghan, C. J., and Aronson, J. K. (2016). Post-marketing withdrawal of 462 medicinal products because of adverse drug reactions: A systematic review of the world literature. *BMC Med.* 14, 10. doi:10.1186/s12916-016-0553-2
- Priest, B. T., Bell, I. M., and Garcia, M. L. (2008). Role of hERG potassium channel assays in drug development. *Channels (Austin)* 2, 87–93. doi:10.4161/chan.2.2.6004
- Refaeilzadeh, P., Tang, L., and Liu, H. (2009). "Cross-validation," in *Encyclopedia of database systems*. Editors L. Liu and M. T. Özsu (Boston, MA: Springer US), 532–538. doi:10.1007/978-0-387-39940-9_565
- Rogers, D., and Hahn, M. (2010). Extended-connectivity fingerprints. *J. Chem. Inf. Model.* 50, 742–754. doi:10.1021/ci100050t
- Roy, K., Kar, S., and Das, R. N. (2015). "Chapter 7 - validation of QSAR models," in *Understanding the basics of QSAR for applications in pharmaceutical Sciences and risk assessment*. Editors K. Roy, S. Kar, and R. N. Das (Boston: Academic Press), 231–289. doi:10.1016/B978-0-12-801505-6.00007-7
- Ryu, J. Y., Lee, M. Y., Lee, J. H., Lee, B. H., and Oh, K.-S. (2020). DeepHIT: A deep learning framework for prediction of hERG-induced cardiotoxicity. *Bioinformatics* 36, 3049–3055. doi:10.1093/bioinformatics/btaa075
- Sheridan, R. P. (2013). Time-split cross-validation as a method for estimating the goodness of prospective prediction. *J. Chem. Inf. Model.* 53, 783–790. doi:10.1021/ci400084k
- Sinha, N., and Sen, S. (2011). Predicting hERG activities of compounds from their 3D structures: Development and evaluation of a global descriptors based QSAR model. *Eur. J. Med. Chem.* 46, 618–630. doi:10.1016/j.ejmech.2010.11.042
- Siramshetty, V. B., Chen, Q., Devarakonda, P., and Preissner, R. (2018). The catch-22 of predicting hERG blockade using publicly accessible bioactivity data. *J. Chem. Inf. Model.* 58, 1224–1233. doi:10.1021/acs.jcim.8b00150
- Siramshetty, V. B., Nguyen, D.-T., Martinez, N. J., Southall, N. T., Simeonov, A., and Zakharov, A. V. (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. doi:10.1021/acs.jcim.0c00884
- Slavov, S., Stoyanova-Slavova, I., Li, S., Zhao, J., Huang, R., Xia, M., et al. (2017). Why are most phospholipidosis inducers also hERG blockers? *Arch. Toxicol.* 91, 3885–3895. doi:10.1007/s00204-017-1995-9
- Snoek, J., Larochelle, H., and Adams, R. P. (2012). *Practical bayesian optimization of machine learning algorithms*. arXiv:1206.2944 [cs, stat]. Available at: <http://arxiv.org/abs/1206.2944> (Accessed March 16, 2022).
- Steinbeck, C., Han, Y., Kuhn, S., Horlacher, O., Luttmann, E., and Willighagen, E. (2003). The chemistry development kit (CDK): An open-source java library for chemo- and bioinformatics. *J. Chem. Inf. Comput. Sci.* 43, 493–500. doi:10.1021/ci025584y
- Su, Z., Chen, J., Martin, R. L., McDermott, J. S., Cox, B. F., Gopalakrishnan, M., et al. (2006). Block of hERG channel by ziprasidone: Biophysical properties and molecular determinants. *Biochem. Pharmacol.* 71, 278–286. doi:10.1016/j.bcp.2005.10.047
- Sun, H., Huang, R., Xia, M., Shahane, S., Southall, N., and Wang, Y. (2017). Prediction of hERG liability - using SVM classification, bootstrapping and jackknifing. *Mol. Inf.* 36, 1600126. doi:10.1002/minf.201600126
- Tanaka, H., Takahashi, Y., Hamaguchi, S., Iida-Tanaka, N., Oka, T., Nishio, M., et al. (2014). Effect of terfenadine and pentamidine on the hERG channel and its intracellular trafficking: Combined analysis with automated voltage clamp and confocal microscopy. *Biol. Pharm. Bull.* 37, 1826–1830. doi:10.1248/bpb.b14-00417
- Ting, K. M. (2017). "Confusion matrix," in *Encyclopedia of machine learning and data mining*. Editors C. Sammut and G. I. Webb (Boston, MA: Springer US), 260. doi:10.1007/978-1-4899-7687-1_50
- Tropsha, A., Gramatica, P., and Gombar, V. K. (2003). The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb. Sci.* 22, 69–77. doi:10.1002/qsar.200390007
- Vapnik, V. (1963). Pattern recognition using generalized portrait method. *Automation Remote Control* 24, 774–780.
- Villoutreix, B. O., and Taboureau, O. (2015). Computational investigations of hERG channel blockers: New insights and current predictive models. *Adv. Drug Deliv. Rev.* 86, 72–82. doi:10.1016/j.addr.2015.03.003
- Walker, B. D., Singleton, C. B., Bursill, J. A., Wyse, K. R., Valenzuela, S. M., Qiu, M. R., et al. (1999). Inhibition of the human ether-a-go-go-related gene (HERG) potassium channel by cisapride: Affinity for open and inactivated states. *Br. J. Pharmacol.* 128, 444–450. doi:10.1038/sj.bjp.0702774
- Wang, S., Sun, H., Liu, H., Li, D., Li, Y., and Hou, T. (2016). ADMET evaluation in drug discovery. 16. Predicting hERG blockers by combining multiple pharmacophores and machine learning approaches. *Mol. Pharm.* 13, 2855–2866. doi:10.1021/acs.molpharmaceut.6b00471
- Willett, P., Barnard, J. M., and Downs, G. M. (1998). Chemical similarity searching. *J. Chem. Inf. Comput. Sci.* 38, 983–996. doi:10.1021/ci9800211
- Xiong, G., Wu, Z., Yi, J., Fu, L., Yang, Z., Hsieh, C., et al. (2021). ADMETlab 2.0: An integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic Acids Res.* 49, W5–W14. doi:10.1093/nar/gkab255
- Zhang, C., Zhou, Y., Gu, S., Wu, Z., Wu, W., Liu, C., et al. (2016). *In silico* prediction of hERG potassium channel blockade by chemical category approaches. *Toxicol. Res.* 5, 570–582. doi:10.1039/C5TX00294J
- Zhang, H., Wu, W., Li, S. P. G., and Zhang, H. (2018). "A comparative analysis of convergence rate for imbalanced datasets of active learning models," in *2018 IEEE 23rd international conference on digital signal processing (DSP)*, 1–5. doi:10.1109/ICDSP.2018.8631877
- Zhou, Z., Vorperian, V. R., Gong, Q., Zhang, S., and January, C. T. (1999). Block of HERG potassium channels by the antihistamine astemizole and its metabolites desmethylastemizole and norastemizole. *J. Cardiovasc. Electrophysiol.* 10, 836–843. doi:10.1111/j.1540-8167.1999.tb00264.x
- Zhu, Q., and Lu, P. (2020). Stem cell transplantation for amyotrophic lateral sclerosis. *Adv. Exp. Med. Biol.* 136, 71–97. doi:10.1007/978-981-15-4370-8_6
- Zolotoy, A. B., Plouvier, B. P., Beatch, G. B., Hayes, E. S., Wall, R. A., and Walker, M. J. A. (2003). Physicochemical determinants for drug induced blockade of HERG potassium channels: Effect of charge and charge shielding. *Curr. Med. Chem. - Cardiovasc. Hematological Agents* 1, 225–241. doi:10.2174/1568016033477432