npj | digital medicine

Article

Published in partnership with Seoul National University Bundang Hospital

https://doi.org/10.1038/s41746-024-01370-8

Fast and accurate prediction of drug induced proarrhythmic risk with sex specific cardiac emulators

Paula Dominguez-Gomez[1,2,5] ✉, Alberto Zingaro[1,5], Laura Baldo-Canut[1,5], Caterina Balzotti[1,5], Borje Darpo[3], Christopher Morton[1], Mariano Vázquez[1,4] & Jazmin Aguado-Sierra[1]

In silico trials for drug safety assessment require many high-fidelity 3D cardiac simulations to predict drug-induced QT interval prolongation, which is often computationally prohibitive. To streamline this process, we developed sex-specific emulators for a fast prediction of QT interval, trained on a dataset of 900 simulations. Our results show significant differences between 3D and 0D single-cell models as risk levels increase, underscoring the ability of 3D modeling to capture more complex cardiac responses. The emulators demonstrated an average error of 4% compared to simulations, allowing for efficient global sensitivity analysis and fast replication of in silico clinical trials. This approach enables rapid, multi-dose drug testing on standard hardware, addressing critical industry challenges around trial design, assay variability, and cost-effective safety evaluations. By integrating these emulators into drug development, we can improve preclinical reliability and advance the practical application of digital twins in biomedicine.

Arrhythmias manifest as irregular heart rhythms stemming from abnormal cardiac electrical activity, ranging from benign palpitations to severe conditions like ventricular fibrillation[1]. A pivotal biomarker associated with increased arrhythmic risk is a prolonged QT interval, as measured on electrocardiograms (ECGs) from the beginning of the QRS complex to the end of the T wave. The change from baseline QT, the ΔQT, is notably linked to torsades de pointes (TdP), a potentially lethal, polymorphic ventricular tachycardia[2]. Prolongation of the QT interval, caused by many drugs beyond cardiac treatments, often arises from unintended ion channel blockades such as the hERG potassium channel, crucial for cardiac repolarization. Such blockade delays repolarization, thereby prolonging the QT interval. The gold standard for assessing the proarrhythmic risk of a drug traditionally involves measuring its effects on the hERG channel in vitro, a single-dose safety pharmacology study in dogs or monkeys, and evaluation of potential effects on the QT interval in clinical trials. Despite substantial research on this topic[3–6], the precise relationship between drug-induced ion channel blockade at the cellular level and QT prolongation at the organ level remains poorly understood, complicating risk assessments.

Regulatory agencies, such as the US Food and Drug Administration (FDA) and the European Medicines Agency, have recognized the potential of computational models and digital twins to enhance drug safety evaluations, as they offer a promising approach to bridge the gaps in the translation of nonclinical assays to clinical outcomes in the drug discovery phase and in the early preclinical stage[6]. The Comprehensive in Vitro Proarrhythmia Assay (CiPA) initiative, for instance, promotes the use of in silico models alongside in vitro and clinical data to complement proarrhythmic risk assessment[7].

Computational models for predicting proarrhythmic risk vary in complexity[8]. At the simplest level, single-cell 0D models (zero dimensional, where the spatial dependence of variables is neglected in favor of time dependence only) simulate the electrical activity of individual cardiac cells, providing insights into ion channel behavior and drug effects. However, these 0D models cannot capture QT interval prolongation as they do not account for the holistic behavior of the heart. Given that QT interval measurement is the gold standard for assessing proarrhythmic risk in clinical practice, employing 3D cardiac models for in silico clinical trials is essential. These models account for the intricate anatomy of the heart and simulate cardiac function at the organ level, providing a more comprehensive and accurate representation of drug-induced effects on cardiac electrophysiology. Several studies have predicted proarrhythmic risk using 0D models[9–14], while others have advanced to 3D representations of the heart for assessing ΔQT[15–21]. However, gaps persist in the literature in systematically determining the necessity of 3D models for accurate drug-induced proarrhythmic risk prediction compared to 0D single-cell models.

[1]ELEM Biotech S.L., Pier 07, Via Laietana, 26, Barcelona, 08003, Spain. [2]University Pompeu Fabra, Carrer de Tànger, 122-140, Barcelona, 08018, Spain. [3]Clario, 1818 Market St Suite 2600, Philadelphia, 19103, USA. [4]Barcelona Supercomputing Center, Plaça d'Eusebi Güell, 1–3, Barcelona, 08034, Spain. [5]These authors contributed equally: Paula Dominguez-Gomez, Alberto Zingaro, Laura Baldo-Canut, Caterina Balzotti. ✉e-mail: pdominguez@elem.bio

In addition, despite several advancements in 3D cardiac models, very few models account for sex-specific differences[21–24], even though substantial evidence indicates that female are more susceptible to arrhythmias[25]. While incorporating sex differences may be less critical for drugs causing only mild QT prolongation at therapeutic doses, it becomes essential for evaluating drugs with a high risk of TdP or substantially differing absorption rates between sexes. This gap underscores the urgent need to integrate anatomical and phenotypical sex differences into proarrhythmic risk predictions.

Despite their high accuracy, detailed cardiac 3D models can pose significant computational challenges for application in large-scale or real-time contexts. For drug discovery and at the preclinical level, the ability to predict ΔQT in real-time is crucial for gaining an early understanding of a drug's safety profile, identifying safety margins, and streamlining subsequent development stages. This capability facilitates protocol modifications, dose regimen adjustments, or prompt discontinuation of problematic drugs, thereby optimizing resource allocation and ensuring that only the most promising drug candidates move forward. Advancing the concept of digital twins in biomedicine requires accelerating numerical simulations to achieve real-time virtual representations of living systems[26]. In the context of clinical trials, such advancements could enable rapid, real-time screening methods, particularly during the preclinical stage. These digital twins could act as fast-response systems for designing computational trials, allowing for faster and more efficient evaluations of drug safety and efficacy.

To address these challenges, surrogate models based on artificial intelligence methods, offer a promising solution. These surrogate models, or emulators, approximate the outputs of high-fidelity simulations at a fraction of the computational cost. Previous works have demonstrated the utility of emulators in cardiac modeling[27–35] and drug safety assessment[36,37]. Notably, Costabal et al.[36] have pioneered the development of an emulator for ΔQT prediction, trained on a combination of 3D and 1D electrophysiological simulations.

In this study, we build on our previous work[23] by developing two emulators for real-time proarrhythmic risk assessment, specifically for ΔQT estimation. Each emulator is tailored for each sex, and is trained exclusively using data from 3D simulations. This process involved 900 electrophysiological runs, requiring approximately 2.1 million CPU hours globally. Prior to developing the emulators, we compared results from the 3D and 0D models. Our analysis demonstrated that 3D modeling is significantly more sensitive than 0D modeling in capturing the onset of abnormal electrical propagation in the context of proarrhythmic risk assessment for drugs, especially as the risk level increases. Leveraging these insights, we developed the emulators which achieved high accuracy with an average relative error of less than 4% compared to the simulator results, and a computational speed-up of five orders of magnitude. We then applied the emulators to replicate clinical trials for four benchmark drugs, comparing their predictions with both simulator results from different anatomies and real clinical data. This comparison validated the accuracy and generalizability of the emulators, demonstrating also their reliability in practical scenarios.

Results

We present a schematic overview of the general methodology used to build our emulators for real-time cardiac ΔQT prediction (see Fig. 1). The emulators are generated using data derived from 3D high-fidelity electrophysiological simulations (i.e., obtained with the simulator). To simulate the drugs' effect, we sample blockades of the seven most relevant ionic channels for cardiac proarrhythmic assessment[38]: I_{CaL}, I_{NaL}, I_{to}, I_{Ks}, I_{K1}, I_{Na}, and I_{Kr}. We account for sex differences by employing detailed biventricular geometries for both male and female patients, and include respective phenotypes[23]. Using

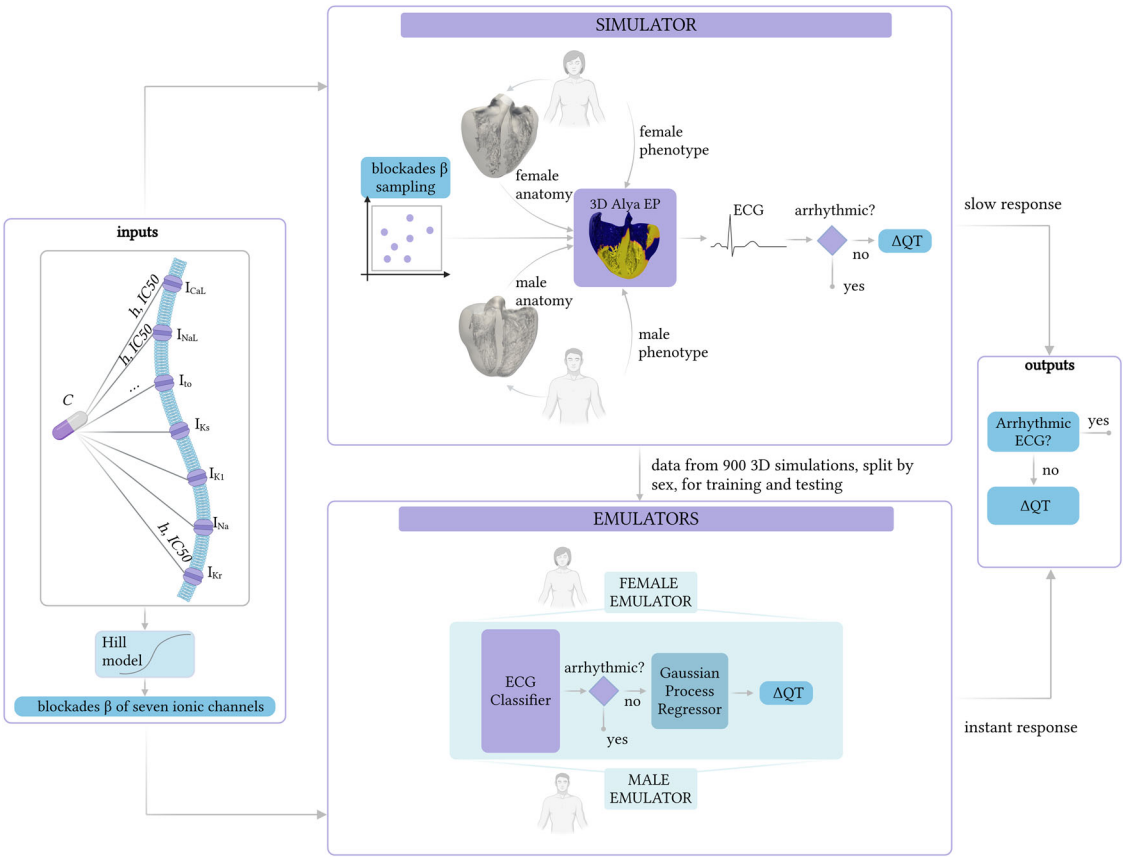

Fig. 1 | Overview of the methodology employed to build the ΔQT emulators. This methodology uses a simulator and sex-specific emulators to predict QT prolongations based on blockade levels of seven ionic channels, derived from the Hill model. The simulator performs 3D cardiac electrophysiological simulations for male and female anatomies, generating ECG data to classify arrhythmia and producing ΔQT outputs for training the emulators. The emulators consist of an ECG classifier followed by a Gaussian Process Regression model for fast ΔQT predictions.

Alya, our finite element simulator[39], we performed 3D simulations to compute the ECG and to measure QT interval prolongation relative to a baseline configuration (ΔQT). This data is then used to train our emulators. We develop two separate emulators, one for males and one for females, each predicting ΔQT in real-time based on the blockades of the seven ionic channels. Each emulator comprises a classifier and a Gaussian process regression (GPR) model. The former classifies the data into non-arrhythmic or arrhythmic; the latter predicts ΔQT. With the term "arrhythmic", in the context of these emulators, we refer to cases where the ΔQT cannot be computed due to an abnormal electrical propagation pattern in the ECG or where the ΔQT computed in silico exceeds a threshold[40]. The last criterion excludes cases producing values much larger than those typically measured in clinical practice and associated with the presence of arrhythmic events.

Simulator results

In Fig. 2, we present the simulation data that we obtain, and that we then use to train our emulators. This dataset comprises 900 simulations, with 450 simulations per sex, encompassing both arrhythmic and non-arrhythmic cases. To create our dataset, we excluded all the cases in which the 0D O'Hara-Rudy (ORd) initialization model[41] did not converge, i.e., it did not produce a periodic solution in terms of calcium concentration of each cell type (0 cases for males, 37 for females). Figure 2a displays the ECG signals, while Fig. 2b presents the distributions resultant from the computation of ΔQT values for the entire dataset. Notably, the presence of arrhythmia precludes the computation of ΔQT; consequently, these cases are not included in the ΔQT distributions but their ECG patterns are depicted in grey in Fig. 2a. The results in this plot suggest that, on average, women tend to have longer QT interval durations. Additionally, Fig. 2b shows a broader range of ΔQT values for men, likely because women are more prone to developing arrhythmic behavior under the same drug conditions. As a result, their data are more often excluded from the distribution. Figure 3 showcases examples for both males and females, highlighting differences through electrical propagation maps and corresponding ECG patterns based on the presence and absence of arrhythmias. It is worth

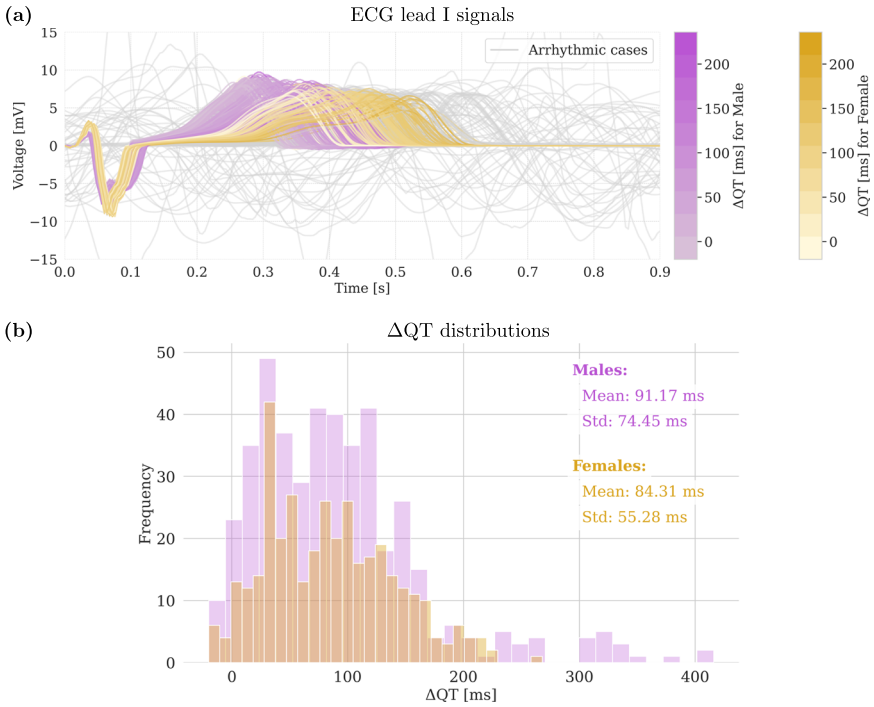

Fig. 2 | ECG signals and ΔQT distributions computed with the simulator. a ECG lead I signals for male and female subjects, with grey signals representing arrhythmic cases. **b** ΔQT distributions for male and female subjects.

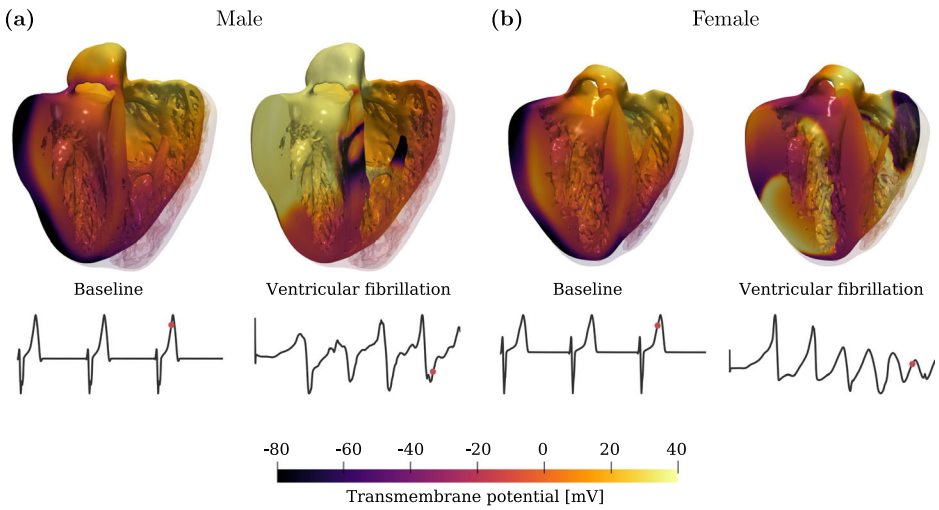

Fig. 3 | Examples of 3D electrophysiological simulations obtained with detailed biventricular anatomies. Both for the male (**a**) and female (**b**) anatomies, we report on the left the baseline simulation (without drug), and on the right the simulation in ventricular fibrillation conditions (due to drug). The signals represent the computed ECG lead I for each simulation and the red point on the ECG curve denotes the time in which the 3D images are taken.

noting that 11% of the female cohort exhibited arrhythmic behavior, compared to only 2% of the male cohort. The higher prevalence of non-convergent and arrhythmic cases among female subjects resulted in a reduced female sample size, accounting for the difference in distribution areas between males and females in Fig. 2b.

Modeling proarrhythmic risk in 0D and 3D

3D cardiac models, while computationally expensive, capture spatial details and allow for the computation of ΔQT, aligning with clinical outcomes. In contrast, 0D models rely on metrics that cannot be measured clinically, such as the action potential duration (APD) or qNet (computational biomarker based on the sum of the net ionic currents over the course of the action potential[42]). To assess the value added by 3D models, we analyze the relationship between 3D and 0D simulation outcomes for evaluating proarrhythmic risk, as depicted in Figs. 4 and 5. The 0D results are obtained from the initialization of the 3D simulation with the 0D ORd model run in endocardial, mid-myocardial, and epicardial cells.

Figure 4 illustrates a comparison of biomarkers from 3D and 0D simulations, with Fig. 4a and b depicting the relationship between ΔAPD (i.e., the difference between APD after and before drug administration, computed for different cell types) and ΔQT for females and males, respectively. The data points are differentiated according to the cell types: endocardial (Endo), mid-myocardial (Mid), and epicardial (Epi) cells. The colored areas represent low and high-risk regions. The high-risk threshold for ΔQT is set at 10 ms as it is commonly accepted that changes in QT interval of less than this value are generally considered to be within the normal physiological range[43], and for ΔAPD at 13.4 ms. This threshold is based on the established relationship where QT change is approximately 1.34 times the APD change, as reported by Mirams et al.[44]. Linear regression lines are fitted for each cell type, along with an overall average. The relationship between ΔAPD and ΔQT appears linear at lower values, but the data points spread as the values increase, indicating that the linear regression model's fit decreases with higher risk. This trend is quantified by the R^2 scores[45], which show the extent to which the model fits the data linearly. The R^2 scores for the overall average and individual cell types highlight that while there is a general linear trend, significant deviations occur at higher risk levels. To further investigate this, we performed a cumulative sum test to assess the loss of linearity in the 0D model predictions compared to the 3D predictions. The test revealed that for ΔQT values of 100 ms in females and 125 ms in males, the residuals increase significantly and begin to diverge. This divergence suggests a significant difference in behavior between the 0D and 3D models beyond these thresholds, marking the onset of what we refer to as the arrhythmogenic window. The analysis suggests that while 0D models can reliably predict outcomes below this arrhythmogenic window, their ability to detect electrical propagation abnormalities diminishes considerably within the window as the risk level rises, when compared to the 3D model.

Figure 4c and d illustrate the relationship between qNet and ΔQT for females and males, respectively. qNet is intended to classify the risk of drug-

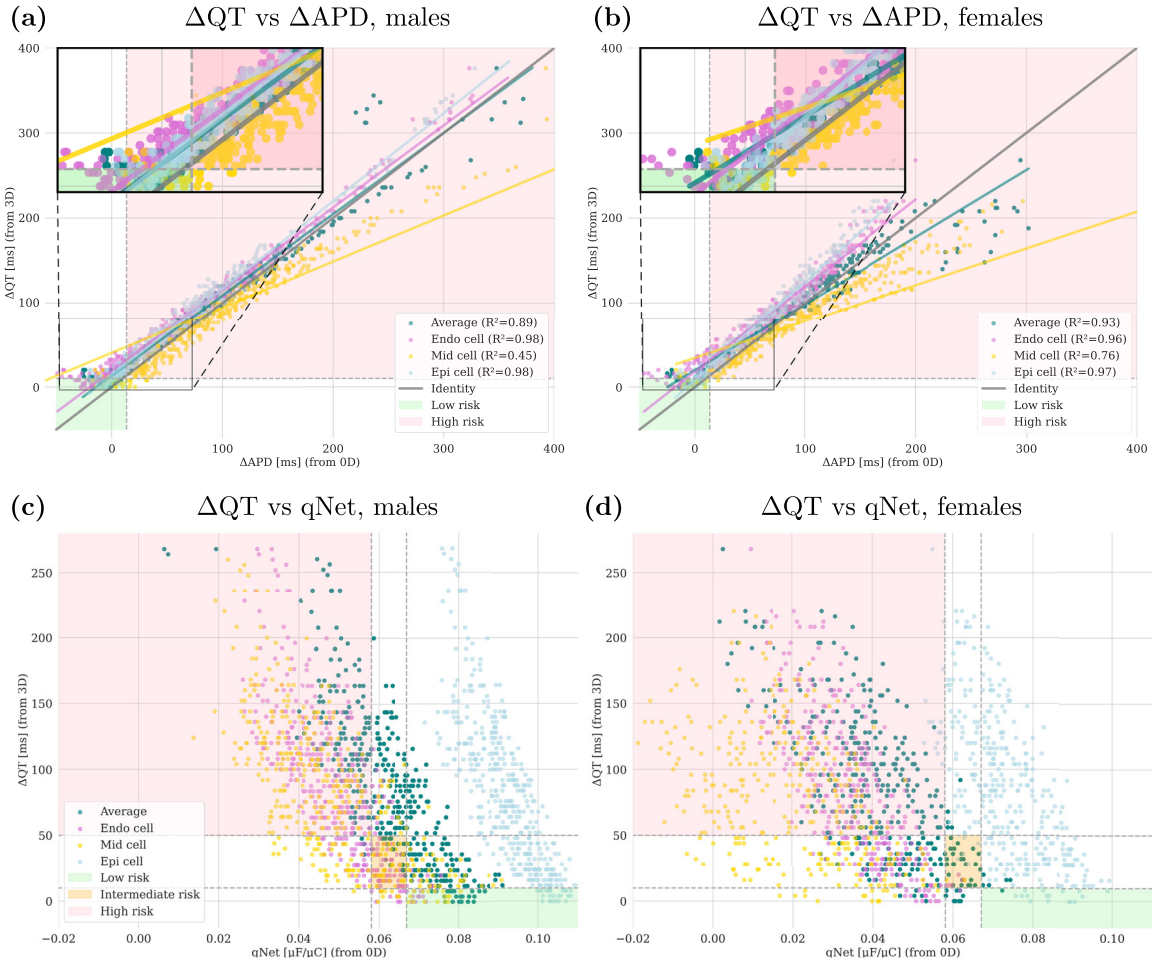

Fig. 4 | Comparison of biomarkers from 3D and 0D simulations to assess proarrhythmic risk. The top panels show a comparison of ΔQT and ΔAPD for males (**a**) and females (**b**). The bottom panels illustrate the relationship between ΔQT and qNet for males (**c**) and females (**d**). Data points are colored according to the cellular type: endocardial (Endo cell), mid-myocardial (Mid cell) and epicardial (Epi cell) types. Average represents the average of the three cellular types. Linear regression lines for each cell type indicate the general trend, and corresponding R^2 scores are reported in the legends. The shaded areas mark the low, intermediate, and high-risk regions.

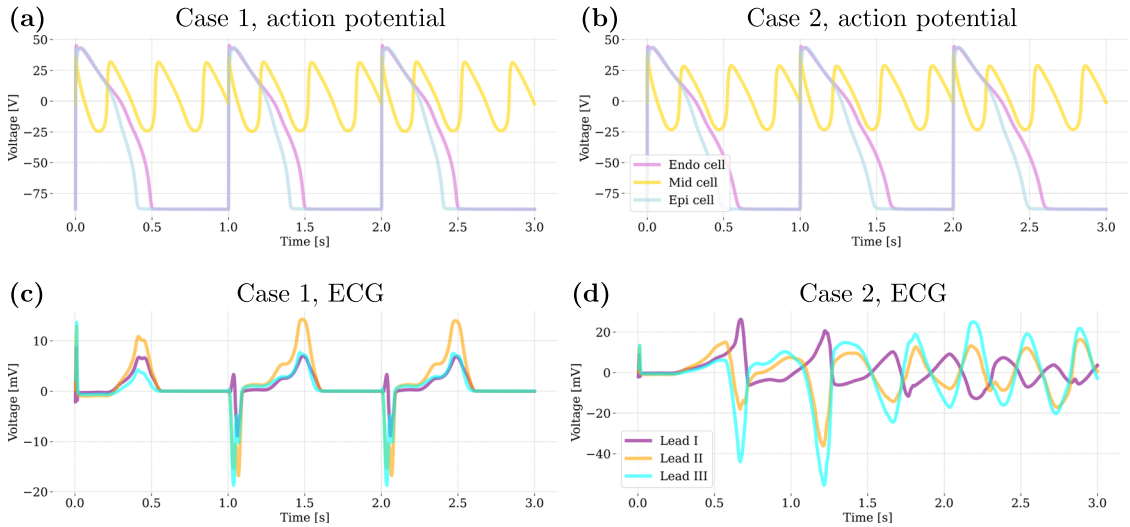

Fig. 5 | Comparison of two cases with similar action potentials leading to different ECG outcomes. These cases correspond to the following choices of the input parameters (see Eq. (1)): $\boldsymbol{x}$ = [1.00, 0.44, 1.00, 1.00, 0.45, 0.55, 1.00] and $\boldsymbol{x}$ = [0.94, 0.42, 0.90, 0.44, 0.63, 0.42, 0.91], for Case 1 and Case 2, respectively. Action potentials in (**a**, **b**) with abnormal behavior of the mid cell, are obtained from the last 3 beats of the 0D model (serving as initialization of the 3D simulation). However, the ECGs show different responses: (**c**) presents QT prolongation (170 ms) and ST elevation, while (**d**) manifests ventricular fibrillation. This illustrates that very similar cellular-level events can produce different effects on overall cardiac electrophysiological function.

induced arrhythmias into low, intermediate, and high-risk categories by evaluating the net charge carried by ions during the cardiac action potential. The plots delimit these three risk categories (intermediate risk: qNet ∈ [0.0581, 0.0671] μF/μC and ΔQT ∈ [10, 50] ms), by coloring the different corresponding regions. These thresholds are based on the FDA's CiPA initiative guidelines[46] and statistics on drug-induced risk of life-threatening arrhythmias[47]. Despite its utility for risk classification, qNet shows limited accuracy and fails to predict risk correctly for any female cell type. This constraint is evident in the plots, where the classification boundaries do not align well with the actual data points for females: our results suggest that qNet often misclassifies female cell types into incorrect risk categories.

Figure 5a and b show the action potentials for two subjects with abnormal mid-cell behavior. Despite similar cellular-level events, the ECG signals in Fig. 5c and d demonstrate significantly different outcomes: case 1 exhibits QT prolongation, while case 2 experiences ventricular fibrillation. These comparisons highlight that even with similar cellular behaviors, the overall cardiac outcomes can vary greatly. QT prolongation of case 1 suggests a significant but non-lethal proarrhythmic risk, whereas ventricular fibrillation characterizing case 2 indicates a life-threatening condition. The shortcomings of 0D models become particularly apparent under conditions of high QT prolongation, where they fail to capture the complexities of whole heart interactions and tissue-level electrical impulse propagation that are crucial for accurate risk assessment.

Accuracy of the emulators compared to simulator results

First, a preliminary version of the emulators is introduced and used for database design with a conservative approach to the ion channels' blockade range. Then, a global sensitivity analysis (GSA) is performed to understand the main ion channels that influence our model response and consequently to expand and refine our database. This provides the ability to create an enhanced version of the emulators based on this last dataset, which are evaluated for their accuracy and computational performance in comparison to the simulator. With this, the emulators are capable of predicting the ΔQT responses even with higher blockades of the ion channels, promoting the occurrence of arrhythmic events.

Emulator preliminary version. We begin by designing the dataset for predicting ΔQT using a GPR model. Denoting by N the sample size, the input of the GPR model are vectors $\boldsymbol{x}_i \in \mathbb{R}^7$, containing the current blockades of the seven ionic channels:

$$\boldsymbol{x}_i = \left[1-\beta_{\mathrm{CaL}}, 1-\beta_{\mathrm{NaL}}, 1-\beta_{\mathrm{to}}, 1-\beta_{\mathrm{Ks}}, 1-\beta_{\mathrm{K1}}, 1-\beta_{\mathrm{Na}}, 1-\beta_{\mathrm{Kr}}\right]_i, \quad \text{with } \mathrm{i} = 1, \ldots, \mathrm{N}. \tag{1}$$

The output $y_i \in \mathbb{R}$, is the corresponding QT prolongation :

$$y_i = \mathrm{QT}_i, \text{ with } i = 1, \ldots, N. \tag{2}$$

In Eq. (1), we consider blockades β_{k} of the seven channels $\beta_{\mathrm{k}} \in [0.0, 0.6]$, with k = CaL, NaL, to, Ks, K1, Na, Kr, as done in ref. 36. The input vectors are selected by performing Latin hypercube sampling[48] as N varies in {150, 250, 350}. To compare the results obtained for the different values of N and determine the ideal sample size, we fix the test set as $\mathscr{D}^{\mathrm{test}} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{42}$ for all the three cases, where 42 represents the 28% of the data when N = 150. The training set $\mathscr{D}^{\mathrm{train}}$ consists of 72% of the remaining part of the dataset $\mathscr{D} \backslash \mathscr{D}^{\mathrm{test}}$. Then, for each N, we tune the sex-specific optimal hyperparameters through an automatic exhaustive search that maximizes the R^2 score. Table 1 presents the variation in the sample size N of the R^2 score, the mean absolute error $\varepsilon^{\mathrm{MAE}}$, the mean absolute percentage error $\varepsilon^{\mathrm{MAPE}}$, and the root mean squared error $\varepsilon^{\mathrm{RMSE}}$. The scores are high for both sexes, regardless of N. In the male case, the errors show less sensitivity to N, while they exhibit a considerable improvement as N increases for females.

The GPR model, tuned with N = 350, represents the preliminary version of the emulators used to perform the GSA[49]. The objective is to identify the ionic channels that most significantly impact the output, specifically focusing on inputs that predominantly induce a high risk of QT

Table 1 | R^2 scores and prediction errors on the test set $\mathscr{D}^{\mathrm{test}}$ for males (M) and females (F) using different values of N

N	R^2 score		$\varepsilon^{\mathrm{MAE}}$ [ms]		$\varepsilon^{\mathrm{MAPE}}$ [%]		$\varepsilon^{\mathrm{RMSE}}$ [ms]	
	M	F	M	F	M	F	M	F
150	0.999	0.995	1.409	3.052	3.364	5.851	1.680	3.961
250	0.999	0.998	1.384	1.895	3.575	3.997	1.750	2.557
350	0.999	0.998	1.349	1.877	3.411	3.829	1.705	2.412

prolongation. Based on[47] (as we also do in the analysis of Fig. 4), we fix the high-risk threshold to 50 ms and use the GSA to examine the channels that primarily contribute to achieving a ΔQT exceeding this threshold. As shown in Table 2, the three most influential ionic channels for both males and females are I_{Kr}, I_{NaL}, and I_{Ks}, with I_{Kr} providing the most substantial contribution, both individually and through interactions with other channels.

In light of this, we leverage the results of the GSA to expand our database. Specifically, we conduct an additional 100 simulations per sex, focusing on the blockades of these three critical ion channels $\beta_k \in [0.4, 0.9]$, with k = Kr, NaL, Ks, while setting the remaining blockades to zero. Throughout the rest of the paper, we present results from these enhanced emulators.

Table 2 | GSA with the preliminary version of the emulators ($\beta_k \in [0.0, 0.6]$, with k = CaL, NaL, to, Ks, K1, Na, Kr)

Index	I_{Kr}	I_{NaL}	I_{Ks}	I_{Na}	I_{CaL}	I_{K1}	I_{to}
GSA on males							
S_1	0.713	0.020	0.005	0.024	0.007	0.005	0.000
S_T	0.943	0.170	0.134	0.130	0.126	0.102	0.007
Index	**I_{Kr}**	**I_{Ks}**	**I_{NaL}**	**I_{K1}**	**I_{CaL}**	**I_{Na}**	**I_{to}**
GSA on females							
S_1	0.652	0.013	0.011	0.017	0.012	0.002	0.000
S_T	0.920	0.195	0.193	0.165	0.101	0.074	0.011

S_1 and S_T are the first and total Sobol' indices[49], respectively.

Table 3 | Optimal hyperparameters for XGBC and KNC methods

Sex	Model	Hyperparameter	Tuned value
M	XGBC	colsample_bytree	0.7
		gamma	0.5
		subsample	0.7
		learning_rate	0.1
		max_depth	3
		n_estimators	100
F	KNC	metric	manhattan
		weights	distance
		n_neighbors	7

Table 4 | Test scores of the classifiers across the considered metrics for males (M) and females (F)

Accuracy		Precision		Recall		F1 score	
M	F	M	F	M	F	M	F
1.00	0.99	1.00	0.94	1.00	1.00	1.00	0.97

Emulator enhanced version. Expanding the dataset with higher blockade levels leads to an increase in arrhythmic cases. To predict ΔQT, we design our enhanced emulator using a classifier followed by a GPR model. The classifier enables us to filter out inputs that lead to arrhythmias so that we train the regressor only on relevant data.

The first step of the emulators consists of a classifier to determine whether the ECG is arrhythmic: it is characterized by an abnormal signal—for which the ΔQT is not computable—or the QT prolongation exceeds a fixed threshold τ. Specifically, following[40], we set $\tau = 240$ ms and $\tau = 196$ ms for males and females, respectively. This choice allows us to limit the QT to the threshold value of 600 ms for both sexes.

Considering the same inputs of Eq. (1), the output of the classifier is $y_i^{\text{class}} \in \{0, 1\}, i = 1, \ldots, N$, where a value of 0 (negative) indicates a non-arrhythmic case, while 1 (positive) indicates an arrhythmic one. In our data, 92% of the outputs are negative for males, compared to 85% for females. We randomly divide our data into a training and test set, allocating 72% of data for training and 28% for testing. Note that the data is split in a way that preserves the proportion of positive and negative values in both the training and test sets. We use XGBoost Classifier (XGBC)[50] and *k*-Neighbors Classifier (KNC)[51] for males and females, respectively. We tune the optimal hyperparameters on the training set through an exhaustive search, looking for those that maximize the F1 score[52], as shown in Table 3. The performance of the classifier is reported in Table 4. The evaluation of accuracy, precision, recall, and F1 score metrics[52] on the test set reveals perfect scores for males and consistently high scores for females. The classifier's strong performance is further illustrated in Fig. 6, which presents the confusion matrix of predicted versus actual values. Here we observe perfect predictions for males and only one false positive for females.

The second step of the emulators is based on a GPR model that predicts the ΔQT values. Inputs and outputs of this model are the same as those defined in Eq. (1) and Eq. (2), respectively. We split our data as done for the classifier and we exclude inputs that are identified as positive (arrhythmic). Similarly to the preliminary version of the emulators, we first tune the optimal hyperparameters through an exhaustive search on the training set. These hyperparameters are selected to maximize the R^2 score of predictions and are reported in Table 5.

Table 5 | Optimal hyperparameters for the GPR model, for males (M) and females (F)

Model	Hyperparameters	Tuned value	
		M	F
GPR	kernel	Matérn 1.5	Matérn 2.5
	gamma	10^{-3}	10^{-3}
	subsample	20	5

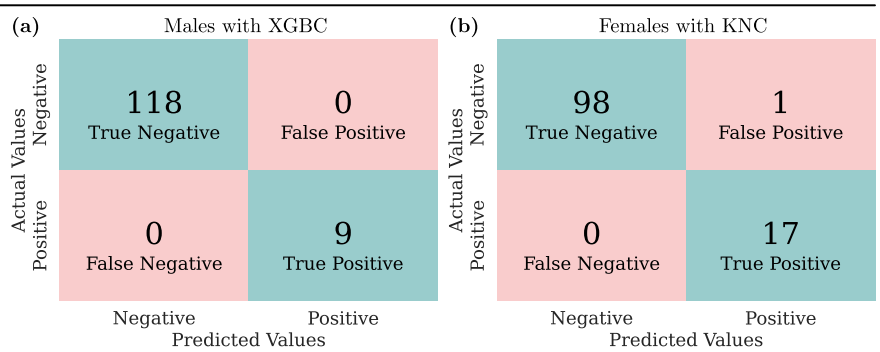

Fig. 6 | Confusion matrices of classifiers. Plots (**a**, **b**) show the results for males and females, respectively. True positives (negatives) represent the number of correctly predicted positive (negative) cases. False positives (negatives) indicate the number of incorrect positive (negative) predictions.

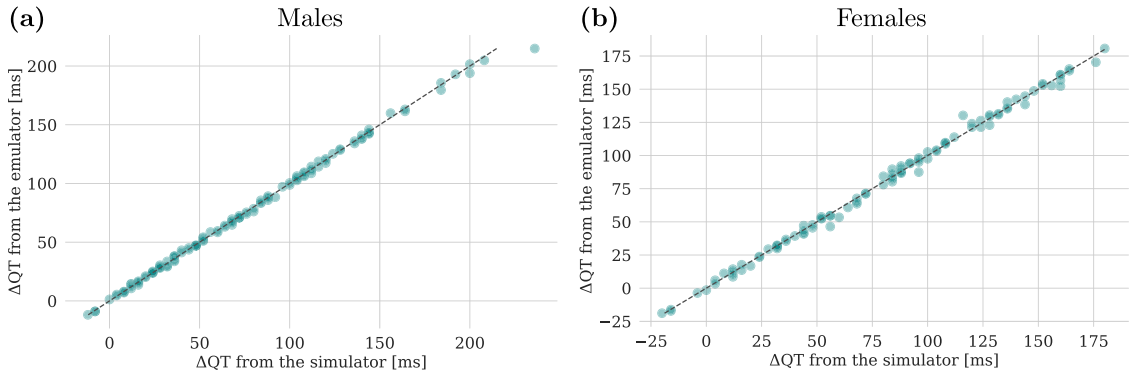

Fig. 7 | Comparison of ΔQT predictions between the simulator and emulators. Both for males (**a**) and females (**b**), the *x*-axis represents the ΔQT values computed with the simulator and the *y*-axis the corresponding predictions from the emulator. The close alignment along the diagonal indicates strong agreement between the simulator and emulator predictions.

Table 6 | R^2 scores and prediction errors on the test set for males (M) and females (F)

R^2 score		ε^{MAE} [ms]		ε^{MAPE} [%]		ε^{RMSE} [ms]	
M	F	M	F	M	F	M	F
0.998	0.996	1.491	2.107	3.063	4.433	2.582	3.083

The tuned GPR model is used to predict the ΔQT values corresponding to the ionic channel blockades of the test set. Figure 7 shows that the ΔQT predictions are highly accurate both for males and females, a result that is also confirmed in Table 6, where we report the R^2 score and the prediction errors.

Outcomes of global sensitivity analysis

After fine-tuning our emulators, we perform again the GSA to address two key questions:

- Which ionic channels mainly influence the output?
- Which ionic channels primarily contribute to a QT prolongation that is classified as high-risk?

The first question is directly determined by the emulator output, while the second is assessed based on whether the predicted ΔQT exceeds the high-risk threshold of 50 ms[47].

Figure 8 shows the GSA results, separated by sex, with the first analysis displayed on the top and the second one on the bottom. In each plot, the solid bars represent the S_1 indices, indicating the main individual effects, while the striped bars indicate the S_T indices, reflecting the total effects, i.e., describing the interactions between different ion channel blockades. Confidence intervals are also included to show the uncertainty in the estimation of the Sobol' indices. From the top plots, we observe that I_{Kr} is the most relevant channel in predicting QT prolongation, regardless of sex. It is important to note that I_{Kr} refers to the rapidly activating delayed rectifier potassium current, which is largely mediated by the hERG (human Ether-à-go-go Related Gene) channel, often referred to simply as I_{Kr} channel. The other channels have considerably less influence on the output, as indicated by their much smaller Sobol' indices and the absence of overlap in their confidence intervals with I_{Kr}. Additionally, interactions between them are negligible since S_1 and S_T are very similar for each channel considered. On the other hand, the second analysis highlights that I_{Kr}, I_{NaL}, and I_{Ks} are the three most influential channels in contributing to a high-risk QT prolongation. Notably, interactions between different channel blockades are significant, particularly for the most influential input, I_{Kr}.

Evaluating emulator and simulator outcomes for benchmark drugs

In Fig. 9, we present the concentration-QT prolongation (C-ΔQT) predictions of the emulators, simulator and clinical data under the influence of dofetilide. The slope of the C-ΔQT relationship is the state of the art, therefore crucial and extensively used in regulatory practices, as it measures the rate of change in the QT interval per unit increase in drug concentration.

Our emulators are developed using single reference geometries that represent healthy adult anatomies, as they are derived from data from middle-aged adult deceased donors with no history of cardiac disease and anatomically normal ventricles. Consequently, the simulator results shown in Fig. 9a are based on the same anatomies used for training the emulators. Leveraging the performance of the emulators, we extend the concentration range beyond the typical experimental limits. This illustrates their potential to predict drug effects at higher concentrations in immediate response time, as we will better discuss in Section Computational performance. Results demonstrate a strong similarity between the outputs of the emulators and the simulator, confirming the accuracy of the emulators. We calculate the slopes of the C-ΔQT relationships for clinical, simulator, and emulator data, and then compute the relative errors between these slopes as percentages. For females, the relative error between the emulator and simulator slopes is 9.0%, the error between the simulator and clinical slopes is 31.1%, and the error between the emulator and clinical slopes is 19.2%. For males, the relative error between the emulator and simulator slopes is 7.7%, the error between the simulator and clinical slopes is 2.4%, and the error between the emulator and clinical slopes is 5.2%. These results underscore the consistent alignment among all datasets, especially considering the limited female representation in the clinical trial[53].

Then, to assess the robustness of our emulators, we compare their predictions against simulator results from six different donor cases (three per each sex) incorporating different anatomical characteristics, BMI and age, see Fig. 9b. Despite being trained on a single average anatomy, the emulator predictions remain highly accurate across different anatomies. Specifically, for female anatomies, the emulator predictions show relative errors ranging from 1.8% to 8.0% compared to the simulator. For male anatomies, relative errors range between 7.2% and 9.6%. When comparing the emulator slopes to clinical data, we observe relative errors of approximately 20.6% for females and 7.0% for males. The larger error in females is closely tied to the fact that a really small number of women were included in the clinical trial. These results suggest that, even when trained on a single average geometry, the emulators provide predictions that are comparable to the simulator alignment with the usually sparse clinical data. This suggests that the emulators are effective and capable of generalizing well across various anatomical conditions, offering valuable insights despite the inherent variability in clinical data.

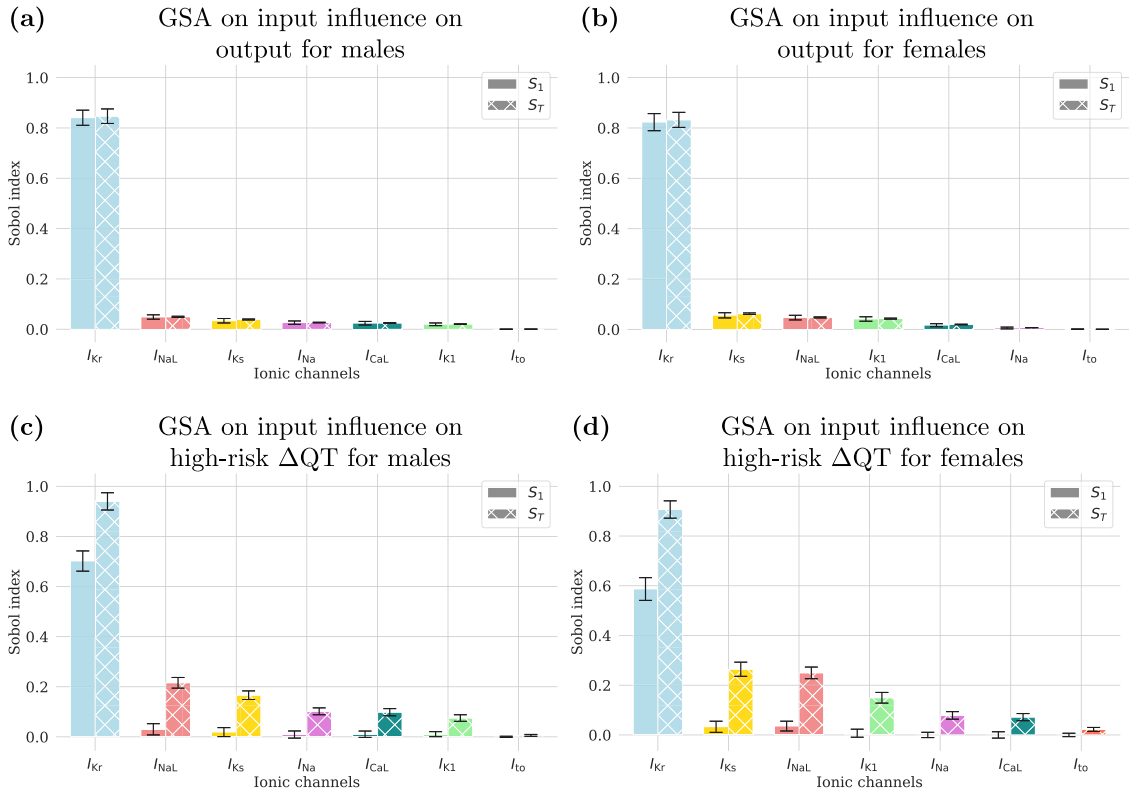

Fig. 8 | First (S_1) and total (S_T) Sobol' indices from the GSA on the influence of ionic channel blockades on ΔQT predictions. The top plots show the influence on the predicted ΔQT for males (**a**) and females (**b**). The bottom plots illustrate the influence in determining if the QT prolongation is of high-risk level for males (**c**) and females (**d**). Each bar represents the impact of a specific ionic channel blockade, highlighting channels that most affect QT prolongation and high-risk outcomes.

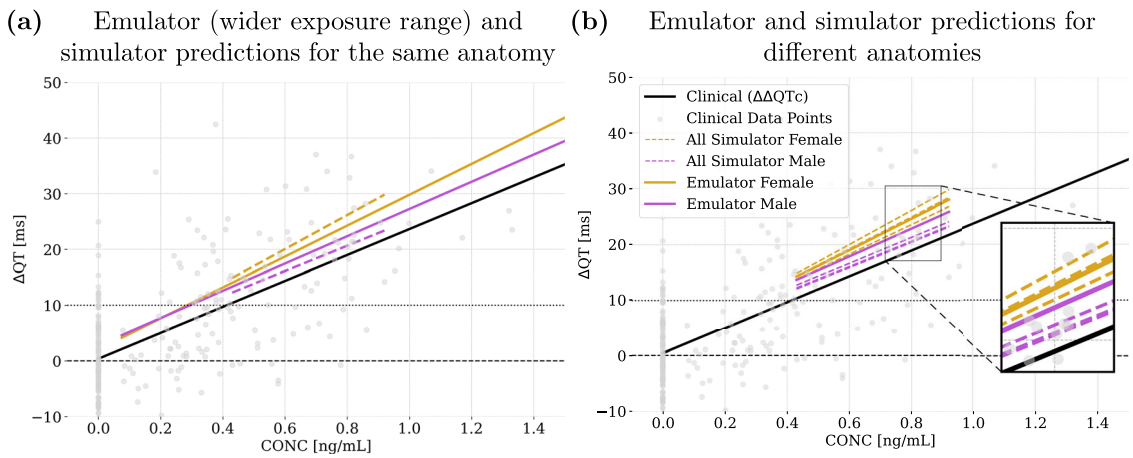

Fig. 9 | Application of the emulators to predict C-ΔQT response for dofetilide. Results in (**a**) are obtained from the same anatomy used to train the emulators, while different anatomies are considered in (**b**). ΔΔQTc clinical, ΔQT simulated and ΔQT emulated as a function of plasma concentration. The grey dots denote the observed ΔΔQTc with respect to plasma concentration for each measurement taken in the clinical trial extracted from Darpo et al. (2015)[53]. The magnified region in plot (**b**) highlights six dashed lines, three per sex, corresponding to the three distinct anatomies considered.

To further evaluate the accuracy of the emulators, we predict ΔQT for four benchmark drugs: moxifloxacin, ondansetron, dofetilide, and verapamil. We compare the emulator results with clinical data and with simulator results from our previous study[23], which used six different anatomies. To ensure consistency and avoid bias, we use the same concentration protocol for each drug, involving three ascending concentrations, except for verapamil, which is tested at four concentrations. The calculated C-ΔQT relationships are shown in Fig. 10. Our analysis shows that the slopes of the emulator regression lines are remarkably similar to those produced by the simulator across all benchmark drugs. This similarity is observed alongside the expected differences among sexes. Table 7 provides a comprehensive summary of the regression analysis results for each drug, comparing clinical data with the simulator and emulator results. The latter closely mirror the simulator results for all drugs. Specifically, the slopes of the emulator predictions fall within the confidence intervals of the simulated slopes, achieving critical concentrations within a 0.25-fold range of simulated results. This alignment extends to most clinical trends as well, supporting the emulator reliability and accuracy in predicting QT interval prolongation

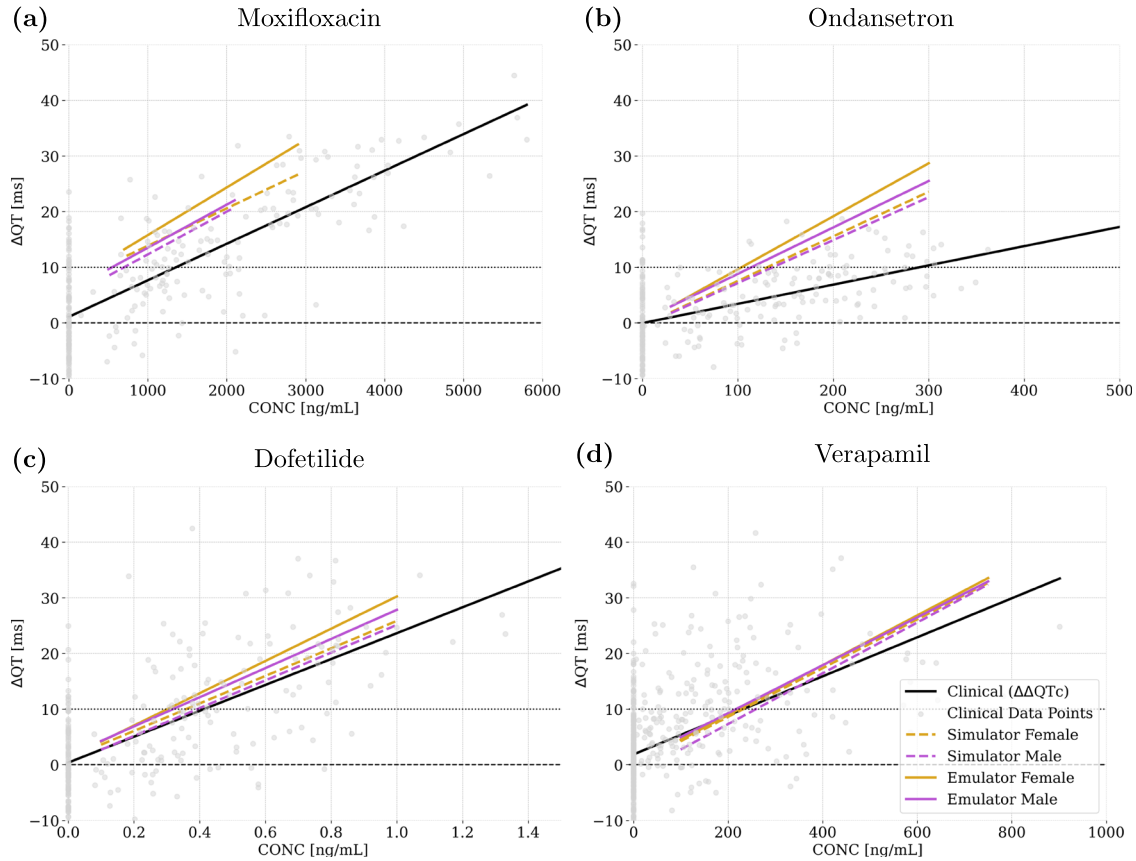

Fig. 10 | Comparison of ΔΔQTc clinical, ΔQT simulated and ΔQT emulated as a function of plasma concentration for different drugs. Plot (**a**) shows the results for moxifloxacin, plot (**b**) for ondansetron, plot (**c**) for dofetilide, and plot (**d**) for verapamil. The grey dots denote the observed ΔΔQTc with respect to plasma concentration for each measurement taken in the clinical trial extracted from Darpo et al. (2015)[53] and Vicente et al. (2019)[92].

Table 7 | Summary of regression analysis results for four benchmark drugs

Drug	Data Type	Slope	Intercept	Std Err	Confidence Interval	Critical Conc. [ng/mL]
Dofetilide	Clinical	23.252	0.387	1.653	± 3.239	0.413
	Simulator Female	24.720	1.131	0.886	± 1.737	0.359
	Simulator Male	24.925	0.201	0.979	± 1.918	0.393
	Emulator Female	28.983	1.254	0.768	± 1.505	0.302
	Emulator Male	26.185	1.650	1.059	± 2.076	0.319
Moxifloxacin	Clinical	0.007	1.094	0.000	± 0.001	1356.450
	Simulator Female	0.007	7.125	0.000	± 0.001	427.089
	Simulator Male	0.008	4.625	0.001	± 0.001	699.382
	Emulator Female	0.009	7.216	0.001	± 0.001	325.114
	Emulator Male	0.008	5.766	0.001	± 0.001	546.715
Ondansetron	Clinical	0.035	−0.033	0.004	± 0.007	290.285
	Simulator Female	0.080	−0.489	0.003	± 0.005	130.858
	Simulator Male	0.077	−0.560	0.003	± 0.006	136.854
	Emulator Female	0.095	0.090	0.000	± 0.001	103.977
	Emulator Male	0.083	0.455	0.002	± 0.003	114.323
Verapamil	Clinical	0.035	1.863	0.003	± 0.006	232.144
	Simulator Female	0.044	−0.150	0.002	± 0.005	231.871
	Simulator Male	0.046	−1.810	0.002	± 0.003	259.001
	Emulator Female	0.045	0.009	0.001	± 0.003	223.479
	Emulator Male	0.043	0.684	0.001	± 0.001	216.686

For each drug, we present the following metrics: Slope, which quantifies the rate of change in QT prolongation with respect to drug concentration; Intercept, the baseline QT prolongation when drug concentration is zero; Std Err, the standard error of the slope, indicating the precision of the regression estimate; Confidence Interval, providing the range within which the true slope is expected to fall with 95% confidence; and Critical Conc., the drug concentration at which the QT prolongation reaches 10 ms. Results are provided for both clinical data and predictions from the simulator and emulators, differentiated by sex, when applicable.

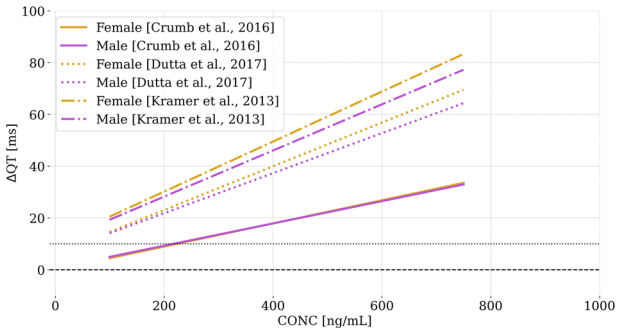

Fig. 11 | Application of the emulators to predict C-ΔQT response for verapamil with input data from three different sources[38,54,55]. The emulated ΔQT values are shown as a function of plasma concentration.

Table 8 | Computational times to build the emulators: CPU time to run 900 simulations, time to find the optimal hyperparameters of the emulators, and time to train the emulators

Simulator (900 cases)	Emulator optimal hyperparameters	Emulator training
123 d	48 min	23 s

Male and female results are summed. The simulator time has been computed excluding the cases where the initialization of the cellular model did not converge to a periodic solution.

Table 9 | Comparison of the simulator and emulator computational performance to predict ΔQT for a single case and to reproduce in silico clinical trials of benchmark drugs (average cost for the clinical trial of a single drug)

		Simulator	Emulators	Speed-up
Single case		3.43 h	$8 \cdot 10^{-2}$ s	$1.5 \cdot 10^{5}$
Benchmark drugs	Moxifloxacin Dofetilide Ondansetron Verapamil	9 d	$3.2 \cdot 10^{-1}$ s	$2.4 \cdot 10^{6}$

The speed-up is computed as the ratio between reported simulator and emulator times.

under the effect of different drugs. Ondansetron and moxifloxacin present some discrepancies with respect to the clinical data in slope and intercept, respectively, leading to differences in critical concentrations. Drug experimental variability and clinical data intrinsic biases may justify the observed differences, as also discussed in ref. 23.

Finally, in Fig. 11 we include an analysis that compares ion channel *IC*50 and Hill coefficient values from three different publicly available sources for verapamil[38,54,55]. This comparison demonstrates the ability of our tool to assess the impact of drug-related parameter variability with minimal computational cost. While we recognize that these values can vary significantly across different laboratories, our emulators enable a thorough exploration of such variations efficiently. This capability is particularly valuable, as it allows researchers to evaluate the potential effects of inter-laboratory differences without the need for extensive computational resources.

Computational performance

We present the results in terms of the computational performance of both the simulator and the emulators. Table 8 shows the computational times required to build the emulators: running 900 simulations took approximately 123 days of CPU time. The time to find the optimal hyperparameters was about 48 minutes, while the emulator training took only seconds. It is important to note that these operations are performed only once.

In Table 9, we present the computational costs associated with reproducing a single case and conducting clinical trials with benchmark drugs using both the simulator and the emulators. A single simulation took, on average, 3.43 h, whereas the emulators can predict the ΔQT in mere centiseconds, providing a speed-up of five orders of magnitude. Generating a clinical trial for a single drug required approximately 9 days of CPU time using the simulator. In contrast, the emulators significantly reduce this cost, delivering accurate results in less than a second.

Discussion

We introduced sex-specific emulators capable of predicting drug-induced proarrhythmic risk. Developed separately for each sex, these emulators leveraged a comprehensive dataset derived from high-fidelity 3D electrophysiological simulations, capturing crucial anatomical and phenotypical differences. By taking ionic channel blockades as input, they provided real-time predictions of QT prolongation, ensuring accurate and sex-specific assessments.

To construct our dataset, we leveraged the computational power of Alya[39,56] to perform 900 electrophysiological simulations (450 per sex), sampling the space of seven ionic channel blockades as input, which totaled approximately 2.1 million CPU hours. This comprehensive dataset, entirely composed of high-fidelity 3D simulations, was essential for developing accurate emulators. Notably, our dataset is comparable in size to that built by Costabal et al.[36], who used a mix of 400 low-fidelity (1D) and 45 high-fidelity (3D) simulations.

The usage of 3D cardiac models introduces additional complexity and computational cost compared to 0D single-cell models. However, our analysis demonstrated the clear advantages of 3D modeling for predicting drug-induced proarrhythmic risk, underscoring its importance despite the higher computational demands. 3D models provide greater physiological relevance by capturing spatial heterogeneity and anatomical details crucial for simulating drug effects on cardiac electrophysiology. These models can compute ΔQT, which makes them comparable to clinical trial outcomes, thus enhancing the reliability of computational models for cardiac safety assessment. In contrast, 0D modeling relies on in vitro metrics such as APD or qNet, which are not directly translatable to clinical settings. Another significant advantage of 3D models is their capacity to incorporate disease states and comorbidities, particularly through detailed anatomical representations. This is especially valuable because such patients are often excluded from real clinical trials. By using cardiac models, we can ensure that these underrepresented groups are considered in the assessment of drug safety. Our comprehensive dataset allowed us to systematically compare 3D versus 0D modeling in drug proarrhythmic risk assessment. By comparing ΔQT and ΔAPD, we found that while 0D models can generally predict outcomes accurately, their ability to identify electrical propagation abnormalities decreased significantly as the risk level increased. This highlights that the 0D model may not provide the same clinical insights as the 3D model. While it is challenging to determine which model is more accurate without common ground truth data, the clinical gold standard for assessing proarrhythmic risk is the quantification of ΔQT. Therefore, a model capable of computing this measure is preferable for establishing credibility and validation. Since the 0D and 3D models yield different outcomes beyond a certain risk threshold, we believe the 3D model may be more suitable for assessing drug-induced proarrhythmic risk, particularly in higher-risk scenarios. Furthermore, our comparison between ΔQT and qNet revealed that the latter frequently misclassifies female cell types into incorrect risk categories. This limitation stems from qNet being derived from the original ORd model, which is based on a generalized, endocardial model[41]. As a result, qNet is more accurate for male endocardial cells but fails to account for the sex-specific differences in cardiac electrophysiology that are crucial for accurately assessing risk in females. Additionally, comparing action potentials (from the 0D model) with the ECGs (from the 3D simulation) revealed that similar patterns at the 0D level can lead to different outcomes at the 3D level, especially within the arrhythmogenic window, again

emphasizing the importance of 3D models in proarrhythmic risk assessment.

Our study underscored the critical importance of incorporating sex differences in computational models, aligning with previous research findings[21–24]. Our results showed that females are more susceptible to arrhythmias, consistent with earlier studies[25,57]. The use of 3D electrophysiological models enabled a comprehensive approach to account for sex-specific anatomies and phenotypes. Thus, although 3D cardiac models demand greater computational resources compared to 0D models, their capacity to capture complex physiological details and sex differences—thereby providing more accurate and personalized risk assessments—justifies their use in predicting drug-induced proarrhythmic risk. Thus, our findings emphasized the necessity of high-fidelity 3D simulations to enhance the reliability and accuracy of computational cardiac safety evaluations. One way to mitigate the computational expense of 3D simulations is by developing emulators, or surrogate models, which can predict the outcomes of computationally costly 3D simulations at a fraction of the computational cost.

To compute QT prolongation induced by drugs in real-time, we developed two emulators that utilized a two-step approach: a classifier followed by a GPR model. The classifier is used to filter cases, improving the robustness and accuracy of our emulators by addressing the challenges inherent in cardiac electrophysiological models, which often display bifurcations and discontinuous responses. Ghosh et al.[58] noted that such discontinuities could complicate predictions when using GPR models, as they assume smooth and continuous responses to changes in parameters. By implementing the classifier, we effectively segregated non-arrhythmic from arrhythmic cases, thereby minimizing the complexity and discontinuities presented to the GPR model. Additionally, sex-specific ΔQT thresholds were crucial to avoid the introduction of physiological anomalies that could mislead the training process and predictions. This consideration is particularly relevant for Thorough QT studies, where identifying significant QT prolongation is essential for further clinical development. Literature highlights that a QT prolongation exceeding 20 ms is often significant enough to warrant extra monitoring and labeling[59]. Furthermore, patients experiencing TdP usually show more pronounced QT prolongation before its onset. Sex differences add another layer of complexity, with women being more sensitive to QT prolongation[25]. Expert guidelines suggest that QT intervals longer than 550 ms are associated with a high arrhythmic risk[40]. Consequently, our chosen thresholds—240 ms for males and 196 ms for females—are set conservatively to account for cases that could lead to a QT interval of 600 ms, ensuring that rare but critical extreme cases are also identified. This approach allowed the regressor to work with a more homogeneous and relevant dataset, thereby enhancing the accuracy and reliability of our predictions.

The sex-based emulators demonstrated high accuracy when compared to simulator results, with errors in the range of 2.6 ms, a relatively minor discrepancy given the magnitude of drug-induced changes. These emulators enabled us to replicate clinical trials for four benchmark drugs and validate our findings against in vivo data. This consistency underscored the emulator reliability in forecasting drug-induced QT prolongation across varying concentrations and compounds. The emulators effectively captured the sex-based differences in drug response observed with the simulator, which is crucial for developing sex-specific therapeutic strategies, especially when it comes to high QT-prolonging drugs. Notably, emulators facilitated an in-depth exploration of dofetilide responses at higher concentrations without additional computational expense. Standard in silico clinical trials typically require around 9 days of CPU time, but the emulators reduced this time to mere fractions of a second, achieving a speed-up of six orders of magnitude in predicting QT prolongation. To assess the robustness of our model across different anatomical contexts, we compared emulator predictions—trained on fixed anatomy—with simulator results using different anatomical models. In addition, different from the emulators, the simulator results are obtained by combining phenotypical variability[60]. The emulators produced results with errors ranging from 1.8% to 9.6%, a level of error we considered acceptable in comparison to the existent interstudy variability in clinical practice[61]. This demonstrated that our emulators maintain robustness and reliability, even when tested against data with a larger variability.

In addition, these emulators are also tools to identify and prioritize influential inputs by carrying out GSA. This analysis was used to achieve two main objectives: first, to study the influence of each input on the ΔQT response; and second, to identify the key inputs that determine whether there is a high risk of QT prolongation. We found that I_{Kr} is the primary predictor of ΔQT, consistent with the findings reported in ref. 36. The results also demonstrate that the seven inputs can be linearly combined to reconstruct the QT prolongation. However, when evaluating clinically relevant quantities, such as assessing whether the inputs lead to a QT prolongation that is classified as high-risk, requires accounting for high-order interactions between the channels, with I_{Kr}, I_{NaL}, and I_{Ks} being the most significant contributors. These findings underscore the importance of considering multichannel interactions in proarrhythmic risk evaluation, suggesting that a sole reliance on hERG in vitro assessment may be insufficient.

It is important to highlight that the high accuracy of our emulators is largely attributable to the quality and reliability of the underlying 3D cardiac simulator[39]. This simulator is the result of nearly two decades of continuous development, research, and an extensive validation process across multiple studies[62–68]. Its development has involved close collaboration with clinicians and pharmaceutical experts, ensuring that it reflects the complexity of human cardiac physiology and drug interactions[23,69]. Achieving such precise emulator performance would not have been possible without this foundation of well-established computational models, which have been refined through years of iterative improvements and cross-disciplinary expertise[39,56,70–72]. As such, while the emulators significantly reduce computational costs, their effectiveness is intrinsically linked to the robustness and accuracy of the simulator from which they are derived.

This study has some limitations. Firstly, we relied solely on ΔQT to estimate proarrhythmic risk, whereas other electrocardiogram features, such as the Tpeak-Tend interval[73], might provide additional insights into drug effects. Secondly, our dataset does not account for variability in phenotypic expression or anatomical differences, which could be crucial for a more comprehensive risk assessment. Future research should focus on incorporating these variabilities to better represent the spectrum of patient responses and further improve the accuracy of our results when compared to simulator results. Furthermore, this study assessed the QT effects of drugs without considering the potential impact of their metabolites, which may be the primary contributors to these effects. However, in the drug discovery phase, detailed information of metabolites is rarely available and a reasonable starting point is therefore to test the effect of the parent drug. An interesting direction for future work would be to perform sensitivity analysis to account for potential changes in ionic currents triggered by active metabolites. Additionally, our current models use a uniform activation protocol across all cases.

A key aspect highlighted by our study is the observed sensitivity to calcium alternans across female cells, which may be partially attributed to the interplay between sex-specific ionic channel conductances proposed by Fogli Iseppe et al.[74] and ionic channel blockages. This combination of factors could increase the likelihood of calcium alternans, impacting cardiac action potential stability and raising susceptibility to arrhythmic events. We believe that hormonal variations like progesterone and estrogen might modulate these differences, given their significant influence on cardiac ionic channel expression and function. Future work should aim to quantify these effects, allowing for a deeper understanding of female ionic channel modulation contribute to the observed arrhythmic sensitivity in females.

Moreover, clinical trials include placebo groups to eliminate confounding effects that allow them to compute placebo-corrected ΔQT. Further developments consist of including placebo corrections within the computational models. Future directions also involve performing uncertainty quantification, especially on parameters such as $IC50$, h and the free fraction of the drug. Incorporating variability in these parameters could help refine uncertainty ranges in model predictions and enhance the overall

robustness of the emulators. An interesting application of the emulators would be to integrate ionic channel margin distributions into QT prolongation risk assessments, as suggested by Leishman et al.[75]. This approach would enable the definition of safety margins by accounting for the variability observed in in vitro assay-measured parameters. By doing so, it could serve as an effective filter in early drug development, providing a robust sensitivity analysis that informs decision-making processes and better characterizes the risk landscape.

To conclude, our emulators provide a valuable tool for evaluating QT prolongation from the discovery phase and the very early stages of drug development, offering high accuracy in simulating drug-induced effects on cardiac physiology. Serving as an fast-response preliminary design tool for computational clinical trials, these emulators create a virtual trial environment where outcomes can be predicted and concentration ranges refined—all before committing to computationally expensive full-scale simulations. However, it is crucial to underscore that the emulators do not replace the simulator. The simulator remains indispensable, especially in later stages of development, due to its ability to account for a wider array of physiological factors that the emulators, particularly given their current limitations, cannot fully capture. This distinction is vital, as the emulators are primarily intended to complement the simulator by accelerating early-stage assessments, not to substitute the comprehensive analyses that only the simulator can provide. Moreover, the use of surrogate models allows for the comprehensive execution of studies with uncertainty quantification across all variable drug-related parameters, enabling thorough compound evaluation, informed early-stage decision-making, and the efficient design of 3D virtual trials with a larger control over experimental conditions. Furthermore, emulators also address the growing concern of climate impact linked to preclinical and clinical studies, an item that is getting higher on the agenda of large pharmaceutical companies. The emulators unleash the potential to analyze, investigate, and innovate at a marginal/insignificant cost for the environment, representing a negligible carbon footprint. By integrating high-fidelity simulations with immediate prediction capabilities, our emulators advance the concept of digital twins in biomedicine, improving the efficiency and sustainability of drug development.

Methods

The electrophysiological simulator

Anatomical models. We collect retrospective data from the Visible Heart Lab library at the University of Minnesota[76]. Specifically, we consider two biventricular cardiac geometries (male and female) reconstructed from high-resolution MRI scans. We use human heart anatomies from adult deceased donors with no history of cardiac disease and anatomically normal ventricles in order to represent average healthy individuals.

The electrophysiological model. To model cardiac electrophysiology, we consider the monodomain model[77] coupled with the ORd cellular model[41]. We selected the ORd model because it is recognized as the consensus in silico model by the CiPA initiative[54]. Specifically, we used a modified version of the ORd model by Passini et al.[78] with modified conductances as described in Dutta et al.[54]. The location of the activation points was instead set following the work by Durrer et al.[79], as we explained in our previous study[64]. Cardiac fibers are modeled using the outflow tract rule-based method[80]. This method is particularly suited for our study due to its ability to accurately assign fiber directions within high-resolution biventricular geometries, which include complex structures such as trabeculae and papillary muscles[64,80]. We incorporate transmural myocyte heterogeneity by assigning distinct electrophysiological and cellular properties to various regions of the myocardium: endocardial (inner, 30%), mid-myocardial (middle, 40%), and epicardial (outer, 30%) cells. A larger diffusion is assigned to a one-element layer on the endocardial surface to account for the fast conduction of the Purkinje fibers. We set a constant heart rate of 60 bpm.

To generate male and female phenotypes, we apply sex-specific ion channel subunit expression as described in refs. 23,74,81 to the two sex-specific anatomies considered in the study.

Modeling drugs. To model the effect of drugs, we use a multi-channel conductance-block formulation[9,23]. Given the conductance g_k of one of the seven most influent ionic channels I_{CaL}, I_{NaL}, I_{to}, I_{Ks}, I_{K1}, I_{Na}, and I_{Kr}[38], we define the ion channel conductance after the drug administration with the following Hill model[9]:

$$g_k^{drug} = g_k \beta_k, \text{ with } \beta_k = \left[1 + \left(\frac{C}{IC50_k}\right)^{h_k}\right]^{-1}, \quad (3)$$

with k = CaL, NaL, to, Ks, K1, Na, Kr. In the equations above, g_k^{drug} is the conductance of the k–th channel after drug administration, C the drug concentration, $IC50_k$ the concentration required to have a 50% current blockade of the k–th channel, and h_k the corresponding Hill exponent. Thus, the blockade β_k of channel k is identified by the Hill parameters h_k, the $IC50_k$ and the drug concentration C. Notice that C is defined as the concentration of the drug in the plasma that is not bound to plasma proteins and is therefore available to exert a pharmacological effect. Therefore, it is expressed as

$$C = \frac{f_u}{100}\widetilde{C},$$

where $\widetilde{C}$ is the total concentration of the drug in the plasma and f_u—expressed in percentage—is the free fraction of the drug: the ratio of the unbound drug concentration to the total drug concentration.

Initialization of the electrophysiological model. The 3D simulation is initialized by solving the 0D ORd model for each cell type until the intracellular calcium concentration converges to a periodic solution. Periodicity is determined when the RMSE between consecutive beats is smaller than 10^{-7} μmol for three consecutive beats. The results from the 0D model then serve as initial conditions for the 3D simulation. Cases where periodicity is never achieved due to the presence of calcium alternans are excluded from our analysis (0 cases for males, 37 for females).

For the analysis in Section Modeling proarrhythmic risk in 0D and 3D, we compared the output from the initialization of the 0D ORd model against the ΔQT obtained in the subsequent 3D simulation.

Pseudo ECG and ΔQT computation. To compute the ECG, we assume isotropic electrical conductivity in the torso and we rely on a pseudo-ECG approach, for which the potential in a given generic point $\boldsymbol{x}_*$ of the body (where the electrode is positioned) is computed as[82]:

$$u(\boldsymbol{x}_*, t) = \int D\nabla v(\boldsymbol{x}, t) \cdot \nabla\left(\frac{1}{||\boldsymbol{x} - \boldsymbol{x}_*||}\right)\mathrm{d}\boldsymbol{x},$$

where $\Omega \times (0, T)$ is the spatio-temporal domain, Ω represents the biventricular geometry and T is the simulation final time. The function $v : \times(0, T) \to \mathbb{R}$ is the transmembrane potential and D is the orthotropic tensor of local diffusivities. Three leads are then computed as[64]:

$$\begin{aligned} \text{lead}_{\text{I}}(t) &= u(\boldsymbol{x}_{\text{LA}}, t) - u(\boldsymbol{x}_{\text{RA}}, t), \\ \text{lead}_{\text{II}}(t) &= u(\boldsymbol{x}_{\text{LL}}, t) - u(\boldsymbol{x}_{\text{RA}}, t), \\ \text{lead}_{\text{III}}(t) &= u(\boldsymbol{x}_{\text{LL}}, t) - u(\boldsymbol{x}_{\text{LA}}, t), \end{aligned}$$

where $\boldsymbol{x}_{\text{LA}}$, $\boldsymbol{x}_{\text{RA}}$, and $\boldsymbol{x}_{\text{LL}}$ are the positions of the electrodes at the left arm (LA), right arm (RA), and left leg (LL), respectively. We refer the interested reader to[64] for additional information on how we compute the pseudo-ECG. The latter is used to compute the QT interval duration at a baseline

configuration (QT^{bsl}) and after drug administration (QT^{drug}). This computation is done automatically for all the cases using an automatic algorithm as we explain in ref. 23. The QT prolongation is then defined as

$$\Delta QT = QT^{drug} - QT^{bsl}. \quad (4)$$

Our computational approach isolates and quantifies the direct impact of drugs on ΔQT, eliminating the confounding effects present in clinical trial cohorts, such as varying patient demographics, concomitant medications, and individual physiological differences. Consequently, our computational data does not require placebo correction, as it inherently excludes these variables. This approach ensures that the observed ΔQT effects are solely due to the drug itself, providing a clearer comparison to clinical data where placebo adjustments account for such confounders[43]. Additionally, by using a fixed heartbeat period of 1 second (60 bpm), the ΔQT values computed from both the simulator and emulators do not require heart rate correction, as no variability in heart rate is present. Typically, QT interval measurements fluctuate with changes in heart rate, necessitating correction methods like the Fridericia formula, which are often population-based[83]. Instead, maintaining a constant heart rate allowed us to directly compare the QT effects of different drugs without relying on such corrections, which we see as a significant advantage. This means that the ΔΔQTc from clinical data is directly comparable to the ΔQT from the simulator and emulators.

Computational aspects. The computational model is implemented in the multiphysics and multiscale finite element library Alya[39,56], developed at the Barcelona Supercomputing Center and ELEM Biotech SL. Alya is optimized for efficient execution on supercomputers within a high-performance computing framework. Our study uses biventricular meshes comprising approximately 58 million tetrahedral linear elements, with an average mesh size of 300 μm, and a constant time-step size of 20 μs. Simulations are conducted on the Nord3 machine at the Barcelona Supercomputing Center, simulating three beats with a period of 1 s each. The computational cost of a single simulation, utilizing 672 cores, is approximately 3.5 h. To develop our emulators, we perform 450 simulations per sex, totaling 900 simulations and corresponding to 2.1 million CPU hours globally.

The emulators

To assess real-time QT prolongation caused by drugs, we develop sex-specific emulators that integrate a classifier followed by a regressor. The classifier enhances the robustness and accuracy of our emulators by filtering out non-arrhythmic and arrhythmic cases, reducing dataset complexity and discontinuity. This is crucial given the bifurcations and discontinuous responses in cardiac electrophysiological models, as noted by Ghosh et al.[58]. We also employ sex-specific ΔQT thresholds to avoid physiological unfeasible prolongations that could mislead the training and prediction processes. Our thresholds of 240 ms for males and 196 ms for females, which are designed to account for a QT interval of 600 ms, provide a conservative approach to identifying at-risk individuals. This methodology ensures the regressor operates on a more homogeneous subset of data, enhancing the reliability of the predictions.

Let $\mathscr{D} = (\boldsymbol{X}, \boldsymbol{y})$ be a dataset of size N, where $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$ is the input matrix, and $\boldsymbol{y} = (y_1, \ldots, y_N)$ is the output vector. The input of our model are the current blockages $(1 - \beta_k)$ defined in Eq. (3), with k = CaL, NaL, to, Ks, K1, Na, Kr. The output consists of a binary vector for the classifier (where 1 denotes an arrhythmic ECG and 0 non-arrhythmic ECG) and the ΔQT, computed as in Eq. (4), for the regressor. This results in a dataset of the following size: $\boldsymbol{X} \in \mathbb{R}^{7\times N}$, $\boldsymbol{y} \in \mathbb{R}^N$. We begin by splitting the dataset $\mathscr{D}$ into two parts: the training set $\mathscr{D}^{train}$ and the test set $\mathscr{D}^{test}$. The training set, comprising 72% of the data selected randomly, is used to tune the hyperparameters of both the classifiers and the regressor. The remaining 28% forms the test set, which is used to evaluate their performance. To ensure reproducibility and consistency across the consecutive steps of the emulators, we use the same sets $\mathscr{D}^{train}$ and $\mathscr{D}^{test}$ for classifiers and regressors. The input data are standardized in each step of the procedure, while the output vectors need to be normalized only for the regressor step. Data normalization is performed using the StandardScaler package from the scikit-learn Python library[84], which removes the mean and scales the data to unit variance.

To tune the optimal hyperparameters of the emulators, and to evaluate their performance, we employ a nested k-fold cross-validation strategy on the training set[85], with $k = 5$.

Table 10 | Hyperparameters description and set of selected possible values to explore with the automated exhaustive search for XGBC and KNC methods

Sex	Model	Hyperparameter	Values to explore
M	XGBC	n_estimators	{100, 200, 300}
		max_depth	{3, 5, 7}
		learning_rate	{0.01, 0.1, 0.3}
		subsample	{0.7, 0.9}
		colsample_bytree	{0.7, 0.9}
		gamma	{0, 0.1, 0.5}
		n_neighbors	{3, 5, 7, 9, 11}
F	KNC	weights	{uniform, distance}
		metric	{Euclidean, Manhattan, Minkowski}

Table 11 | Metrics used to evaluate the classifiers' performance[52]

Accuracy	Precision	Recall	F1 score
$\frac{TP + TN}{\#\text{predictions}}$	$\frac{TP}{TP + FP}$	$\frac{TP}{TP + FN}$	$\frac{2TP}{2TP + FP + FN}$

Classifier. We considered several classifiers, ranging from Random Forests, XGBC, Linear Regression, Support Vector Machine, and KNC. Of these, XGBC provided the most accurate results for males, while KNC performed best for females. For a detailed explanation of these methodologies, refer to[50,51].

The tuning of the optimal hyperparameters for the XGBC and the KNC methods is performed through the scikit-learn library and the GridSearchCV package. For XGBC and KNC we seek to determine the optimal hyperparameters among the possibilities reported in Table 10.

To define the evaluation metrics, we first introduce the following terminology to categorize the classifier outcomes:

True Positive (TP): predicted values correctly identified as positive.
True Negative (TN): predicted values correctly identified as negative.
False Positive (FP): predicted values identified as positive, while the actual ones are negative.
False Negative (FN): predicted values identified as negative, while the actual ones are positive.

Table 11 presents the metrics used for evaluation. Specifically, the optimal hyperparameters are selected to maximize the F1 score, as it is the most representative metric for our purposes. Indeed, the F1 score balances the impact of both false negatives and false positives, which are critical factors in medical applications, and offers greater robustness when dealing with imbalanced datasets[86]. After selecting the optimal hyperparameters, we use the test set to evaluate the quality of predictions on unseen data.

Gaussian process regression model. Here we focus on the regression model, whose goal is to capture the underlying relation between inputs and outputs. This model is designed to provide efficient approximations of the outputs for novel input data that are not contained in the training set. We evaluated several machine learning methods, including GPR model, multi-layer perceptron regressor, random forests, and XGBoost regression. Among these, the GPR model yielded the most accurate

results. A GPR model is a supervised learning method that assumes that the relation between inputs and outputs can be described by a Gaussian distribution. This means that for any set of input values, the corresponding outputs are assumed to follow a joint Gaussian distribution. We refer to[87] [Chapter 2] for further information on GPR models and to[27,29,31] for their application in cardiac modeling.

Consider again the training set $\mathscr{D}^{\text{train}}$ and the test set $\mathscr{D}^{\text{test}}$. GPR model calibration consists of tuning the hyperparameters of the kernel function on $\mathscr{D}^{\text{train}}$. This is done by maximizing the log-marginal-likelihood, that is the probability of reproducing the given output values with the emulators.

The tuning of the hyperparameters of the emulators is based again on a nested 5-fold cross-validation strategy on the training set. We perform the exhaustive search with the GridSearchCV class, which aims to find the optimal hyperparameters among the possibilities shown in Table 12.

To define the metrics used to tune the optimal hyperparameters and evaluate the model performance, we introduce some notation. Given the input data $\boldsymbol{X}^{\text{test}}$, we denote by $\boldsymbol{y}^{\text{pred}}$ the corresponding prediction of ΔQT intervals. The ground-truth output vector is denoted by $\boldsymbol{y}^{\text{test}}$. We select the hyperparameters that maximize the R^2 score, defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{N^{\text{test}}} \left(y_i^{\text{test}} - y_i^{\text{pred}}\right)^2}{\sum_{i=1}^{N^{\text{test}}} \left(y_i^{\text{test}} - \bar{y}_i\right)^2},$$

Table 12 | Hyperparameters description and set of selected possible values to perform the automated exhaustive search for GPR model

Hyperparameter	Values to explore
kernel	{RBF, Matérn 1.5, Matérn 2.5}
alpha	$\{10^{-5}, 10^{-3}, 10^{-2}, 10^{-1}, 1\}$
n_restarts_optimizer	{0, 5, 10, 20, 50}

See [87] [Section 4.2] for more details on kernel functions.

where $\bar{y}$ is the average of $\boldsymbol{y}^{\text{test}}$. Evaluating the accuracy of the emulators involves comparing the predicted values with the actual ones. To this end, we use the following error metrics, namely the mean absolute error (MAE), the mean absolute percentage error (MAPE), and the root mean squared error (RMSE), defined as:

$$\begin{aligned}
\varepsilon^{\text{MAE}} &= \frac{1}{N^{\text{test}}} \sum_{i=1}^{N^{\text{test}}} \left| y_i^{\text{test}} - y_i^{\text{pred}} \right|, \\
\varepsilon^{\text{MAPE}} &= \frac{100}{N^{\text{test}}} \sum_{i=1}^{N^{\text{test}}} \frac{\left| y_i^{\text{test}} - y_i^{\text{pred}} \right|}{\left| y_i^{\text{test}} \right|}, \\
\varepsilon^{\text{RMSE}} &= \sqrt{\frac{1}{N^{\text{test}}} \sum_{i=1}^{N^{\text{test}}} \left(y_i^{\text{test}} - y_i^{\text{pred}} \right)^2}.
\end{aligned}$$

Finally, we evaluate the performance of the model with the optimized hyperparameters on the test set.

Performing global sensitivity analysis

In this work, we perform GSA[49,88] twice: firstly to find the most influential channels from the preliminary emulators, allowing us to later enhance the database with higher blockades; secondly, to identify and prioritize influential inputs from our enhanced emulators. GSA is divided into the following steps:

1. *Sampling with Sobol' sequences*: Sobol' sequences are quasi-random sequences that ensure a more even and thorough exploration of the parameter space compared to traditional random sampling methods. We use these sequences to generate a set of diverse input parameter combinations (ionic channels' blockades). Specifically, we sample $2^{12} = 4096$ points in the parameter space (this number is typically chosen as a power of 2 to ensure robust coverage and convergence properties).
2. *Model evaluation*: For each sampled parameter combination, we evaluate our emulators to obtain corresponding model outputs (ΔQT).
3. *Variance decomposition and sensitivity indices*: Sobol' sensitivity analysis decomposes the total variance of the model output into contributions from individual parameters (first-order sensitivity indices,

Table 13 | *IC*50, expressed in nmol/L, is the concentration of the drug that inhibits 50% of its target ion channel activity; *h* (Hill coefficient, dimensionless) describes the steepness of the drug's concentration-response curve

		Moxifloxacin	Ondansetron	Dofetilide	Verapamil		
I_{CaL}	*IC*50	–	22551	–	202	201.8	200
	h	–	0.8	–	1.1	1.1	0.8
I_{Kr}	*IC*50	93041	1492	4.9	499	288	250
	h	0.6	1.0	0.9	1.1	1.0	0.89
I_{K1}	*IC*50	–	–	–	–	$349 \cdot 10^6$	–
	h	–	–	–	–	0.3	–
I_{to}	*IC*50	–	–	18.8	–	13429.2	–
	h	–	–	0.8	–	0.8	–
I_{Ks}	*IC*50	50321	–	–	–	–	–
	h	1.0	–	–	–	–	–
I_{NaL}	*IC*50	382337	19181	–	–	7028	–
	h	1.1	1.0	–	–	1.0	–
I_{Na}	*IC*50	–	–	–	–	–	32500
	h	–	–	–	–	–	1.33
Ionic block profile		38	38	54	38	54	55
$\tilde{C}$ [ng/mL]		700, 1500, 2900 (F) 500, 1000, 2100 (M) (Fig. 10)	30, 100, 300 (Fig. 10)	0.43, 0.92 (Fig. 9) 0.1, 0.3, 1 (Fig. 10)	99.6, 199.2, 398.4, 750 (Figs. 10 & 11)		
f_u [%]		65	27	35	12		

The ionic block profile refers to the sources of these two parameters. $\tilde{C}$ is the total concentration of the drug in the plasma, and f_u is the free fraction of the drug. For moxifloxacin, different concentrations are used for males and females, according to the reported higher maximum concentrations observed in females[91]. The concentrations were selected based on reported plasma levels in clinical trials. Different dofetilide concentrations are used for the study in Fig. 9 and the benchmark analysis in Fig. 10. The former focuses on evaluating emulator performance across various physiological conditions, whereas the latter aims to thoroughly assess drug safety profiles. The illustrated verapamil response in Fig. 10 incorporates data exclusively from Crumb et al.[38], among the reported ionic block profile sources.

S_1) and their interactions (total-order sensitivity indices, S_T). This decomposition provides quantitative measures of the relative influence of each parameter and interaction on the variability of the model output.

We run GSA simulations using the SALib Python library[89,90], performing two key analyses. The first examines how input parameters influence the emulator outputs (i.e., ΔQT), while the second identifies the most influential factors contributing to QT prolongation risk. In this case, after computing the ΔQT, we categorize the results into a binary vector, labeling them as high or low risk based on whether the predicted ΔQT exceeds the specified high-risk threshold of 50 ms[47].

For the preliminary version of the emulators, we focus on the second analysis. Using Sobol' sequences, we perform the GSA to identify the three ionic channels that have the most significant impact on the drug-induced high risk of QT prolongation. The primary objective is to isolate regions within the sample space that surround the thresholds for arrhythmic events, thereby enhancing the training dataset in this critical area. Sobol' sensitivity analysis is particularly well-suited for this purpose due to its variance-based approach, which systematically decomposes the total variance in model output into contributions from individual parameters and their interactions.

After developing and tuning the final version of the emulators, we repeat the GSA to further explore the impact of ionic channel blockades on QT prolongation. The enhanced version of the emulators is used to perform both the analyzes outlined above.

Comparison of in silico and in vivo clinical trials for benchmark drugs

We evaluate the emulator performance by comparing their predictions of C-ΔQT for various benchmark drugs with simulator results and publicly available clinical data[53]. The simulator results, shown in our analysis, follow the population approach outlined in ref. 60, which incorporates phenotypic variability, as we presented in ref. 23. Additionally, the simulator incorporates anatomical variability using data from the Visible Heart Lab at the University of Minnesota[76], which features a comprehensive collection of patient cases with varying BMI, age, and anatomical characteristics. In contrast, our emulators do not account for any of these sources of variability.

To assess the emulator accuracy, we perform linear regressions on the C-ΔQT relationships derived from clinical data, simulator results, and emulator results. For the clinical data, the linear regression is based on ΔΔQTc. This validation analysis is divided into three parts, corresponding to Figures 9, 10 and 11 respectively. Table 13 details the blockades and concentrations used as input data for all of them.

Data availability

The data generated and analyzed during this study are proprietary to ELEM Biotech and cannot be publicly shared due to commercial confidentiality. However, access to the data may be considered on a case-by-case basis. Requests for access can be directed to ELEM Biotech compliance department, and will be subject to a data use agreement ensuring compliance with relevant regulations and restrictions.

Code availability

ELEM Biotech owns the commercial rights to Alya, the computational finite element solver employed in this study for the simulator. However, the methodology can be replicated using any finite element solver given all the parameterization information provided in this paper. The emulators are implemented in Python using the scikit-learn library, and all necessary steps for replication are described in the Methodology section.

Received: 27 September 2024; Accepted: 3 December 2024;
Published online: 26 December 2024

References

1. National Health Service. Arrhythmia (2021). https://www.nhs.uk/conditions/arrhythmia/. Accessed July 31, 2024.
2. Cubeddu, L. QT Prolongation and Fatal Arrhythmias: A Review of Clinical Implications and Effects of Drugs. *American journal of therapeutics* **10**, 452–7 (2003).
3. Sager, P. T., Gintant, G., Turner, J. R., Pettit, S. & Stockbridge, N. Rechanneling the cardiac proarrhythmia safety paradigm: A meeting report from the cardiac safety research consortium. *American Heart Journal* **167**, 292–300 (2014).
4. Gintant, G., Sager, P. T. & Stockbridge, N. Evolution of strategies to improve preclinical cardiac safety testing. *Nature Reviews Drug Discovery* **15**, 457–471 (2016).
5. Kaye, G. & Lemery, R.*Fast Facts: Cardiac Arrhythmias* (S. Karger AG, 2018).
6. Valentin, J.-P. et al. The Challenges of Predicting Drug-Induced QTc Prolongation in Humans. *Toxicol. Sci.* **187**, 3–24 (2022).
7. Colatsky, T. et al. The Comprehensive in Vitro Proarrhythmia Assay (CiPA) initiative - Update on progress. *Journal of Pharmacological and Toxicological Methods* **81**, 15–20 (2016).
8. Hwang, M., Lim, C.-H., Leem, C. H. & Shim, E. B. In silico models for evaluating proarrhythmic risk of drugs. *APL Bioengineering* **4**, 021502 (2020).
9. Mirams, G. R. et al. Simulation of multiple ion channel block provides improved early prediction of compounds' clinical torsadogenic risk. *Cardiovascular research* **91**, 53–61 (2011).
10. Li, Z. et al. Assessment of an in silico mechanistic model for proarrhythmia risk prediction under the CiPA initiative. *Clinical Pharmacology & Therapeutics* **105**, 466–475 (2019).
11. Passini, E. et al. Human in silico drug trials demonstrate higher accuracy than animal models in predicting clinical pro-arrhythmic cardiotoxicity. *Front. Physiol.* **8**, 668 (2017).
12. Abbasi, M., Small, B. G., Patel, N., Jamei, M. & Polak, S. Early assessment of proarrhythmic risk of drugs using the in vitro data and single-cell-based in silico models: proof of concept. *Toxicology Mechanisms and Methods* **27**, 88–99 (2017).
13. Passini, E. et al. Drug-induced shortening of the electromechanical window is an effective biomarker for in silico prediction of clinical risk of arrhythmias. *British Journal of Pharmacology* **176**, 3819–3833 (2019).
14. Romero, L. et al. In silico QT and APD prolongation assay for early screening of drug-induced proarrhythmic risk. *Journal of Chemical Information and Modeling* **58**, 867–878 (2018).
15. Zemzemi, N. et al. Computational assessment of drug-induced effects on the electrocardiogram: from ion channel to body surface potentials. *British Journal of Pharmacology* **168**, 718–733 (2013).
16. Hwang, M. et al. Three-dimensional heart model-based screening of proarrhythmic potential by in silico simulation of action potential and electrocardiograms. *Fron. Physiol.* **10**, 1139 (2019).
17. Sahli Costabal, F., Yao, J. & Kuhl, E. Predicting drug-induced arrhythmias by multiscale modeling. *International Journal for Numerical Methods in Biomedical Engineering* **34**, e2964 (2018).
18. Okada, J.-I. et al. Arrhythmic hazard map for a 3D whole-ventricle model under multiple ion channel block. *British Journal of Pharmacology* **175**, 3435–3452 (2018).
19. Wilhelms, M., Rombach, C., Scholz, E. P., Dössel, O. & Seemann, G. Impact of amiodarone and cisapride on simulated human ventricular electrophysiology and electrocardiograms. *EP Europace* **14**, v90–v96 (2012).
20. Cranford, J. P. et al. Efficient computational modeling of human ventricular activation and its electrocardiographic representation: A sensitivity study. *Cardiovascular engineering and technology* **9**, 447–467 (2018).
21. Peirlinck, M., Sahli Costabal, F. & Kuhl, E. Sex differences in drug-induced arrhythmogenesis. *Front. Physiol.* **12**, 708435 (2021).
22. Llopis-Lorente, J. et al. Combining pharmacokinetic and electrophysiological models for early prediction of drug-induced

arrhythmogenicity. *Computer Methods and Programs in Biomedicine* **242**, 107860 (2023).
23. Aguado-Sierra, J. et al. Virtual clinical QT exposure-response studies – a translational computational approach. *Journal of Pharmacological and Toxicological Methods* **126**, 107498 (2024).
24. Peirlinck, M., Lee, J., Fovargue, D. & Kuhl, E. Sex matters: A comprehensive comparison of female and male hearts. *Frontiers in Physiology* **13**, 831179 (2022).
25. Darpo, B. et al. Are women more susceptible than men to drug-induced QT prolongation? Concentration-QTc modelling in a phase 1 study with oral rac-sotalol. *British Journal of Clinical Pharmacology* **77**, 522–531 (2014).
26. Subasi, A. & Subasi, M. E. Digital twins in healthcare and biomedicine. In *Artificial Intelligence, Big Data, Blockchain and 5G for the Digital Transformation of the Healthcare Industry*, 365–401 (Elsevier, 2024).
27. Longobardi, S. et al. Predicting left ventricular contractile function via Gaussian process emulation in aortic-banded rats. *Philosophical Transactions of the Royal Society A* **378**, 20190334 (2020).
28. Fresca, S., Manzoni, A., Dedè, L. & Quarteroni, A. Deep learning-based reduced order models in cardiac electrophysiology. *PloS one* **15**, e0239416 (2020).
29. Karabelas, E. et al. Global sensitivity analysis of four chamber heart hemodynamics using surrogate models. *IEEE Transactions on Biomedical Engineering* **69**, 3216–3223 (2022).
30. Regazzoni, F., Salvador, M., Dedè, L. & Quarteroni, A. A machine learning method for real-time numerical simulations of cardiac electromechanics. *Computer methods in applied mechanics and engineering* **393**, 114825 (2022).
31. Strocchi, M. et al. Cell to whole organ global sensitivity analysis on a four-chamber heart electromechanics model using gaussian processes emulators. *PLOS Computational Biology* **19**, e1011257 (2023).
32. Salvador, M., Regazzoni, F., Dedè, L. & Quarteroni, A. Fast and robust parameter estimation with uncertainty quantification for the cardiac function. *Computer Methods and Programs in Biomedicine* **231**, 107402 (2023).
33. Cicci, L., Fresca, S., Manzoni, A. & Quarteroni, A. Efficient approximation of cardiac mechanics through reduced-order modeling with deep learning-based operator approximation. *International Journal for Numerical Methods in Biomedical Engineering* **40**, e3783 (2024).
34. Salvador, M. et al. Whole-heart electromechanical simulations using latent neural ordinary differential equations. *NPJ Digital Medicine* **7**, 90 (2024).
35. Yin, M. et al. A scalable framework for learning the geometry-dependent solution operators of partial differential equations. *Nat. Comput. Sci.* **4**, 928–940 (2024).
36. Sahli Costabal, F., Matsuno, K., Yao, J., Perdikaris, P. & Kuhl, E. Machine learning in drug development: Characterizing the effect of 30 drugs on the QT interval using Gaussian process regression, sensitivity analysis, and uncertainty quantification. *Computer Methods in Applied Mechanics and Engineering* **348**, 313–333 (2019).
37. Grandits, T. et al. Neural network emulation of the human ventricular cardiomyocyte action potential for more efficient computations in pharmacological studies. *Elife* **12**, RP91911 (2024).
38. Crumb, W. J., Vicente, J., Johannesen, L. & Strauss, D. G. An evaluation of 30 clinical drugs against the comprehensive in vitro proarrhythmia assay (CiPA) proposed ion channel panel. *Journal of Pharmacological and Toxicological Methods* **81**, 251–262 (2016). Focused Issue on Safety Pharmacology.
39. Vázquez, M. et al. Alya: Multiphysics engineering simulation toward exascale. *Journal of computational science* **14**, 15–27 (2016).
40. Houltz, B. et al. Electrocardiographic and clinical predictors of Torsades de Pointes induced by almokalant infusion in patients with chronic atrial fibrillation or flutter: A prospective study. *Pacing and Clinical Electrophysiology* **21**, 1044–1057 (1998).
41. O'Hara, T., Virág, L., Varró, A. & Rudy, Y. Simulation of the undiseased human cardiac ventricular action potential: model formulation and experimental validation. *PLoS computational biology* **7**, e1002061 (2011).
42. Park, J.-S., Jeon, J.-Y., Yang, J.-H. & Kim, M.-G. Introduction to in silico model for proarrhythmic risk assessment under the cipa initiative. *Translational and clinical pharmacology* **27**, 12 (2019).
43. Garnett, C. et al. Scientific white paper on concentration-QTc modeling. *Journal of Pharmacokinetics and Pharmacodynamics* **45**, 1–15 (2018).
44. Mirams, G. Action potential durations and QT intervals. https://mirams.wordpress.com/2014/03/21/apd_vs_qt/ (2014). Accessed: 2024-8-1.
45. Lewis-Beck, C. & Lewis-Beck, M.*Applied regression: An introduction*, vol. 22 (Sage publications, 2015).
46. Strauss, D. G. et al. Comprehensive in vitro proarrhythmia assay (CiPA) update from a cardiac safety research consortium/health and environmental sciences institute/FDA meeting. *Therapeutic Innovation & Regulatory Science* **53**, 519–525 (2019).
47. Roden, D. M. Drug-induced prolongation of the QT interval. *The New England journal of medicine* **350**, 1013–1022 (2004).
48. McKay, M. D., Beckman, R. J. & Conover, W. J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **42**, 55–61 (2000).
49. Saltelli, A. et al. *Global sensitivity analysis. The primer* (John Wiley & Sons, Ltd., Chichester, 2008).
50. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794 (ACM, 2016).
51. Cover, T. & Hart, P. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory* **13**, 21–27 (1967).
52. Hossin, M. & Sulaiman, M. N. A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining & Knowledge Management Process* **5**, 1 (2015).
53. Darpo, B. et al. Results from the IQ-CSRC prospective study support replacement of the thorough QT study by QT assessment in the early clinical phase. *Clinical Pharmacology and Therapeutics* **97**, 326–335 (2015).
54. Dutta, S. et al. Optimization of an in silico cardiac cell model for proarrhythmia risk assessment. *Front. Physiol.* **8**, 616 (2017).
55. Kramer, J. et al. Mice models: superior to the herg model in predicting torsade de pointes. *Scientific reports* **3**, 2100 (2013).
56. Santiago, A. et al. Fully coupled fluid-electro-mechanical model of the human heart for supercomputers. *International journal for numerical methods in biomedical engineering* **34**, e3140 (2018).
57. Vicente, J., Zheng, N., Bende, G. & Garnett, C. Chapter 72 - Sex differences in drug-induced QT prolongation. In Malik, M. (ed.) *Sex and Cardiac Electrophysiology*, 799–806 (Academic Press, 2020).
58. Ghosh, S., Gavaghan, D. & Mirams, G. Gaussian process emulation for discontinuous response surfaces with applications for cardiac electrophysiology models. arXiv (2018).
59. Darpo, B. Detection and reporting of drug-induced proarrhythmias: room for improvement. *EP Europace* **9**, iv23–iv36 (2007).
60. Muszkiewicz, A. et al. Variability in cardiac electrophysiology: Using experimentally-calibrated populations of models to move beyond the single virtual physiological human paradigm. *Progress in Biophysics and Molecular Biology* **120**, 115–127 (2016).
61. Pater, C. Methodological considerations in the design of trials for safety assessment of new drugs and chemical entities. *Current controlled trials in cardiovascular medicine* **6**, 1 (2005).
62. Levrero-Florencio, F. et al. Sensitivity analysis of a strongly-coupled human-based electromechanical cardiac model: Effect of mechanical

parameters on physiologically relevant biomarkers. *Computer Methods in Applied Mechanics and Engineering* **361**, 112762 (2020).
63. Margara, F. et al. In-silico human electro-mechanical ventricular modelling and simulation for drug-induced pro-arrhythmia and inotropic risk assessment. *Progress in Biophysics and Molecular Biology* **159**, 58–74 (2021).
64. Gonzalez-Martin, P. et al. Ventricular anatomical complexity and sex differences impact predictions from electrophysiological computational models. *Plos one* **18**, e0263639 (2023).
65. López-Yunta, M. et al. Infarct transmurality as a criterion for first-line endo-epicardial substrate-guided ventricular tachycardia ablation in ischemic cardiomyopathy. *EP Europace* **21**, 55–63 (2019).
66. Wang, Z. J. et al. Human biventricular electromechanical simulations on the progression of electrocardiographic and mechanical abnormalities in post-myocardial infarction. *EP Europace* **23**, i143–i152 (2021).
67. Bragard, J. R. et al. Cardiac computational modelling. *Revista Española de Cardiología (English Edition)* **74**, 65–71 (2021).
68. Gil, D. et al. What a difference in biomechanics cardiac fiber makes. In *Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges: Third International Workshop, STACOM 2012. Lecture Notes in Computer Science*, 253–260 (Springer, 2013).
69. Aguado-Sierra, J. et al. *HPC Framework for Performing in Silico Trials Using a 3D Virtual Human Cardiac Population as Means to Assess Drug-Induced Arrhythmic Risk*, vol. 2716 of *Methods in Molecular Biology*, chap. 14 (Springer US, New York, NY, 2024).
70. Vázquez, M. et al. A massively parallel computational electrophysiology model of the heart. *International journal for numerical methods in biomedical engineering* **27**, 1911–1929 (2011).
71. Lafortune, P., Arís, R., Vázquez, M. & Houzeaux, G. Coupled electromechanical model of the heart: parallel finite element formulation. *International journal for numerical methods in biomedical engineering* **28**, 72–86 (2012).
72. Vázquez, M. et al. Alya Red CCM: HPC-based cardiac computational modelling. In *Selected topics of computational and experimental fluid mechanics*,189–207 (Springer, 2015).
73. Johannesen, L., Vicente, J., Hosseini, M. & Strauss, D. G. Automated algorithm for j-tpeak and tpeak-tend assessment of drug-induced proarrhythmia risk. *PloS one* **11**, e0160502 (2016).
74. Fogli Iseppe, A. et al. Sex-specific classification of drug-induced Torsade de Pointes susceptibility using cardiac simulations and machine learning. *Clinical Pharmacology & Therapeutics* **110**, 380–391 (2021).
75. Leishman, D. J. et al. *Journal of Pharmacological and Toxicological Methods* **128**, 107524 (2024).
76. U. of Minnesota Atlas of Human Cardiac Anatomy. https://www.vhlab.umn.edu/atlas/histories/histories.shtml (2021). [Accessed 02-08-2024].
77. Franzone, P. C., Pavarino, L. F. & Scacchi, S.*Mathematical cardiac electrophysiology*, vol. 13 (Springer, 2014).
78. Passini, E. et al. Mechanisms of pro-arrhythmic abnormalities in ventricular repolarisation and anti-arrhythmic therapies in human hypertrophic cardiomyopathy. *Journal of molecular and cellular cardiology* **96**, 72–81 (2016).
79. Durrer, D. et al. Total excitation of the isolated human heart. *Circulation* **41**, 899–912 (1970).
80. Doste, R. et al. A rule-based method to model myocardial fiber orientation in cardiac biventricular geometries with outflow tracts. *International Journal for Numerical Methods in Biomedical Engineering* **35**, e3185 (2019).
81. Yang, P.-C. & Clancy, C. E. In silico prediction of sex-based differences in human susceptibility to cardiac ventricular tachyarrhythmias. *Frontiers in physiology* **3**, 33341 (2012).
82. Plonsey, R. & Barr, R. C.*Bioelectricity: A Quantitative Approach* (Springer, New York, NY, 2007).
83. Desai, M., Li, L., Desta, Z., Malik, M. & Flockhart, D. Variability of heart rate correction methods for the qt interval. *British Journal of Clinical Pharmacology* **55**, 511–517 (2003).
84. Pedregosa, F. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).
85. Varma, S. & Simon, R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics* **7**, 1–8 (2006).
86. Zheng, A.*Evaluating machine learning models: a beginner's guide to key concepts and pitfalls* (O'Reilly Media, 2015).
87. Rasmussen, C. E. & Williams, C. K. I.*Gaussian processes for machine learning*. Adaptive Computation and Machine Learning (MIT Press, Cambridge, MA, 2006).
88. Saltelli, A. Making best use of model evaluations to compute sensitivity indices. *Computer physics communications* **145**, 280–297 (2002).
89. Herman, J. & Usher, W. SALib: An open-source Python library for Sensitivity Analysis. *The Journal of Open Source Software***2** (2017). https://doi.org/10.21105/joss.00097.
90. Iwanaga, T., Usher, W. & Herman, J. Toward SALib 2.0: Advancing the accessibility and interpretability of global sensitivity analyses. *Socio-Environmental Systems Modelling* **4**, 18155 (2022).
91. Florian, J. A., Tornøe, C. W., Brundage, R., Parekh, A. & Garnett, C. E. Population pharmacokinetic and concentration – QTc models for moxifloxacin: Pooled analysis of 20 thorough QT studies. *The Journal of Clinical Pharmacology* **51**, 1152–1162 (2011).
92. Vicente, J. et al. Assessment of multi-ion channel block in a phase I randomized study design: Results of the CiPA phase I ECG biomarker validation study. *Clinical Pharmacology & Therapeutics* **105**, 943–953 (2019).

Acknowledgements

This project was partially funded by the European Union - EIC Project No 190134524: "ELEM Virtual Heart Populations for Supercomputers" (ELVIS). Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or EISMEA. Neither the European Union nor the granting authority can be held responsible for them.

Author contributions

P.D., A.Z., L.B., and C.B. contributed equally to the conception, design, data acquisition, implementation of the computer code and supporting algorithms, analysis, validation, results visualization, writing of the original manuscript, and its subsequent review. B.D. assisted with study design, clinical validation, and manuscript review. C.M. and M.V. contributed to funding acquisition, provided computational resources, and reviewed the manuscript. J.A. supervised the work and contributed to the manuscript review. All authors have read and approved the manuscript.

Competing interests

P.D., A.Z., L.B., C.B., B.D. and J.A. declare no competing interests. M.V. is CTO and co-founder of ELEM Biotech and C.M. is CEO and co-founder of ELEM Biotech.

Additional information

Correspondence and requests for materials should be addressed to Paula Dominguez-Gomez.