

Reviewed Preprint

Revised by authors after peer review.

About eLife's process

Reviewed preprint version 2

February 13, 2024 (this version)

Reviewed preprint version 1

December 4, 2023

Sent for peer review

August 22, 2023

Posted to preprint server

August 17, 2023

Neural network emulation of the human ventricular cardiomyocyte action potential: a tool for more efficient computation in pharmacological studies

Thomas Grandits , Christoph M. Augustin, Gundolf Haase, Norbert Jost, Gary R. Mirams, Steven A. Niederer, Gernot Plank, András Varró, László Virág, Alexander Jung

Department of Mathematics and Scientific Computing, University of Graz • NAWI Graz, University of Graz • Gottfried Schatz Research Center for Cell Signaling, Metabolism and Aging - Division of Medical Physics and Biophysics, Medical University of Graz • BioTechMed-Graz • Department of Pharmacology and Pharmacotherapy, University of Szeged • HUN-REN-TKI, Research Group of Pharmacology • Centre for Mathematical Medicine & Biology, School of Mathematical Sciences, University of Nottingham • Division of Imaging Sciences & Biomedical Engineering, King's College London

 https://en.wikipedia.org/wiki/Open_access

 Copyright information

Abstract

Abstract

Computer models of the human ventricular cardiomyocyte action potential (AP) have reached a level of detail and maturity that has led to an increasing number of applications in the pharmaceutical sector. However, interfacing the models with experimental data can become a significant computational burden. To mitigate the computational burden, the present study introduces a neural network (NN) that emulates the AP for given maximum conductances of selected ion channels, pumps, and exchangers. Its applicability in pharmacological studies was tested on synthetic and experimental data. The NN emulator potentially enables massive speed-ups compared to regular simulations and the forward problem (find drugged AP for pharmacological parameters defined as scaling factors of control maximum conductances) on synthetic data could be solved with average root-mean-square errors (RMSE) of 0.47 mV in normal APs and of 14.5 mV in abnormal APs exhibiting early afterdepolarizations (72.5% of the emulated APs were aligning with the abnormality, and the substantial majority of the remaining APs demonstrated pronounced proximity). This demonstrates not only very fast and mostly very accurate AP emulations but also the capability of accounting for discontinuities, a major advantage over existing emulation strategies. Furthermore, the inverse problem (find pharmacological parameters for control and drugged APs through optimization) on synthetic data could be solved with high accuracy shown by a maximum RMSE of 0.21 in the estimated pharmacological parameters. However, notable mismatches were observed between pharmacological parameters estimated from experimental data and distributions obtained from the Comprehensive in vitro Proarrhythmia Assay initiative. This reveals larger inaccuracies which can be attributed

particularly to the fact that small tissue preparations were studied while the emulator was trained on single cardiomyocyte data. Overall, our study highlights the potential of NN emulators as powerful tool for an increased efficiency in future quantitative systems pharmacology studies.

eLife assessment

This **valuable** prospective study develops a new tool to accelerate pharmacological studies by using neural networks to emulate the human ventricular cardiomyocyte action potential. The evidence supporting the conclusions is **convincing**, based on using a large and high-quality dataset to train the neural network emulator. There are nevertheless a few areas in which the article may be improved through validating the neural network emulators against extensive experimental data. In addition, the article may be improved through delineating the exact speed-up achieved and the scope for acceleration. The work will be of broad interest to scientists working in cardiac simulation and quantitative system pharmacology.

1 Introduction

Computer models of human physiology are becoming increasingly detailed and mature and the area of ventricular cardiomyocyte electrophysiology (EP) is one of the most advanced. The most updated models include fine representations of ion movements through various important channels, pumps, and exchangers, and take the complex handling of intracellular calcium accurately into account [Grandi et al., 2010 [↗](#), O'Hara et al., 2011 [↗](#), Tomek et al., 2019 [↗](#), Bartolucci et al., 2020 [↗](#)]. While these models have individual strengths and limitations in replicating different aspects of physiology, pathology, and pharmacology [Corrado et al., 2020 [↗](#), Amuzescu et al., 2021 [↗](#)], their degree of credibility has reached a level that has led to an increasing number of applications in academia and beyond. This holds in particular for the pharmaceutical sector, where much effort is spent on using computer modeling to reduce traditional preclinical and clinical methodologies for assessing the efficacy and safety of novel drug candidates [Mirams et al., 2011 [↗](#), Passini et al., 2017 [↗](#), Li et al., 2019 [↗](#), Passini et al., 2021 [↗](#)]. To improve the regulatory assessment of a drug's proarrhythmic potential, the Comprehensive in Vitro Proarrhythmia Assay (CiPA) was proposed in 2013 following a workshop at the US Food and Drug Administration [Sager et al., 2014 [↗](#), Colatsky et al., 2016 [↗](#), Strauss et al., 2021 [↗](#)]. A central component is a computer model of human ventricular cardiomyocyte EP that is coupled to a pharmacological model describing the interaction between a given drug and multiple arrhythmia-relevant channels [Dutta et al., 2017 [↗](#), Li et al., 2017 [↗](#), Li et al., 2019 [↗](#)]. For a given drug, experimental channel block data are collected to inform the pharmacological model and corresponding simulations of the action potentials (AP; time course of the transmembrane potential) are performed to predict the proarrhythmic risk based on a mechanistically motivated biomarker [Chang et al., 2017 [↗](#), Li et al., 2017 [↗](#), Li et al., 2019 [↗](#)]. The prediction is then compared with findings in experimental [Blinova et al., 2018 [↗](#)] and clinical [Vicente et al., 2017 [↗](#)] studies. To compute the drugged AP for given pharmacological parameters is a forward problem, while the corresponding inverse problem is to find pharmacological parameters for given control (before drug administration) and drugged AP. Some relevant examples for the latter have been presented by [Bottino et al., 2006 [↗](#)] who estimated pharmacological parameters from APs of canine Purkinje fibers and by Tveito *et al.* who estimated pharmacological parameters from AP biomarkers measured in human induced pluripotent stem cell-derived cardiomyocytes [Tveito et

al., 2018 [\[1\]](#) and several animal ventricular cardiomyocytes [Tveito et al., 2020 [\[2\]](#)]. Furthermore, [Jæger et al., 2021 [\[3\]](#)] identified the optimal polypharmacological treatment for recovering APs of mutant ventricular cardiomyocytes based on biomarkers of simulated wild type and mutant APs.

When the models are interfaced with experimental data, attention should be paid to the inherent uncertainty in the data that results from beat-to-beat variability (intrinsic variability), cell-to-cell variability (extrinsic variability), and measurement errors (observational uncertainty) [Mirams et al., 2016 [\[4\]](#)]. Uncertainty propagates through the given problem from APs to estimated parameters or from parameters to predicted APs and must be properly quantified to draw reliable conclusions from the results. Multiple methodologies exist for this purpose [Oakley and O'Hagan, 2004 [\[5\]](#), Mirams et al., 2016 [\[4\]](#), Sher et al., 2022 [\[6\]](#)] but usually require many simulations, which even for ordinary differential equation (ODE)-based models of cardiomyocyte EP can become a significant computational burden when considering that each simulation includes a substantial number of beats to reach the model's limit cycle, (also often called steady state). To overcome this problem, surrogate models have emerged which approximate (emulate) chosen outputs for given inputs multiple orders of magnitude faster. In line with uncertainty quantification literature, the cardiomyocyte EP model is from now on termed the “*simulator*”, whereas the surrogate model is termed the “*emulator*”. Earlier work has reported on an emulator based on linear interpolation of a multi-dimensional lookup table [Mirams et al., 2014 [\[7\]](#)] and more recently, Gaussian process (GP) emulators have become popular. Their key advantage is that in- and outputs are modeled as random distributions which allows for rapid sampling of the posterior distributions [Chang et al., 2015 [\[8\]](#), Johnstone et al., 2016 [\[9\]](#), Coveney and Clayton, 2018 [\[10\]](#), Ghosh et al., 2018 [\[11\]](#), Rasmussen and Williams, 2019 [\[12\]](#), Coveney et al., 2021 [\[13\]](#)] and while outputs of recently published GP emulators were relevant biomarkers of the AP [Chang et al., 2015 [\[8\]](#), Johnstone et al., 2016 [\[9\]](#), Coveney and Clayton, 2018 [\[10\]](#), Ghosh et al., 2018 [\[11\]](#), Coveney et al., 2021 [\[13\]](#)], the emulation of the entire AP can also be realized, e.g. through dimensionality reduction techniques such as the principal component analysis or regressing state-transition models [Mohammadi et al., 2019 [\[14\]](#)]. However, GP emulators are not well suited to capture discontinuities of the response surface as standard Gaussian process emulators assume a smooth and continuous response to changes in parameter values. Applying GPs for modeling discontinuous functions therefore remains a largely open problem. Thus, they may fail to capture AP abnormalities, which is a particular drawback for pharmacological studies where bifurcations in behavior such as early afterdepolarizations (EAD) can occur [Ghosh et al., 2018 [\[11\]](#)]. To address this, [Ghosh et al., 2018 [\[11\]](#)] presented a two-step approach for the emulation of the AP duration at 90% repolarization that first sets up a GP for the location of discontinuities and then fits separate GP emulators for the output of interest either side of these boundaries. In contrast, it has been proven that neural networks (NN) can approximate even discontinuous functions with arbitrary precision in theory [Hornik et al., 1989 [\[15\]](#)], while recent works using NNs show empirically promising results for modeling partial differential equations containing discontinuities [Jagtap et al., 2020 [\[16\]](#)]. These features render NN emulators suitable emulation candidates and while [Lei and Mirams, 2021 [\[17\]](#)] have recently investigated NN emulation of hERG channel kinetics, [Jeong et al., 2023 [\[18\]](#)] proposed a neural network using AP shapes as input for the prediction of a drug's proarrhythmic risk. However, to the best of our knowledge, NN emulators have not yet been used as surrogate for cardiomyocyte EP models.

The present study introduces NN emulation of the human ventricular cardiomyocyte AP and investigates the applicability in pharmacological studies. To this end, a NN emulator was developed based on data generated using a state-of-the-art simulator [Tomek et al., 2019 [\[19\]](#), Tomek et al., 2020 [\[20\]](#)] and the evaluation was done for forward and inverse problems on synthetic and experimental data.

2 Methods and material

The methodology of this study including the development of the emulator and the evaluation is outlined in **Figure 1**.

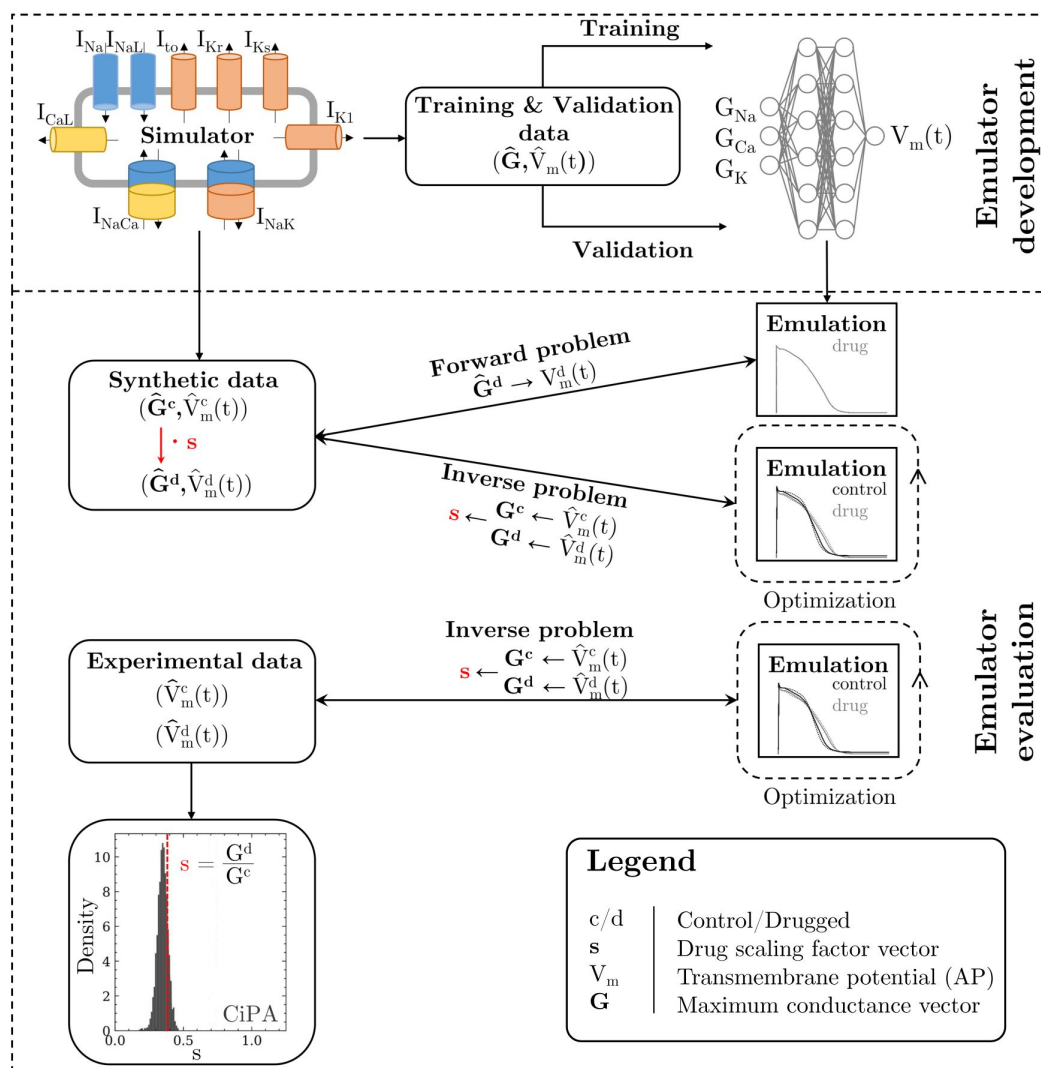


Figure 1

Methodology of this study including the emulator development and the evaluation. The simulator is visualized by a schematic human ventricular cardiomyocyte that includes all currents considered for the emulator training. Inputs of the emulator (see **Figure 3**) are the corresponding maximum conductances (G) but for the sake of illustration, only three representatives are shown. Output is the AP ($V_m(t)$). Training and validation data (maximum conductances $\hat{G}$ and APs $\hat{V}_m(t)$) were generated by the simulator. The evaluation was performed for forward and inverse problems and to this end, the pharmacological parameter s was introduced. This describes the interaction between the drug and a given target and was defined as scaling factor of the respective maximum conductance in control conditions (s & 1: block, $s = 1$: no effect, $s > 1$: enhancement). Synthetic data (control maximum conductances and drugged maximum conductances obtained through scaling, and control and drugged APs) and experimental data (control and drugged APs) were used for the evaluation [Orvos et al., 2019]. The forward problem was only solved for synthetic data, whereas the inverse problem was solved for both synthetic and experimental data. When experimental data were used, estimated pharmacological parameters were compared to distributions derived from data published within the CiPA initiative [Li et al., 2017, Chang et al., 2017].

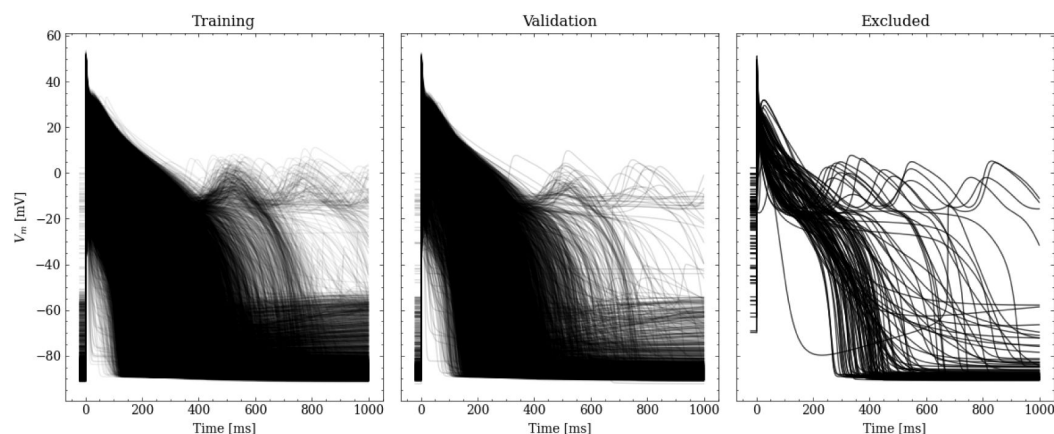


Figure 2

Processed APs used for training and validation (left and center). Additionally we show the excluded APs on the right (see text for description of the exclusion criteria).

2.1 Simulator

The simulator of [Tomek et al., 2020] (ToR-ORD-dynCI simulator) was used. This is available in CARPentry [Vigmond et al., 2008] and was implemented based on the published CellML file for the endocardial subtype¹. Simulations were performed in CARPentry with the single cell tool bench. To compute the gating variables, the Rush-Larsen Method [Rush and Larsen, 1978] was employed, which uses an analytical solution assuming fixed voltage over a small timestep, and the remaining variables were computed by the Forward Euler method. To ensure low computational cost, we found the maximum solver and sampling time steps that still produce accurate results as follows. Various solver and sampling time steps were applied to generate APs and the biomarkers (Appendix A) used in this study were computed and compared with those that correspond to the minimum time steps (solver: 0.005 ms; sampling: 0.01 ms). We considered the time steps with only 2% relative difference for all AP biomarkers (solver: 0.01 ms; sampling: 0.05 ms) to offer a sufficiently good approximation. APs were stimulated at a pacing cycle length of 1000 ms using 1 ms long rectangular transmembrane current density pulses of $53 \frac{\mu A}{cm^2}$ [Tomek et al., 2019]. To approach the simulator's steady state, a series of 1000 stimuli were applied for each new parameter set starting from the initial states specified in the CellML file (when 1000 additional stimuli were applied, the maximum intracellular $[Ca^{2+}]$, $[Cl^-]$, and $[Na^+]$ changed by 1.5%, 0.15%, and 1.7%, respectively). Note, that the simulations can also be performed using open-source software such as Myokit [Clerx et al., 2016], OpenCOR [Garny and Hunter, 2015] and openCARP [Plank et al., 2021].

2.2 Emulator

Here we present the architecture and capabilities of the implemented and trained emulator. Details on the practical implementation and a link to the public code can be found in Section 5.

2.2.1 Data

We generated 3 data sets in the study to train, validate and test the emulator performance.

Training/validation data (#1) The first data set is the supervised training data set, containing pairs of maximum conductance samples $\mathbf{x}$ and the corresponding AP $\hat{V}_m(t)$ that was obtained from the simulator. Sobol' sequences were used to generate 40,000 maximum conductance samples, containing 20,000 maximum conductance samples between 0% and 200% of the original values and 20,000 maximum conductance samples between 50% and 150% of the original values. The first covers a range that was considered plausible in terms of physiology and pathology [Britton et al., 2017 [↗](#), Tomek et al., 2019 [↗](#)], and in terms of pharmacology (where full block is plausible). The latter covers a range that was considered particularly relevant in line with experimental calibration results presented in [Tomek et al., 2019 [↗](#)] and patch clamp measurements of channels that were exposed to 30 clinical drugs (blocks in up to the four-fold of the maximum free therapeutic concentration were analyzed in agreement with the CiPA paradigm [Li et al., 2019 [↗](#)] [Crumb et al., 2016 [↗](#)]). The SALib-Sensitivity Analysis Library [Herman and Usher, 2017 [↗](#)] was used in the entire study for the generation of samples based on Sobol' sequences. For each maximum conductance sample, simulations were performed to obtain the corresponding 40,000 APs. APs with a transmembrane potential difference of more than 10% of the amplitude between $t = 0$ and 1000 ms (indicative of an AP that is far away from full repolarization) were excluded. This, however, applied to only 116 APs.

Starting from the original APs in data set #1, the data were first extended by 10 ms from $t \in [0, 1000]$ ms to $t \in [-10, 1000]$ ms to enable some extrapolation of V_m and hence a better alignment of the depolarization; for $t \in [-10, 0]$ ms the initial resting membrane potential $V_m(0)$ was held constant. Then, the data were non-uniformly resampled from the original uniformly simulated APs, to emphasize the depolarization slope with a high accuracy while lowering the number of repolarization samples. For this purpose, we resampled the APs to 4 kHz for $t \in [-20, -5]$ ms (resting phase) and 10 kHz for $t \in [-5, 20]$ ms (depolarization phase) to 4 kHz. The repolarization phase ($t \in [20, 1000]$ ms) was also resampled to 4 kHz.

From the initial training data set, 20% were randomly chosen to be used for validation leaving 31908 pairs of maximum conductances and corresponding APs for training. Figure 2 shows processed APs that were used for training, validation and the APs excluded due to missing full repolarization as described above.

Synthetic test data (#2/#3) Two sets of synthetic data were created using the simulator and each of the sets consisted of control and drug data with pairs of maximum conductances and corresponding APs.

The control data were the same in both sets. They were created using an experimentally calibrated population of 100 synthetic cardiomyocytes [Britton et al., 2013 [↗](#), Muszkiewicz et al., 2016a [↗](#), Gemmell et al., 2016 [↗](#)]. and to this end, Sobol' sequences were used to generate samples of maximum conductances with values between 50% and 150% of the original values. Maximum conductance samples that produced APs without abnormalities (checked for the last two consecutive APs; see Appendix A [↗](#)), and with seven biomarker values (derived from the last AP; see Appendix A [↗](#)) in agreement with experimental ranges (Table 1 [↗](#)) were included in the population. Please note that the experimental ranges were not derived from the data set described in Section 2.2.1 [↗](#).

AP biomarker	Unit	Min	Max
RMP	mV	−95	−80
$dVmMax$	$mVms^{-1}$	100	1000
$Peak$	mV	10	55
APD_{40}	ms	85	320
APD_{50}	ms	110	350
APD_{90}	ms	180	440
Tri_{90-40}	ms	50	150

Table 1

AP biomarkers and their experimental ranges used to generate the population of synthetic cardiomyocytes. These were adopted from [Passini et al., 2017]. Experimental data were collected at 37 °C in small right ventricular trabeculae and papillary tissue preparations obtained from healthy human hearts during pacing with a cycle length of 1000 ms [Britton et al., 2017, O’Hara et al., 2011].

ID	1	2	3	4	5	6	7	8	9	10
G_{Kr}	0.05	0.06	0.07	0.08	0.09	0.10	0.11	0.12	0.13	0.14
P_{Ca}	1.20	1.22	1.24	1.26	1.28	1.30	1.32	1.34	1.36	1.38

Table 2

Pharmacological parameter samples (synthetic drugs) with scaling factors for G_{Kr} and P_{Ca} to generate the drug data of data set #3.

Data set #2: The motivation for creating data set #2 was to evaluate the emulator on data of normal APs. Drug data were created using 100 synthetic drugs represented by a set of pharmacological parameters. Each synthetic drug was built to have four different targets, with all channels, pumps, and exchangers related to the emulator inputs considered as potential targets. To this end, 100 samples of four pharmacological parameters, each with values between 0.5 (50% block) and 1.5 (50% enhancement) were randomly generated. The synthetic drugs were applied to the entire synthetic cardiomyocyte population by scaling each of the relevant control maximum conductances with the corresponding pharmacological parameter. The samples that produced APs without abnormalities (checked for the last two consecutive APs; see Appendix A) were included in the data set. No sample was excluded and thus, the data set consisted of 100 control data pairs and 10,000 drug data pairs.

Data set #3: The motivation for creating data set #3 was to test the emulator on data of abnormal APs showing the repolarization abnormality EAD. This is considered a particularly relevant AP abnormality in pharmacological studies because of their role in the genesis of drug-induced ventricular arrhythmia’s [Weiss et al., 2010]. Drug data were created using ten synthetic drugs with the hERG channel and the Cav1.2 channel as targets. To this end, ten samples with pharmacological parameters for G_{Kr} and P_{Ca} (Table 2) were generated and the synthetic drugs were applied to the entire synthetic cardiomyocyte population by scaling G_{Kr} and P_{Ca} with the corresponding pharmacological parameter. Of the 1000 APs simulated, we discarded APs with a

transmembrane potential difference of more than 10% of the amplitude between $t = 0$ and 1000 ms (checked for the last AP), indicative of an AP that is far away from fully repolarizing within 1000 ms. This left us with 950 APs, 171 of which exhibit EAD (see [Appendix C](#)).

Experimental data (#4) In the experimental data set, APs were recorded in small right ventricular trabeculae and papillary tissue preparations that were obtained from healthy human hearts [Orvos et al., 2019]. The conventional microelectrode technique was used for recording at a temperature of 37 °C. Stimulation of APs was done at a pacing cycle length of 1000 ms using a pair of platinum electrodes that provided rectangular current pulses of 2 ms duration. To allow the preparations to equilibrate, stimuli were delivered for at least 60 min before the measurements started. Measurements were performed under control conditions and after administration of five channel-blocking drugs at one concentration in multiple preparations. Drugs were cisapride (30 nM), dofetilide (10 nM), sotalol (30 μM), terfenadine (1 μM), and verapamil (300 nM). The last ten consecutive APs of each measurement were analyzed to quantify the beat-to-beat variability. Overall, the beat-to-beat variability was found to be small (standard deviation in all APs below 7 mV before the peak due to time alignment mismatch and below 2 mV after the peak) and thus, the last ten consecutive APs of each measurement were averaged. In most of the preparations, the standard deviation between beats did not vary over time and thus, no temporal correlation of noise was assumed. Averaging also reduced the noise level. The experimental data set contained one pair of averaged control and drugged AP per preparation per drug. Pairs were excluded if the biomarker values (see [Appendix A](#)) of the control or the drugged AP were not in the range found in the training data (see [Section 2.2.3](#)). This applied to seven pairs and the final data set contained three pairs for cisapride, dofetilide, sotalol, and terfenadine, and one pair for verapamil. All measurements were performed at the University of Szeged, Hungary, and conformed to the principles of the Declaration of Helsinki. Protocols were approved by the National Scientific and Research Ethical Review Boards (4991-0/2010-1018EKU [339/PI/010]).

An overview of all utilized data is given in [Table 3](#).

ID	Description	Usage	Origin	Samples
#1	Training/validation data	Training and validating the emulator, choosing the best architecture (Section 2.2.2)	Simulation	39,884
#2	Synthetic drug data, normal APs	Testing forward and inverse performance for normal APs (Section 3.1.1 and Section 3.1.2)	Simulation	10 ⁴
#3	Synthetic drug data, including EAD APs	Testing forward performance of abnormal (EAD) APs (Section 3.1.1)	Simulation	950
#4	Experimental cardiomyocytes	Testing and comparing the inverse performance with data published by the CiPA initiative [Li et al., 2017, Chang et al., 2017] (Section 3.1.3)	[Orvos et al., 2019]	26

Table 3

Summary of the data used in this study, along with their usage and the number of valid samples. Note that each AP is counted individually, also in cases of control/drug pairs.

2.2.2 Architecture

The emulator takes maximum conductances of channels, pumps, and exchangers as inputs and computes the corresponding AP ($V_m(t)$) after the last stimulus as output. It was trained to represent human ventricular cardiomyocytes under control and drugged conditions and the inputs were selected based on two assumptions: 1) The kinetics of channels are preserved, while the number of channels, pumps, and exchangers vary due to different expression levels [Syed et al., 2005 [↗](#), Groenendaal et al., 2015 [↗](#), Krogh-Madsen et al., 2016 [↗](#)]. These numbers are captured in the simulator by the maximum conductance parameters (or permeability parameters but maximum conductance is used as general term here for the sake of simplicity) that determine the respective current densities; 2) Channels, pumps, and exchangers are potential (intended and unintended) drug targets and the interaction between drugs and their targets can be described by a scaling of the related maximum conductances [Brennan et al., 2009 [↗](#)]. The corresponding scaling factors are pharmacological parameters. These assumptions allowed us to focus on maximum conductances and we considered those as inputs which either the AP is sensitive to (G_{Na^+} , G_{NaL} , P_{Ca} , G_{To} , G_{Kr} , G_{K1} , G_{NCX} , P_{NaK}) or which are related to common drug targets (G_{Na^+} , G_{NaL} , P_{Ca} , G_{To} , G_{Kr} , G_{KS} , G_{K1} ; [Crumb et al., 2016 [↗](#)]) leading to the following selection: G_{Na^+} , G_{NaL} , P_{Ca} , G_{To} , G_{Kr} , G_{KS} , G_{K1} , G_{NCX} , and P_{NaK} . AP sensitivity was quantified using a global sensitivity analysis (GSA; see Appendix B [↗](#)) and the inclusion threshold was a total-effect Sobol' sensitivity index (ST) above 0.1 with respect to any of the considered biomarkers (see Appendix A [↗](#)).

Several emulator architectures were tried on the training and validation data sets and the final choice was hand-picked as a good trade-off between high accuracy on the validation set (#1) and low computational runtime cost. We decided to utilize a two-stage emulator architecture: First, the maximum conductances $\mathbf{x}$ – normalized to the range $\mathbf{x}_i \in [-0.5, 0.5]$ – are encoded using a first NN (Θ_1) into a latent representation ϑ . Second, this intermediate representation parameterizes a function $f_\vartheta: \mathbb{R} \rightarrow \mathbb{R}$ defined by a second NN (Θ_2) that can be continuously evaluated to receive the emulated AP at time t . To help the second NN in computing the fast depolarization, a simple depolarization term ($\tanh$) is added to f_ϑ . The parameters of this depolarization function are slope (d_1), offset (d_2), and amplitude (d_3), and are created by encoding the parameters through the first network, similar to the latent code. The AP approximated by the emulator is thus defined by

$$V_m(t) := f_\vartheta(t) + \tanh(d_1^2(t - d_2)) + d_3^2. \quad (1)$$

A schematic drawing of the emulator architecture is provided in Figure 3 [↗](#). Splitting the network into two parts — one for encoding the parameters into a latent space and a second one for evaluating f_ϑ — allowed us to give the emulator enough complexity without markedly increasing the computational cost: in most cases, it is desirable to compute the whole AP in the entire range, e.g. [0, 1000] ms, and not only at a single time step. Creating a single network that computes the mapping from maximum conductances to the transmembrane potential at a single time step (compare Figure 3 [↗](#)) is either orders of magnitudes slower than encoding the parameters into a latent vector (only done once per AP) or would require to reduce the complexity of the network, which led to inaccurate emulations in the validation. The additional depolarization term was introduced to address the difficulty of fitting the depolarization phase during training and decreased the required training time substantially. Note that the mapping from maximum conductances $\mathbf{x}$ to depolarization parameters $\{d_1, d_2, d_3\}$ is also learned through Θ_1 .

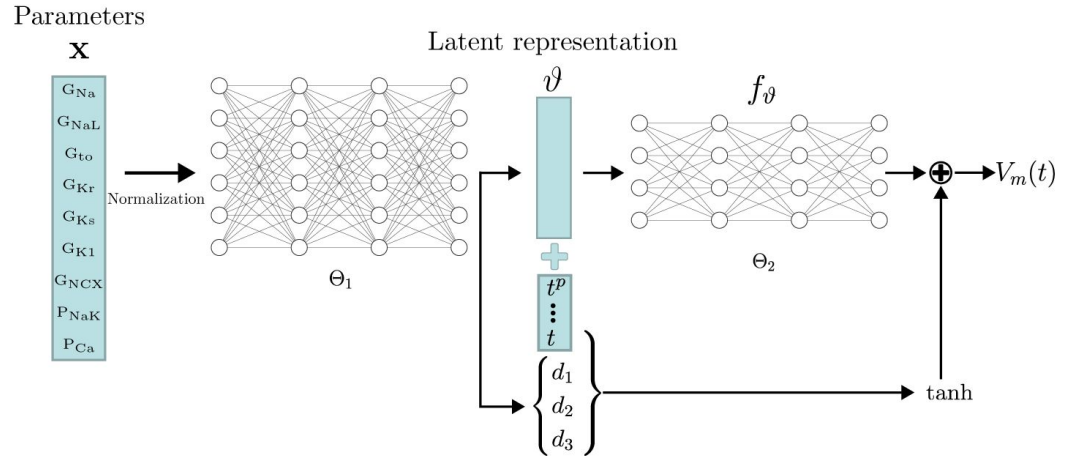


Figure 3

Conceptual architecture of the neural network emulator. The maximum conductances $\mathbf{x}$ are encoded into depolarization parameters d_i and a latent space representation ϑ that uniquely defines the time series functional $f_\vartheta : \mathbb{R} \rightarrow \mathbb{R}$. The time is normalized and encoded in polynomials up to degree 8 (t^p for $p \in \{1, 2, \dots, 8\}$), before being appended to the latent code ϑ . f_ϑ is then used in conjunction with the depolarization helper $\tanh$ to approximate the AP $V_m(t)$.

The exact architecture employed – chosen by a cross-validation approach (see [Section 2.2.4](#)) – comprised a first network (Θ_1) of four fully connected layers of 256 neurons, each to encode the parameters into the latent vector $\vartheta \in \mathbb{R}^{256}$. This first network additionally generates the parameterization for the depolarization model $\{d_1, d_2, d_3\}$. The second network, computing the APs from the latent representation (Θ_2), consisted of four fully connected layers of 64 neurons each. Exponential linear unit (ELU) activation functions [Clevert et al., 2016] were used for all layers, except for the final non-linear layer, which was modeled using a $\tanh$ activation function followed by a $[-1, 1]^{64} \rightarrow \mathbb{R}$.

2.2.3 Training

Although different regularization schemes such as variational losses (e.g. $\frac{dV_m}{dt}$) were tried, the wealth of training data allowed us to define the training loss purely in terms of mean-squared-error (MSE)

$$\mathcal{L}(\Theta_1, \Theta_2) = \frac{1}{2|\mathbb{T}|} \sum_{(\mathbf{x}, \hat{V}_m) \in \bar{X} \subset X} \sum_{t \in \mathbb{T}} \left(f_{\vartheta_{\Theta_1}(\mathbf{x}), \Theta_2}(t) - \hat{V}_m(t) \right)^2 dt, \quad (2)$$

where $f_{\vartheta_{\Theta_1}(\mathbf{x}), \Theta_2}$ describes the output of the emulator using the current NN weights Θ_1/Θ_2 , $\bar{X}$ refers to the current training batch and X is the training data set containing both target APs $\hat{V}_m$ and corresponding maximum conductances $\mathbf{x}$. For the training, increasing batch sizes ($|\bar{X}|$ 1250 to 1800) were used, both in terms of the entire AP and considering only subsets over time $T < T$ from $\frac{1}{16}|\mathbb{T}|$ to $|\mathbb{T}|$, resulting in a reduction of training time needed. The neural network was trained for a total number of 5000 epochs using the first-order gradient-based algorithm ADAM [Kingma and Ba, 2017]. The training time was approximately 4 h on the GPU specified in [Section 2.3.1](#).

2.2.4 Validation

A cross-validation approach was used to quantify and compare the performance of various emulator architectures. The validation was based on 20% of the initial training data set (7976 pairs of maximum conductances and corresponding APs, see [Section 2.2.3](#)). For each pair, the emulated AP V_m was compared against the simulated AP $\hat{V}_m$, given the same maximum conductances. The mismatch was quantified by the root-mean-squared error (RMSE) defined as

$$\text{RMSE}(V_m, \hat{V}_m) := \sqrt{\frac{1}{|T|} \int_T (V_m(t) - \hat{V}_m(t))^2 dt}.$$

The mismatch was also quantified in terms of AP biomarkers $\mathbf{b} \in \mathbb{R}^N$ (see [Appendix A](#)) and normalized maximum conductances $\mathbf{x}$ (see [Section 2.2.5](#)) in which case the RMSE was defined as

$$\text{RMSE}(\mathbf{b}, \tilde{\mathbf{b}}) := \sqrt{\frac{1}{N} \|\mathbf{b} - \tilde{\mathbf{b}}\|_2^2} \quad \text{and} \quad \text{RMSE}(\mathbf{x}, \tilde{\mathbf{x}}) := \sqrt{\frac{1}{N} \|\mathbf{x} - \tilde{\mathbf{x}}\|_2^2},$$

for N samples.

2.2.5 Time series fitting and estimation of maximum conductances and pharmacological parameters

Time series fitting is the basis for solving the inverse problem. To fit a given AP $\hat{V}_m$, defined on a subset of the trained domain $\hat{T} \subset T$, the first step was to choose a trial set of maximum conductances $\mathbf{x}_0$. Then, for the given trial set, the corresponding AP was emulated and the trial set of maximum conductances was iteratively updated by solving the following minimization problem:

$$\min_{\mathbf{x}, t_0} \frac{1}{2|\hat{T}|} \sum_{t \in \hat{T}} (V_m(\mathbf{x}, t - t_0) - \hat{V}_m(t))^2 dt + \frac{\lambda_{\mathbf{x}_0}}{2} \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 + \delta_{[-0.5, 0.5]}(\mathbf{x}), \quad (3)$$

where $V_m(\mathbf{x}, t) := f_{\vartheta_{\Theta_1}(\mathbf{x}), \Theta_2}(t)$ is a shorthand for the emulator approximation function and t_0 is a temporal offset parameter helping in fitting the exact depolarization timing.

Here, $\delta_{[-0.5, 0.5]}(\mathbf{x})$ is the element-wise indicator function on the normalized feasible parameter space $[-0.5, 0.5]$. The minimization was done using ADAM [Kingma and Ba, 2017] combined with a projection on the feasible space.

To estimate maximum conductances for a given control AP, the control AP was fitted using the original maximum conductance values as initial guesses and priors: $\mathbf{x}_0 = \hat{\mathbf{x}} = \mathbf{0}$. To estimate maximum conductances for a given drugged AP, the drugged AP was fitted using the maximum conductances estimated for the respective control AP as initial guesses and priors: $\mathbf{x}_0 = \hat{\mathbf{x}} = \mathbf{x}^c$. The pharmacological parameters (scaling factors of control maximum conductances) were computed as element-wise ratios between the drugged and control maximum conductances $(\mathbf{s})_i := \frac{(\mathbf{G}^d)_i}{(\mathbf{G}^c)_i}$ but here, $\mathbf{G}^d$ and $\mathbf{G}^c$ are the non-normalized maximum conductances, where $(\mathbf{G}^c)_i > 0$.

Since multiple sets of maximum conductances produced similarly good fits of the given AP, the parameter $\lambda_{\mathbf{x}_0}$ was introduced which minimizes the difference between original and control maximum conductances and between control and drugged maximum conductances, respectively. The value was chosen to be $\lambda_{\mathbf{x}_0} = 10$ with respect to the highest accuracy found for the synthetic data set #2 that was generated for the evaluation (see [Section 2.2.1](#)).

2.3 Evaluation

The evaluation was performed for forward and inverse problems in pharmacological studies on synthetic and experimental data.

The raw experimental data were obtained without filtering but some filtering was applied before interfacing the data with the emulator. The APs contained a stimulus artifact between 0 and 1.5 ms that was filtered as follows. For each of the last ten consecutive APs, the transmembrane potential V_m closest to the end of the recorded time series was defined as the resting transmembrane potential RMP and $V_m(t) = RMP$ was set for $t \in [0, 1.75]$ ms. Then, the APs were resampled at 100 kHz and a low-pass filtering was performed with a second-order butterworth filter (cutoff at 2.5 kHz) to reduce the high-frequency noise of the signal. Finally, the filtered APs were averaged and the averaged AP was again resampled at 1 kHz for $t \in [15, 1000]$ ms (repolarization) and 100 kHz in $t \in [0, 15]$ ms (depolarization). An example comparison of a raw and a filtered averaged AP is given in [Figure 4](#).

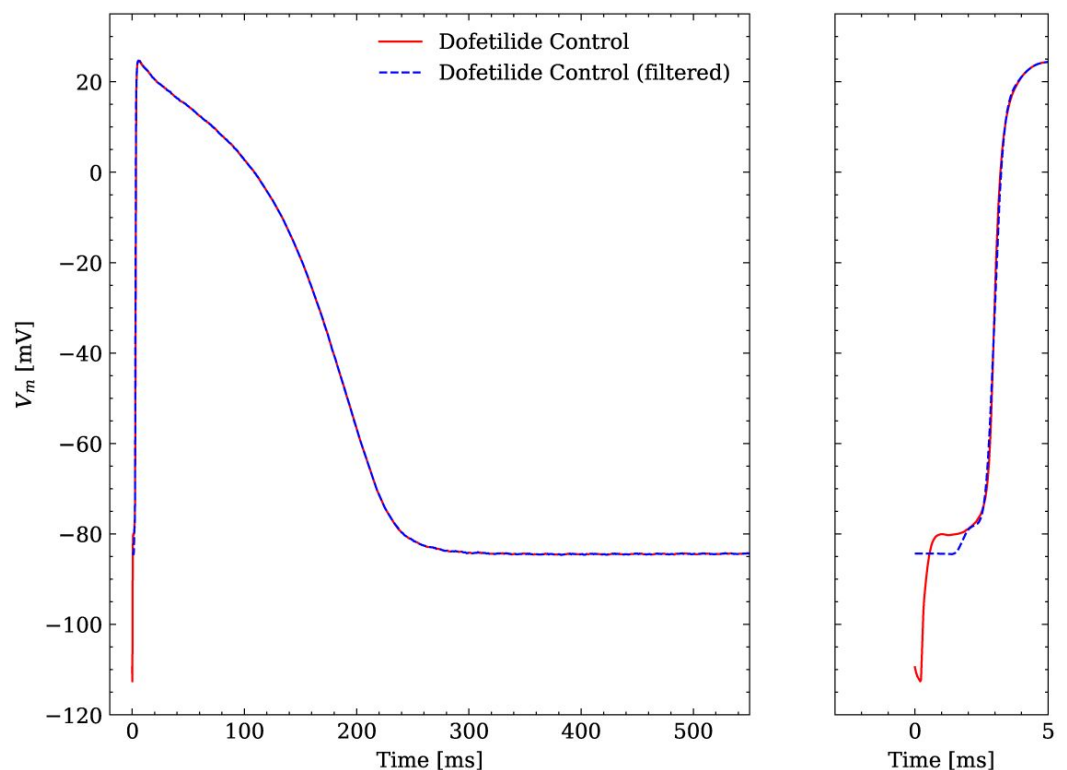


Figure 4

Comparison of an averaged raw and an averaged filtered experimental AP. One dofetilide control AP is shown as example.

2.3.1 Computational performance

The simulation of a single AP (see [Section 2.1](#)) sampled at a resolution of 20 kHz took 293 s on one core of a AMD Ryzen Threadripper 2990WX (clock rate: 3.0 GHz) in CARPentry. Adaptive timestep solver of variable order, such as implemented in Myokit [\[Clerx et al., 2016\]](#), can significantly lower the simulation time (30 s for our setup) by using small step sizes close to the depolarization (phase 0) and increasing the time step in all other phases. The emulation of a

steady state AP sampled at a resolution of 20 kHz for $t \in [-10, 1000]$ ms took 18.7 ms on a AMD Ryzen 7 3800X (clock rate: 3.9 GHz) and 1.2 ms on a Nvidia A100 (Nvidia Corporation, USA), including synchronization and data copy overhead between CPU and GPU.

The amount of required beats to reach the steady state of the cell in the simulator has a major impact on the runtime and is not known a-priori. On the other hand, both simulator and emulator runtime linearly depends on the time resolution, but since the output of the emulator is learned, the time resolution can be chosen at arbitrarily without affecting the AP at the sampled times. This makes direct performance comparisons between the two methodologies difficult. To still be able to quantify the speed-up, we ran Myokit using 100 beats to reach steady state, taking 3.2 s of simulation time. In this scenario, we witnessed a speed-up of 171 and $2 \cdot 10^3$ of our emulator on CPU and GPU respectively (again including synchronization and data copy overhead between CPU and GPU in the latter case). Note that both methods are similarly expected to have a linear parallelization speedup across multiple cells.

For the inverse problem, we parallelized the problem for multiple cells and keep the problem on the GPU to minimize the overhead, achieving emulations (including backpropagation) that run in 120 μ s per AP at an average temporal resolution of 2 kHz. We consider this the peak performance which will be necessary for the inverse problem in [Section 3.1.2](#).

2.3.2 Forward problem

The emulator evaluation for the forward problem, i.e. to find the drugged AP for given pharmacological parameters, was only performed on synthetic data since maximum conductances were not available experimentally. The maximum conductances of data sets #2 and #3 were used to consider data with normal APs and with abnormal APs exhibiting EADs. Pharmacological parameters are not inputs of the emulator but drugged maximum conductances that were computed as control maximum conductances scaled by the given pharmacological parameters (see [Section 2.2.1](#)). These were used to emulate drugged APs. The RMSE was used to quantify the mismatch between the emulated and the ground truth AP.

2.3.3 Inverse problem

The emulator evaluation for the inverse problem, i.e. to find the pharmacological parameters for given control and drugged APs (through optimization), was performed on both synthetic and experimental data. When using synthetic data, the data set #2 was used including data with normal APs. First, control and drugged maximum conductances were estimated based on control and drugged APs and then, pharmacological parameters were computed as ratios of drugged and control conductances (see [Section 2.2.5](#)). The mismatch between estimated and ground truth maximum conductances were quantified using the error that is defined through

$$\mathbf{e}_n := (\mathbf{x} - \hat{\mathbf{x}}), \quad (4)$$

where $\mathbf{x}$ and $\hat{\mathbf{x}}$ are the normalized estimated and ground truth maximum conductances (see [Section 2.2.2](#)). Similarly, we computed the mismatch between estimated and ground truth scaling factor vectors ($\mathbf{s}$ and $\hat{\mathbf{s}}$ respectively) as

$$\mathbf{e}_s := (\mathbf{s} - \hat{\mathbf{s}}). \quad (5)$$

When using experimental data, maximum conductances and pharmacological parameters were estimated in the same way but due to a lack of experimental maximum conductances, the mismatch between estimated and ground truth values could not be quantified. Instead, the estimated pharmacological parameters were compared with distributions computed from data published within the CiPA initiative [Li et al., 2017, Chang et al., 2017] (CiPA distributions).

The data set [2](#) includes 2000 IC_{50} values and Hill coefficients for each drug and for up to seven targets (I_{Na} , I_{NaL} , I_{CaL} , I_{to} , I_{Kr} , I_{KS} , I_{K1}). The pore-block model [Brennan et al., 2009] was used to obtain the corresponding scaling factors.

3 Results

3.1 Evaluation

3.1.1 Forward problem

The emulator evaluation for the forward problem was only done on synthetic data and both data sets #2 and #3 (see Section 2.2.1) were used to analyze the solution accuracy for normal and abnormal APs exhibiting EADs.

The data set #2 was used first and [Figure 5](#) illustrates the distribution of RMSEs between emulated and ground truth drugged APs. In total 10^4 APs were emulated in 0.6 s. The average RMSE over all APs was 0.47 mV and only for a few APs the RMSE was >1 mV with 1.5 mV being the maximum. Largest mismatches were located in the phases 0 and 3 of the AP. While the mismatches in phase 3 were simply a result of imperfect emulation, the mismatches in phase 0 were a result of the difficulty in matching the depolarization time exactly.

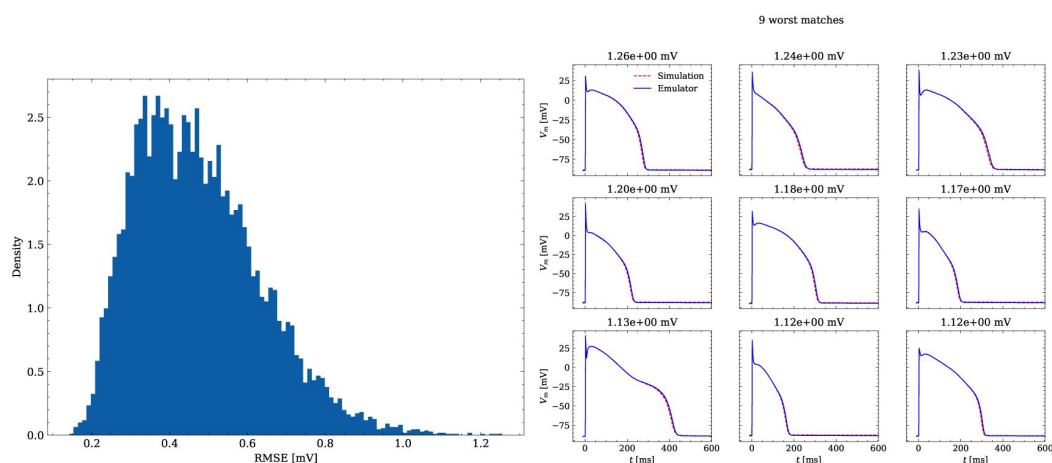


Figure 5

Analysis of solution accuracy of the forward problem on synthetic data including normal APs (drug data of data set #2). Left: histogram of RMSEs for the APs, right: APs with the largest RMSEs. The RMSE is given above each subplot.

[Figure 6](#) shows the distribution of biomarker mismatches between emulated and ground truth drugged APs. The low RMSEs between the APs translated into low RMSEs between the AP biomarkers. Likewise, the difficulty in exactly matching the depolarization time leads to elevated errors and more outliers in the biomarkers influenced by the depolarization phase (TP and dV_{mMax}).

Figure 6

Analysis of solution accuracy of the forward problem on synthetic data including normal APs (drug data of data set #2) with respect to AP biomarkers. Histograms of mismatches for each biomarker are shown and the RMSE is given in the upper left corner. The number in the right upper corner denotes the number of outliers of the 10,000 samples which lie outside the shown ranges.

The data set #3 was used second and **Appendix C** shows all emulated APs, both containing the EAD and non-EAD cases. The emulation of all 950 APs took 0.76 s on the GPU specified in **Section 2.2.3**. We show the emulation of all maximum conductances and the classification of the emulation. The comparison with the actual EAD classification (based on the criterion outlined in **Appendix A**) results in true-positive (EAD both in the simulation and emulation), false-negative (EAD in the simulation, but not in the emulation), false-positive (EAD in the emulation, but not in the simulation) and true-negative (no EAD both in the emulation and simulation). The emulations achieved 72.5% sensitivity (EAD cases correctly classified) and 94.9% specificity (non-EAD cases correctly classified), with an overall accuracy of 90.8% (total samples correctly classified). A substantial amount of wrongly classified APs showcase a notable proximity to the threshold of manifesting EADs. **Figure 7** illustrates the distribution of RMSEs in the EAD APs between emulated and ground truth drugged APs. The average RMSE over all EAD APs was 14.5 mV with 37.1 mV being the maximum. Largest mismatches were located in phase 3 of the AP, in particular in emulated APs that did not fully repolarize.

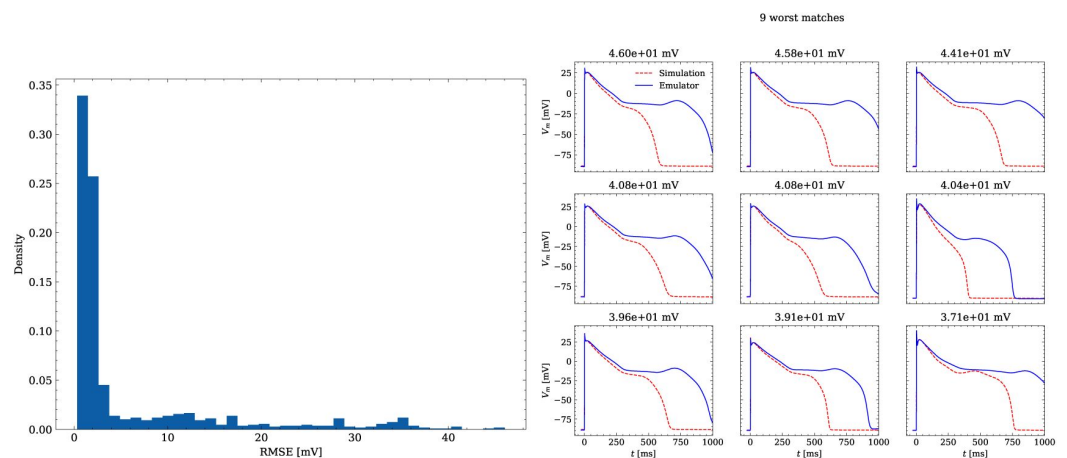


Figure 7

Analysis of solution accuracy of the forward problem on synthetic data including abnormal APs exhibiting EADs (subset of data set #3). Left: histogram of RMSEs for the APs, right: APs with the largest RMSEs. 124 of the 171 emulated APs exhibit the expected EADs (based on the criterion outlined in **Appendix A**). The RMSE is given above each subplot. All emulated APs are shown in **Appendix C**.

3.1.2 Inverse problem based on synthetic data

The emulator evaluation for the inverse problem was first done using synthetic data (data set #2, see **Section 2.2.1**). For this, we minimized (3) by using ADAM [Kingma and Ba, 2017] with no batching for 10^4 iterations for all 100 cardiomyocytes times 100 drugs (i.e. 10^4 APs), resulting in a total of 10^8 emulations, taking approximately

3.5 h on the GPU specified in [Section 2.2.3](#). Control and drugged APs could be fitted with an average RMSE of 0.8 mV. Largest mismatches were located in phase 0 and 3 of the AP for the reasons given above (see [Section 3.1.1](#)). **Figure 8** shows the distribution of the errors between the estimated and the ground truth maximum conductances and pharmacological parameters. For both the maximum conductances ($\text{RMSE} \leq 0.18$) and the related pharmacological parameters ($\text{RMSE} \leq 0.21$) the errors were closely distributed around zero. However, the RMSEs increased from the control maximum conductance over the drugged maximum conductance to the pharmacological parameters and there were distinctive differences among maximum conductances and related pharmacological parameters with the smallest for G_{Kr} and the largest for G_{Ks}

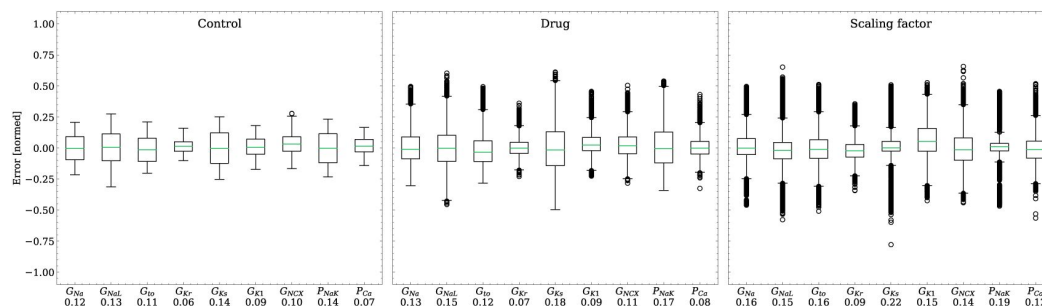


Figure 8

Analysis of solution accuracy of the inverse problem on synthetic data (data set #2). Left: boxplot of errors between normalized estimated and ground truth control maximum conductances, middle: boxplot of errors between normalized estimated and ground truth drugged maximum conductances, right: boxplot of errors between estimated and ground truth pharmacological parameters. Error definitions are given in (4) and (5). The RMSE over all data is given below each parameter.

3.1.3 Inverse problem based on experimental data

The emulator evaluation for the inverse problem was then done on experimental data ([Section 2.2.1](#)). Similar to the synthetic inverse problem, we optimized (3), this time for $5 \cdot 10^4$ and $2.5 \cdot 10^4$ epochs for control and drug APs for each drug sample. The total computational time spent on this task was 2.8 h on the GPU specified in [Section 2.2.3](#).

Figure 9 shows the fitted and the ground truth APs for all drugs. Control and drugged APs could be fitted with average RMSEs shown in [Table 4](#). The largest mismatch was located in phase 0 for most APs, which was the result from an imperfect matching of the exact depolarization timing.

Figure 9

Analysis of fit quality of the inverse problem on experimental data. Comparison of the fitted APs (solid lines) and the experimental APs (dashed lines) at control (red) and after drug administration (blue) for all drugs.

Drug	RMSE Control	RMSE Drug
Cisapride	1.53	2.00
Dofetilide	2.05	1.73
Sotalol	1.40	2.51
Terfenadine	1.22	1.08
Verapamil	1.93	2.21

Table 4

Average RMSE over control and drugged APs measured in all preparations per drug. All values in mV.

Figure 10 compares the estimated pharmacological parameters and the CiPA distributions. Pharmacological parameters that fell into the range spanned by $\mu \pm (0.15 + \sigma)$ of the CiPA distribution, where μ , σ are the distribution's mean and standard deviation respectively, were classified as successfully estimated while the others were classified as unsuccessfully estimated. In total, 50% of the pharmacological parameters could be estimated successfully and while all pharmacological parameters related to G_{KS} could be successfully estimated, unsuccessfully estimated parameters were found across all maximum conductances, in particular related to G_{K1} for which all pharmacological parameters could not be successfully estimated (**Table 5**).

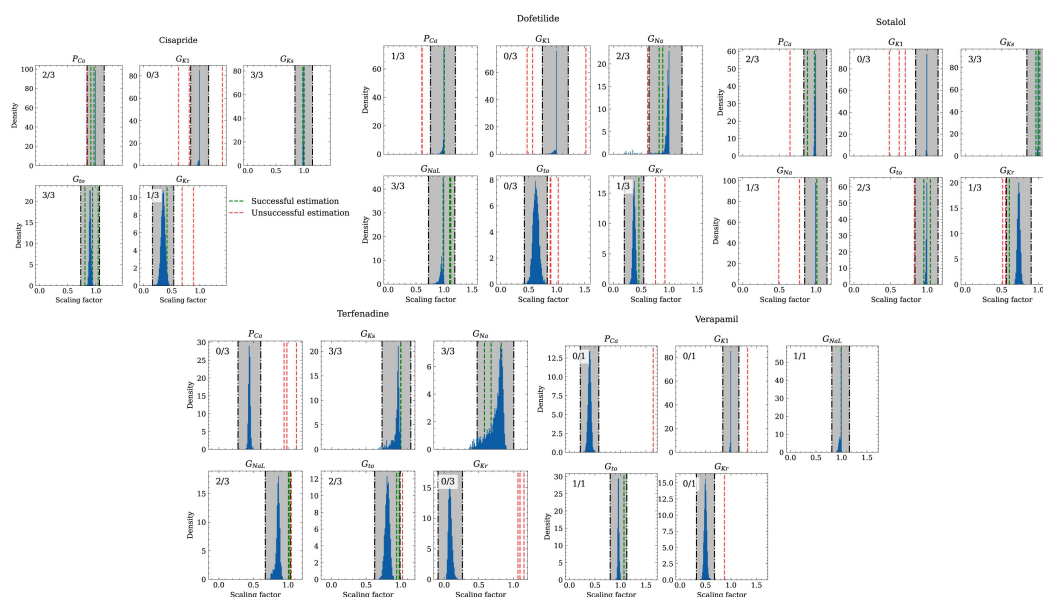


Figure 10

Analysis of solution accuracy of the inverse problem using experimental data. The histograms compare the estimated pharmacological parameters (dashed vertical lines) from data of multiple preparations with the CiPA distributions (blue; see **Section 2.3.3**). The black dash dotted vertical lines are the borders of the range (grey) that

was used to determine if the estimation of the given pharmacological parameter was successful. The range is spanned up by $\mu \pm (0.15 + \sigma)$ of the CiPA distribution. Successfully estimated parameters are shown as green lines and unsuccessfully estimated parameters are shown as red lines. The number in the upper left corner indicates for how many preparations the parameters could be successfully estimated in relation to the total number of preparations for the given drug.

	G _{Na}	G _{NaL}	G _{to}	G _{Kr}	G _{Ks}	G _{K1}	P _{Ca}	Total
Successful	6	6	8	3	9	0	5	37
Unsuccessful	3	1	5	10	0	10	8	37
Ratio	0.67	0.86	0.62	0.23	1.00	0.00	0.38	0.50
Drug	D/S/T	D/T/V	A	A	C/S/T	C/D/S/V	A	

Table 5

Pharmacological parameters related to maximum conductances that were considered successfully or unsuccessfully estimated across all preparations and drugs. For each channel, the drugs are stated for which respective data from the CiPA initiative were available. C, D, S, T, V, A mark cisapride, dofetilide, sotalol, terfenadine, verapamil, all drugs respectively.

4 Discussion

NN emulation of the human ventricular cardiomyocyte AP was introduced and the applicability in pharmacological studies was investigated.

4.1 Evaluation

The evaluated NN emulator showed highly accurate AP emulations for the forward problem on synthetic data. High accuracy was found in normal APs in data set #2 (average RMSE was 0.47 mV; **Figure 5**) and to a lesser extent also in abnormal APs exhibiting EADs: of the emulated EAD APs, 72.5% exhibited alignment with the abnormality, and the substantial majority of the remaining APs demonstrated pronounced proximity, while the average RMSE among the EAD APs was 14.5 mV (**Figure 7**). In comparison, the normal APs exhibiting no EAD in data set #3 could be reconstructed with an average RMSE of 4.02 mV and the overall accuracy of classifying EAD on emulated APs was 90.8%. Increasing the amount of training data within the relevant range could lead to further enhancements in accuracy for abnormal APs. Nevertheless, this observation demonstrates that the emulator is also capable of accounting for discontinuities of the response surface. This is particularly useful in pharmacological studies and a key advantage over existing emulation approaches [Chang et al., 2015, Johnstone et al., 2016, Coveney and Clayton, 2018, Ghosh et al., 2018, Rasmussen and Williams, 2019, Coveney et al., 2021].

The emulator was further evaluated for the inverse problem on synthetic and also on experimental data. Maximum conductances and related pharmacological parameters could be widely estimated with high accuracy on synthetic data ($\text{RMSE} \leq 0.18$ and ≤ 0.21 for all maximum conductances and pharmacological parameters, respectively; **Figure 8**).

The RMSEs increased from the control maximum conductance over the drugged maximum conductance to the pharmacological parameters which may be because the estimation of drugged maximum conductances depends on the control maximum conductances and the computation of pharmacological parameters depends on both control and drugged maximum conductances (see

Section 2.2.5 allowing that errors can propagate and amplify. Distinctive differences were observed among the maximum conductances and related pharmacological parameters and the largest RMSEs were found for G_{Ks} throughout. This can be attributed to various degrees of parameter identifiability [Sarkar and Sobie, 2010, Zaniboni et al., 2010, Groenendaal et al., 2015, Jæger et al., 2019, Whittaker et al., 2020] and the results agree with the GSA that indicates almost non-identifiability of G_{Ks} (Appendix B).

Larger inaccuracies were found in the inverse problem solutions on experimental data (**Figure 10**, **Table 4**). The first reason may be low parameter identifiability and we want to highlight inaccuracies in estimating the pharmacological parameters related to G_{Kr} , P_{Ca} , and G_{NaL} when the hERG channel was blocked in parallel to the Cav1.2 channel (verapamil) or in parallel to both the Cav1.2 and the Nav1.5-late channel (terfenadine). The hERG channel block (prolongation of the AP), and the Cav1.2 and Nav1.5-late channel block (shortening of the AP) are known to have opposite effects on the AP [Orvos et al., 2019]. At the given drug concentrations, these effects were apparently counterbalancing, which resulted in negligible changes of the AP (**Figure 9**). This situation made the estimation of pharmacological parameters very challenging and led to particularly large inaccuracies for terfenadine. The accurate estimations of the pharmacological parameters related to G_{Ks} are surprising at first in light of the almost non-identifiability. This was due to the combination of two factors: 1) different from synthetic data, the drugs at the given concentrations did not affect the corresponding KCNQ1-MinK channel and 2) the difference between control and drugged maximum conductance was weakly enforced to be minimal (see **Section 2.2.5**) which leads to almost no difference in non-identifiable parameters and hence, to a pharmacological parameter of one. The accuracy will likely be much lower in drugs that affect the KCNQ1-MinK channel.

The second and probably main reason for the inaccuracies may be the fact that the data were collected in small tissue preparations, whereas the emulator was trained on data generated by a simulator that represents single cardiomyocytes. APs in small tissue preparations are slightly different from those in single cardiomyocytes. Differences can arise from electrotonic coupling and the mixture of cells including fibroblasts that are able to modify the EP [Kohl and Gourdie, 2014, Mayer et al., 2017, Hall et al., 2021]. This can hamper the fitting of the APs and consequently, the estimation of the maximum conductances and pharmacological parameters.

4.2 Emulator

The presented NN emulator enables a massive speed-up compared to regular simulations and the evaluation for the forward problem on synthetic data showed also highly accurate AP emulations. Cardiomyocyte EP models are already very quick to evaluate in the scale of seconds (see **Section 2.3.1**), but the achieved runtime of emulations allows to solve time consuming simulation protocols markedly more efficient. One such scenario is the presented inverse maximum conductance estimation problem (see **Section 3.1.2** and **Section 3.1.3**), where for estimating maximum conductances of a single AP, we need to emulate the steady state AP at least several hundred times as part of an optimization procedure. Further applications include the probabilistic use of cardiomyocyte EP models with uncertainty quantification [Chang et al., 2017, Johnstone et al., 2016] where thousands of samples of parameters are potentially necessary to compute a distribution of the steady-state properties of subsequent APs, and the creation of cell populations [Muszkiewicz et al., 2016b, Gemmell et al., 2016, Britton et al., 2013]. In addition to the aforementioned strengths of the presented NN emulator, some further valuable features are worth mentioning that arise from the continuous nature of the emulator. First, the AP can be emulated and fitted at any desired resolution. Second, timing offsets, e.g. between stimuli in the data to be fitted and the training data, can be accounted for using t_0 in (3) without retraining. Last but not least, the transmembrane potential gradient $\frac{dV_m}{dt}$, recently highlighted in terms of proarrhythmic potential prediction [Jeong et al., 2022], can be continuously derived and is not dependent on the temporal discretization.

4.3 Limitations and future work

Some limitations have to be considered. First, the emulator has only maximum conductances as inputs. Although these explain much of the AP variability seen between cardiomyocytes [Britton et al., 2013 [↗](#), Muszkiewicz et al., 2016a [↗](#)], the inclusion of parameters related to the channel kinetics might enable a more detailed consideration of drug effects in pharmacological studies. In general, the question of complete parameter identifiability utilizing only APs remains an open challenge [Zaniboni et al., 2010 [↗](#)]. Channel kinetics determine the contribution of the corresponding current to the AP generation in different phases and can thus also modulate drug effects, but were however neglected, as the expansion of the input space might be unsuitable for solving the inverse problem when only AP data are used. Second, the interaction between drugs and their targets is solely captured through scaling of the related maximum conductance at control, which is mostly adequate but in fact an oversimplification [Brennan et al., 2009 [↗](#)]. The interaction can be dependent on time, voltage, and channel state, which requires the use of Markov models with many more parameters [Brennan et al., 2009 [↗](#), Li et al., 2017 [↗](#), Lei et al., 2023 [↗](#)]. Again, this expands the input space and might be unsuitable for solving the inverse problem when only AP data are used. Moreover, drugs that are applied over a longer period of time can also cause modifications of the maximum conductances through changes in gene expression [Shim et al., 2023 [↗](#)]. This requires attention to avoid misinterpretations of found blocking or enhancement effects, e.g. by estimating the control maximum conductances again after a washout procedure. Third, the inverse problem was only solved for AP data obtained from one single stimulation protocol. [Johnstone et al., 2016 [↗](#)] have shown that the usage of AP data obtained from various stimulation protocols can improve the parameter identifiability and thus, the accuracy of parameter estimates. To be able to use those data in the presented approach, the pacing cycle length must be included as additional input in the emulator and the emulator may be trained on more than the last AP of the pacing series. This would also allow to capture alternans. Last but not least, the number of drugs and concentrations considered in the inverse problem on experimental data poses a limitation. The ultimate goal is to have a tool that provides highly accurate solutions for drugs with different targets and concentrations. To this end, analyses must be extended by data obtained from a series of available and well characterized drugs. The data should be collected in single cardiomyocytes in order to minimize the discussed inaccuracies that stem from the use of tissue preparation data. Further simulation studies of the inverse problem on tissue slabs versus cardiomyocyte EP model could be integrated to assess the impact of differences in setups. This should be addressed in future work. Of note, the presented approach can also be straightforwardly applied to other transients, e.g. intracellular $[Ca^{2+}]$ or sarcomere length.

4.4 Conclusion

This paper introduced NN emulation of the human ventricular cardiomyocyte AP and tested its applicability in pharmacological studies. The computational cost of the NN emulator was compared to that of the simulator, revealing a massive speed-up of more than 10^3 . The accuracy of solving the forward problem on synthetic data was found to be high for normal APs and this hold mostly true for abnormal APs exhibiting EADs. This advantage distinguishes our novel approach from existing emulation methods. While larger inaccuracies were observed when utilizing experimental data – a limitation thoroughly discussed and particularly inherent to the fact that small tissue preparations were studied while the emulator was trained on single cardiomyocyte data – the accuracy of solving the inverse problem on synthetic data remained high. Collectively, these findings underscore the potential of NN emulators in improving the efficiency of future quantitative systems pharmacology studies.

5 Code availability

The trained emulator is available as a python package under the link https://github.com/thomgrand/cardiomyocyte_emulator [↗](#)³[↗](#). The trained emulator is provided as a Python package, heavily utilizing PyTorch [Paszke et al., 2019 [↗](#)] for the neural network execution, allowing it to be

executed on both CPUs and NVidia GPUs. Additionally, NumPy [Harris et al., 2020] interfaces are provided for easy interfacing with other libraries. Note that the provided repository is a re-implementation of the code-base used in this study and thus may deviate in performance or runtime.

6 Additional information

6.1 Funding

National Research, Development and Innovation Office Project, No. 142738 HUN-REN Hungarian Research Network

Norbert Jost András Varró

German Research Foundation, Walter Benjamin Fellowship, No. 468256475

Alexander Jung

ERA-NET co-fund action no. 680969 funded by the Austrian Science Fund (FWF), grant I 4652-B

Christoph M. Augustin

Wellcome Senior Research Fellowship, No. 212203/Z/18/Z

Gary R. Mirams

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

This research was funded in whole, or in part, by the Wellcome Trust [212203/Z/18/Z], the Austrian Science Fund (FWF) [I 4652-B] and the National Institutes of Health R01HL158667. For the purpose of open access, the author has applied a CC-BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

6.2 Author contributions

Thomas Grandits: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing

Christoph M. Augustin: Funding acquisition, Writing – review & editing

Gundolf Haase: Funding acquisition, Resources, Writing – review & editing

Norbert Jost: Funding acquisition, Resources, Writing – review & editing

Gary Mirams: Funding acquisition, Methodology, Writing – review & editing

Steven Niederer: Funding acquisition, Writing – review & editing

Gernot Plank: Funding acquisition, Resources, Writing – review & editing

László Virág: Resources, Writing – review & editing

András Varró: Funding acquisition, Resources, Writing – review & editing

Alexander Jung: Conceptualization, Data curation, Funding acquisition, Formal analysis, Investigation, Project administration, Validation, Visualization, Writing – original draft, Writing – review & editing

A AP biomarkers and abnormalities

The AP biomarkers used in this study were selected such that the key characteristics of the depolarization and the repolarization phase can be quantified. They include *RMP* (resting transmembrane potential measured just before stimulation), *dV mMax* (maximum transmembrane potential slope during the upstroke), *Peak* (peak transmembrane potential at the end of the upstroke), *TP* (time-to-peak from the stimulus until *Peak* is reached), *APD_x* (AP duration at $x \in \{30, 40, 50, 60, 70, 80, 90\}\%$ repolarization relative to the AP amplitude (*Peak* – *RMP*) measured from the instant of *dV mMax*), and *Tri_{90–40}* (triangulation defined as the difference between *APD₉₀* and *APD₄₀* [Britton et al., 2017]).

For the GSA (see Appendix B) and the creation of synthetic data (see Section 2.2.1), AP abnormalities were also considered which included depolarization abnormalities, repolarization abnormalities, and alternans. Depolarization abnormalities were defined as an upstroke peak below 0 mV and an AP that does not reach 0 mV before 100 ms after stimulation [Passini et al., 2017]. Repolarization abnormalities were defined as a transmembrane potential rate of rise of more than $\frac{\text{mV}}{\text{ms}}$ from 150 ms after the upstroke peak onwards (representative of early afterdepolarizations) and as a transmembrane potential that does not fall below –40 mV [Passini et al., 2017] (representative of repolarization failure). Alternans were defined as *APD₉₀* difference of more than 5 ms between two consecutive APs [Morotti et al., 2021].

B Global sensitivity analysis

A variance-based Sobol’ global sensitivity analysis (GSA) [Sobol’, 2001] was performed on the simulator to quantify the sensitivities of the maximum channel conductances (inputs) with respect to the AP biomarkers (outputs; see Appendix A). This informed the decision on which inputs to consider in the emulator. Furthermore, it was used for the interpretation of the solutions of the inverse problem since parameters that are insensitive with respect to the outputs indicate non-identifiability [Guillaume et al., 2019].

The maximum conductances used for building the model population in [Tomek et al., 2019] were considered and Saltelli’s sampling scheme [Saltelli, 2002] was applied with $N = 1024$ to generate 20,480 input samples with values between 50% and 150% of the original values. Simulations were performed for each input sample and biomarker values (see Appendix A) derived from the last AP were used for the analysis. However, data were excluded if not all biomarkers could be determined or abnormalities (see Appendix A) were detected in the last two consecutive APs. First-order (S1) and total-effect (ST) Sobol’ sensitivity indices were computed using the Saltelli method [Homma and Saltelli, 1996, Saltelli, 2002]. This requires outputs for each input sample and to take this into account, excluded outputs were assigned the mean values of included outputs. The SALib-Sensitivity Analysis Library [Herman and Usher, 2017] was used for the GSA.

The GSA could be performed on the data of all input samples as no data were excluded. S1 and ST were mostly very similar which indicates only little interactions among the maximum conductivities relative to the AP biomarkers (Figure 11). The only exception was *TP*. As was to be expected, the analysis underlines the predominant relative sensitivity of G_{Na} with respect to biomarkers of the depolarization phase, the predominant relative sensitivity of G_{NaL} , G_{Kr} , and

G_{NCX} with respect to biomarkers of the repolarization phase and the predominant relative sensitivity of G_{K1} , P_{NaK} to the resting transmembrane potential. However, G_{Ks} has a negligible relative sensitivity to all biomarkers. This indicates almost non-identifiability.

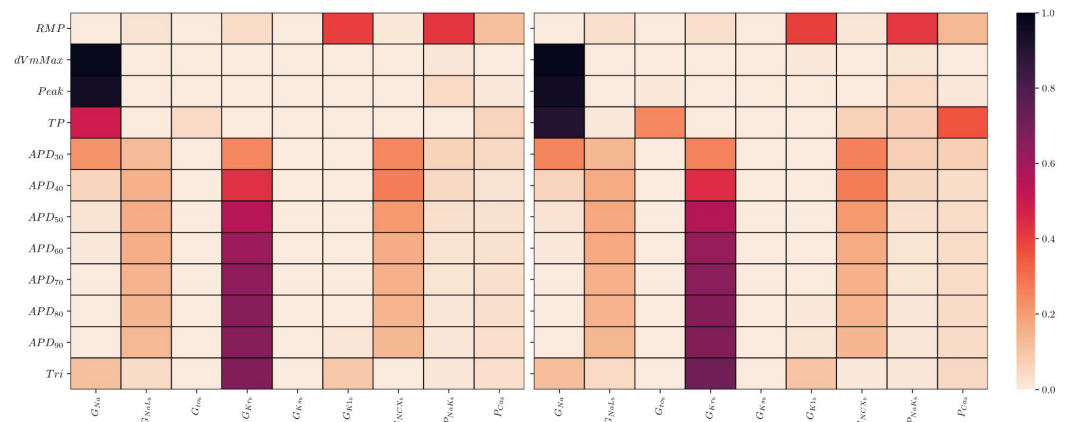


Figure 11

Global sensitivity analysis of the ToR-ORd simulator. Sobol' sensitivity indices are shown for each maximum conductance relative to each AP biomarker. Left: first-order (S1), right: total-effect (ST) Sobol' sensitivity coefficient.

C EAD classification

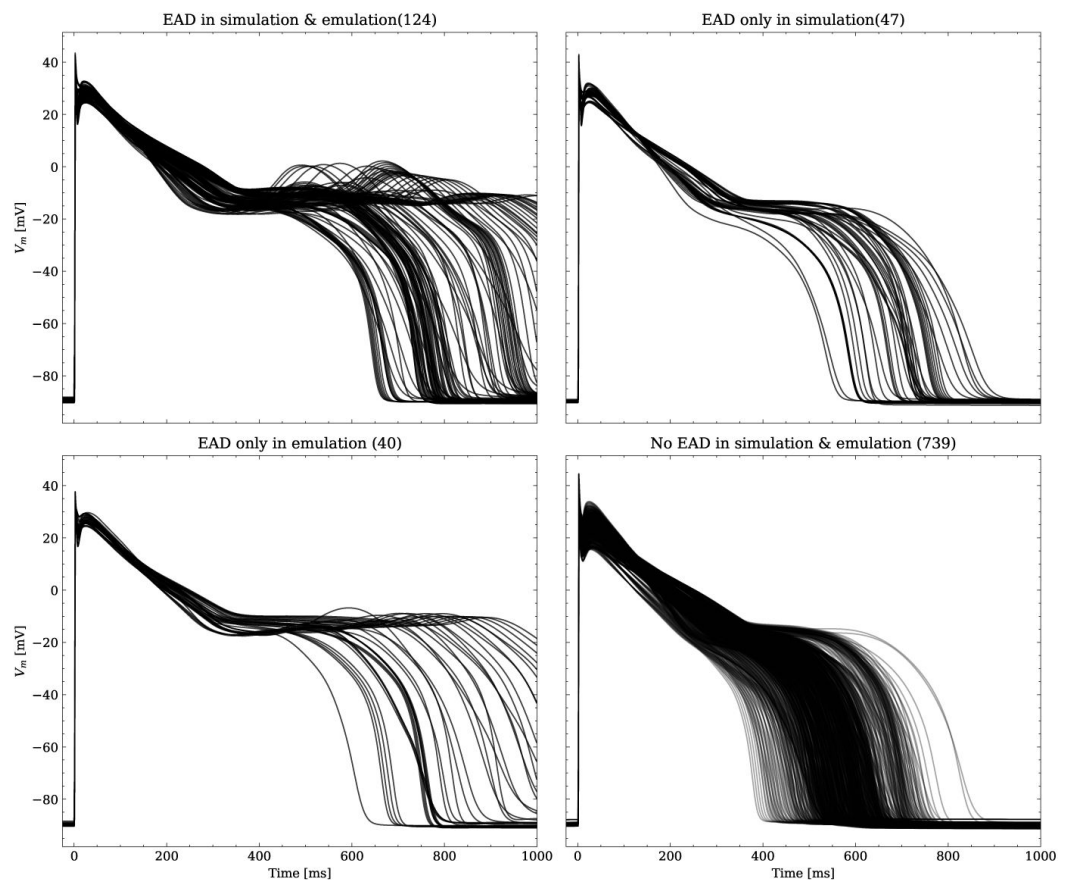


Figure 12

Emulated APs based on the pharmacological parameters of data set #3. See also [Section 3.1.1](#) and [Figure 7](#). From left to right and top to bottom, the plot shows the true positive, false negative, false positive and true negative samples. The number next to the title specifies the number of samples belonging to each category. The classification criterion is outlined in Appendix A.

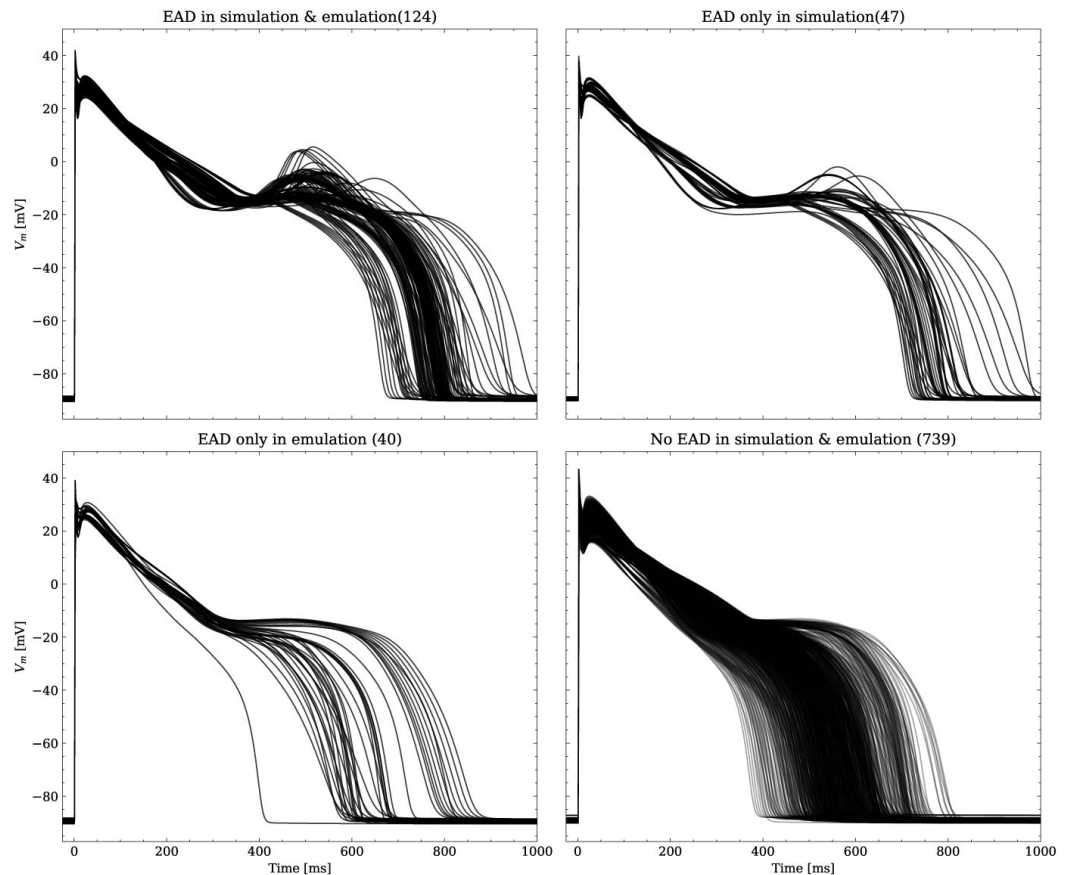


Figure 13

Same as [Figure 12](#), but showing the simulated APs.

References

- Amuzescu B., Airini R., Epureanu F. B., Mann S. A., Knott T., Radu B. M (2021) **Evolution of mathematical models of cardiomyocyte electrophysiology** *Mathematical Biosciences* **334**
- Bartolucci C., Passini E., Hyttinen J., Paci M., Severi S (2020) **Simulation of the effects of extracellular calcium changes leads to a novel computational model of human ventricular action potential with a revised calcium handling** *Frontiers in Physiology* **11**
- Blinova K. *et al.* (2018) **International multisite study of human-induced pluripotent stem cell-derived cardiomyocytes for drug proarrhythmic potential assessment** *Cell Reports* **24**:3582–3592
- Bottino D., Penland R. C., Stamps A., Traebert M., Dumotier B., Georgieva A., Helmlinger G., Lett G. S (2006) **Preclinical cardiac safety assessment of pharmaceutical compounds using an integrated systems-based computer model of the heart** *Progress in Biophysics and Molecular Biology* **90**:414–443
- Brennan T., Fink M., Rodriguez B (2009) **Multiscale modelling of drug-induced effects on cardiac electrophysiological activity** *European Journal of Pharmaceutical Sciences* **36**:62–77
- Britton O. J., Bueno-Orovio A., Ammel K. V., Lu H. R., Towart R., Gallacher D. J., Rodriguez B (2013) **Experimentally calibrated population of models predicts and explains intersubject variability in cardiac cellular electrophysiology** *Proceedings of the National Academy of Sciences* **110**
- Britton O. J., Bueno-Orovio A., Virág L., Varró A., Rodriguez B (2017) **The electrogenic Na^+/K^+ pump is a key determinant of repolarization abnormality susceptibility in human ventricular cardiomyocytes: A population-based simulation study** *Frontiers in Physiology* **8**
- Chang E. T. Y., Strong M., Clayton R. H (2015) **Bayesian sensitivity analysis of a cardiac cell model using a gaussian process emulator** *PLOS ONE* **10**
- Chang K. C. *et al.* (2017) **Uncertainty quantification reveals the importance of data variability and experimental design considerations for in silico proarrhythmia risk assessment** *Frontiers in Physiology* **8**
- Clerx M., Collins P., de Lange E., Volders P. G. A. (2016) **Myokit: A simple interface to cardiac cellular electrophysiology** *Progress in Biophysics and Molecular Biology* **120**:100–114
- Clevert D.-A., Unterthiner T., Hochreiter S. (2016) **Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)** *arXiv:1511.07289*
- Colatsky T., Fermini B., Gintant G., Pierson J. B., Sager P., Sekino Y., Strauss D. G., Stockbridge N (2016) **The comprehensive in vitro proarrhythmia assay (CiPA) initiative — update on progress** *Journal of Pharmacological and Toxicological Methods* **81**:15–20
- Corrado C., Avezù A., Lee A. W. C., Costa C. M., Roney C. H., Strocchi M., Bishop M., Niederer S. A (2020) **Using cardiac ionic cell models to interpret clinical data** *WIREs Mechanisms of Disease* **13**

- Coveney S., Clayton R. H (2018) **Fitting two human atrial cell models to experimental data using bayesian history matching** *Progress in Biophysics and Molecular Biology* **139**:43–58
- Coveney S., Corrado C., Oakley J. E., Wilkinson R. D., Niederer S. A., Clayton R. H (2021) **Bayesian calibration of electrophysiology models using restitution curve emulators** *Frontiers in Physiology* **12**
- Crumb W. J., Vicente J., Johannesen L., Strauss D. G (2016) **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
- Dutta S., Chang K. C., Beattie K. A., Sheng J., Tran P. N., Wu W. W., Wu M., Strauss D. G., Colatsky T., Li Z (2017) **Optimization of an in silico cardiac cell model for proarrhythmia risk assessment** *Frontiers in Physiology* **8**
- Garny A., Hunter P. J (2015) **OpenCOR: a modular and interoperable approach to computational biology** *Frontiers in Physiology* **6**
- Gemmell P., Burrage K., Rodríguez B., Quinn T. A (2016) **Rabbit-specific computational modelling of ventricular cell electrophysiology: Using populations of models to explore variability in the response to ischemia** *Progress in Biophysics and Molecular Biology* **121**:169–184
- Ghosh S., Gavaghan D. J., Mirams G. R. (2018) **Gaussian process emulation for discontinuous response surfaces with applications for cardiac electrophysiology models**
- Grandi E., Pasqualini F. S., Bers D. M (2010) **A novel computational model of the human ventricular action potential and ca transient** *Journal of Molecular and Cellular Cardiology* **48**:112–121
- Groenendaal W., Ortega F. A., Kherlopian A. R., Zygmunt A. C., Krogh-Madsen T., Christini D. J (2015) **Cell-specific cardiac electrophysiology models** *PLOS Computational Biology* **11**
- Guillaume J. H. *et al.* (2019) **Introductory overview of identifiability analysis: A guide to evaluating whether you have the right type of data for your modeling purpose** *Environmental Modelling & Software* **119**:418–432
- Hall C., Gehmlich K., Denning C., Pavlovic D (2021) **Complex relationship between cardiac fibroblasts and cardiomyocytes in health and disease** *Journal of the American Heart Association* **10**
- Harris C. R. *et al.* (2020) **Array programming with NumPy** *Nature* **585**:357–362
- Herman J., Usher W (2017) **SALib: An open-source python library for sensitivity analysis** *J. Open Source Softw* **2**
- Homma T., Saltelli A (1996) **Importance measures in global sensitivity analysis of nonlinear models** *Reliab. Eng. Syst. Saf* **52**:1–17
- Hornik K., Stinchcombe M., White H (1989) **Multilayer feedforward networks are universal approximators** *Neural networks* **2**:359–366

- Jæger K. H., Edwards A. G., Giles W. R., Tveito A (2021) **A computational method for identifying an optimal combination of existing drugs to repair the action potentials of SQT1 ventricular myocytes** *PLOS Computational Biology* **17**
- Jæger K. H., Wall S., Tveito A (2019) **Detecting undetectables: Can conductances of action potential models be changed without appreciable change in the transmembrane potential?** *Chaos: An Interdisciplinary Journal of Nonlinear Science* **29**
- Jagtap A. D., Kawaguchi K., Karniadakis G. E (2020) **Adaptive activation functions accelerate convergence in deep and physics-informed neural networks** *Journal of Computational Physics* **404**
- Jeong D. U., Yoo Y., Marcellinus A., Kim K.-S., Lim K. M (2022) **Proarrhythmic risk assessment of drugs by dvm /dt shapes using the convolutional neural network** *CPT: Pharmacometrics & Systems Pharmacology* **11**:653–664
- Jeong D. U., Yoo Y., Marcellinus A., Lim K. M (2023) **Application of convolutional neural networks using action potential shape for in-silico proarrhythmic risk assessment** *Biomedicines* **11**
- Johnstone R. H., Chang E. T., Bardenet R., de Boer T. P., Gavaghan D. J., Pathmanathan P., Clayton R. H., Mirams G. R. (2016) **Uncertainty and variability in models of the cardiac action potential: Can we build trustworthy models?** *Journal of Molecular and Cellular Cardiology* **96**:49–62
- Kingma D. P., Ba J (2017) **Adam: A Method for Stochastic Optimization** *arXiv:1412.6980*
- Kohl P., Gourdie R. G (2014) **Fibroblast-myocyte electrotonic coupling: Does it occur in native cardiac tissue?** *Journal of Molecular and Cellular Cardiology* **70**:37–46
- Krogh-Madsen T., Sobie E. A., Christini D. J (2016) **Improving cardiomyocyte model fidelity and utility via dynamic electrophysiology protocols and optimization algorithms** *The Journal of Physiology* **594**:2525–2536
- Lei C. L., Mirams G. R (2021) **Neural Network Differential Equations For Ion Channel Modelling** *Frontiers in Physiology* **12**
- Lei C. L., Whittaker D. G., Mirams G. R (2023) **The impact of uncertainty in hERG binding mechanism on in silico predictions of drug-induced proarrhythmic risk** *British Journal of Pharmacology, TBA* <https://doi.org/10.1111/bph.16250>
- Li Z., Dutta S., Sheng J., Tran P. N., Wu W., Chang K., Mdluli T., Strauss D. G., Colatsky T (2017) **Improving the in silico assessment of proarrhythmia risk by combining hERG (human ether-áa-go-go-related gene) channel-drug binding kinetics and multichannel pharmacology** *Circulation: Arrhythmia and Electrophysiology* **10**
- Li Z. *et al.* (2019) **Assessment of an in silico mechanistic model for proarrhythmia risk prediction under the CiPA initiative** *Clinical Pharmacology & Therapeutics* **105**:466–475
- Mayer A., Bittihn P., Luther S (2017) **Complex restitution behavior and reentry in a cardiac tissue model for neonatal mice** *Physiological Reports* **5**

- Mirams G. R., Cui Y., Sher A., Fink M., Cooper J., Heath B. M., McMahon N. C., Gavaghan D. J., Noble D (2011) **Simulation of multiple ion channel block provides improved early prediction of compounds' clinical torsadogenic risk** *Cardiovascular Research* **91**:53–61
- Mirams G. R., Davies M. R., Brough S. J., Bridgland-Taylor M. H., Cui Y., Gavaghan D. J., Abi-Gerges N (2014) **Prediction of thorough QT study results using action potential simulations based on ion channel screens** *Journal of Pharmacological and Toxicological Methods* **70**:246–254
- Mirams G. R., Pathmanathan P., Gray R. A., Challenor P., Clayton R. H (2016) **Uncertainty and variability in computational and mathematical models of cardiac physiology** *The Journal of Physiology* **594**:6833–6847
- Mohammadi H., Challenor P., Goodfellow M (2019) **Emulating dynamic non-linear simulators using gaussian processes** *Computational Statistics & Data Analysis* **139**:178–196
- Morotti S. *et al.* (2021) **Quantitative cross-species translators of cardiac myocyte electrophysiology: Model training, experimental validation, and applications** *Science Advances* **7**
- Muszkiewicz A. *et al.* (2016) **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
- Muszkiewicz A. *et al.* (2016) **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
- Oakley J. E., O'Hagan A (2004) **Probabilistic sensitivity analysis of complex models: a bayesian approach** *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66**:751–769
- O'Hara T., Virág L., Varró A., Rudy Y (2011) **Simulation of the undiseased human cardiac ventricular action potential: Model formulation and experimental validation** *PLoS Computational Biology* **7**
- O'Hara T., Virág L., Varró A., Rudy Y (2011) **Simulation of the Undiseased Human Cardiac Ventricular Action Potential: Model Formulation and Experimental Validation** *PLOS Computational Biology* **7**
- Orvos P. *et al.* (2019) **Evaluation of possible proarrhythmic potency: Comparison of the effect of dofetilide, cisapride, sotalol, terfenadine, and verapamil on hERG and native iKr currents and on cardiac action potential** *Toxicological Sciences* **168**:365–380
- Passini E., Britton O. J., Lu H. R., Rohrbacher J., Hermans A. N., Gallacher D. J., Greig R. J. H., Bueno-Orovio A., Rodriguez B (2017) **Human in silico drug trials demonstrate higher accuracy than animal models in predicting clinical pro-arrhythmic cardiotoxicity** *Frontiers in Physiology* **8**
- Passini E., Zhou X., Trovato C., Britton O. J., Bueno-Orovio A., Rodriguez B (2021) **The virtual assay software for human in silico drug trials to augment drug cardiac testing** *Journal of Computational Science* **52**

- Paszke A. *et al.* (2019) **PyTorch: An Imperative Style, High-Performance Deep Learning Library** *Advances in Neural Information Processing Systems*
- Plank G. *et al.* (2021) **The openCARP simulation environment for cardiac electrophysiology** *Computer Methods and Programs in Biomedicine* **208**
- Rasmussen C. E., Williams C. K. I (2019) **Gaussian processes for machine learning** *Adaptive Computation and Machine Learning Series*
- Rush S., Larsen H (1978) **A practical algorithm for solving dynamic membrane equations** *IEEE Transactions on Biomedical Engineering, BME* **25**:389–392
- Sager P. T., Gintant G., Turner J. R., Pettit S., Stockbridge N (2014) **Rechanneling the cardiac proarrhythmia safety paradigm: A meeting report from the cardiac safety research consortium** *American Heart Journal* **167**:292–300
- Saltelli A (2002) **Making best use of model evaluations to compute sensitivity indices** *Computer Physics Communications* **145**:280–297
- Sarkar A. X., Sobie E. A (2010) **Regression analysis for constraining free parameters in electrophysiological models of cardiac cells** *PLoS Computational Biology* **6**
- Sher A., Niederer S. A., Mirams G. R., Kirpichnikova A., Allen R., Pathmanathan P., Gavaghan D. J., van der Graaf P. H., Noble D. (2022) **A quantitative systems pharmacology perspective on the importance of parameter identifiability** *Bulletin of Mathematical Biology* **84**
- Shim J. V. *et al.* (2023) **Predicting individual-specific cardiotoxicity responses induced by tyrosine kinase inhibitors** *Frontiers in Pharmacology* **14**
- Sobol', I (2001) **Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates** *Mathematics and Computers in Simulation* **55**:271–280
- Strauss D. G., Wu W. W., Li Z., Koerner J., Garnett C (2021) **Translational models and tools to reduce clinical trials and improve regulatory decision making for QTc and proarrhythmia risk (ICH e14/s7b updates)** *Clinical Pharmacology & Therapeutics* **109**:319–333
- Syed Z., Vigmond E., Nattel S., Leon L. J (2005) **Atrial cell action potential parameter fitting using genetic algorithms** *Medical and Biological Engineering and Computing* **43**:561–571
- Tomek J. *et al.* (2019) **Development, calibration, and validation of a novel human ventricular myocyte model in health, disease, and drug block** *eLife* **8**
- Tomek J., Bueno-Orovio A., Rodriguez B. (2020) **Tor-orddync1: an update of the tor-ord model of human ventricular cardiomyocyte with dynamic intracellular chloride** *bioRxiv*
- Tveito A., Jæger K. H., Huebsch N., Charrez B., Edwards A. G., Wall S., Healy K. E (2018) **Inversion and computational maturation of drug response using human stem cell derived cardiomyocytes in microphysiological systems** *Scientific Reports* **8**
- Tveito A., Jæger K. H., Maleckar M. M., Giles W. R., Wall S (2020) **Computational translation of drug effects from animal experiments to human ventricular myocytes** *Scientific Reports* **10**

- Vicente J. *et al.* (2017) **Mechanistic model-informed proarrhythmic risk assessment of drugs: Review of the “CiPA” initiative and design of a prospective clinical validation study** *Clinical Pharmacology & Therapeutics* **103**:54–66
- Vigmond E., dos Santos R. W., Prassl A., Deo M., Plank G. (2008) **Solvers for the cardiac bidomain equations** *Progress in Biophysics and Molecular Biology* **96**:3–18
- Weiss J. N., Garfinkel A., Karagueuzian H. S., Chen P.-S., Qu Z (2010) **Early afterdepolarizations and cardiac arrhythmias** *Heart Rhythm* **7**:1891–1899
- Whittaker D. G., Clerx M., Lei C. L., Christini D. J., Mirams G. R (2020) **Calibration of ionic and cellular cardiac electrophysiology models** *WIREs Systems Biology and Medicine* **12**
- Zaniboni M., Riva I., Cacciani F., Groppi M (2010) **How different two almost identical action potentials can be: A model study on cardiac repolarization** *Mathematical Biosciences* **228**:56–70

Article and author information

Thomas Grandits

Department of Mathematics and Scientific Computing, University of Graz, NAWI Graz, University of Graz

For correspondence: thomas.grandits@uni-graz.at

Christoph M. Augustin

Gottfried Schatz Research Center for Cell Signaling, Metabolism and Aging - Division of Medical Physics and Biophysics, Medical University of Graz, BioTechMed-Graz

Gundolf Haase

Department of Mathematics and Scientific Computing, University of Graz

Norbert Jost

Department of Pharmacology and Pharmacotherapy, University of Szeged, HUN-REN-TKI, Research Group of Pharmacology

Gary R. Mirams

Centre for Mathematical Medicine & Biology, School of Mathematical Sciences, University of Nottingham

ORCID iD: [0000-0002-4569-4312](https://orcid.org/0000-0002-4569-4312)

Steven A. Niederer

Division of Imaging Sciences & Biomedical Engineering, King’s College London

Gernot Plank

Gottfried Schatz Research Center for Cell Signaling, Metabolism and Aging - Division of Medical Physics and Biophysics, Medical University of Graz, BioTechMed-Graz

András Varró

Department of Pharmacology and Pharmacotherapy, University of Szeged, HUN-REN-TKI, Research Group of Pharmacology

László Virág

Department of Pharmacology and Pharmacotherapy, University of Szeged

Alexander Jung

Gottfried Schatz Research Center for Cell Signaling, Metabolism and Aging - Division of Medical Physics and Biophysics, Medical University of Graz

Copyright

© 2023, Grandits et al.

This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

Editors

Reviewing Editor

Yongliang Yang

Dalian University of Technology, Dalian, China

Senior Editor

Aleksandra Walczak

École Normale Supérieure - PSL, Paris, France

Reviewer #1 (Public Review):

Summary:

The authors present a neural network (NN)-based approach to computationally cheaper emulation of simulations of biophysically relatively detailed cardiac cell models based on systems of ordinary differential equations. Relevant case studies are used to demonstrate the performance in prediction of standard action potentials, as well as action potentials manifesting early depolarizations. Application to the "reverse problem" (inferring the effect of pharmacological compounds on ion channels based on action potential data before and after drug treatment) is also explored, which is a task of generally high interest.

Strengths:

This is a well-designed study, which explores an area that many in the cardiac simulation community will be interested in. The article is well written and I particularly commend the authors on transparency of methods description, code sharing, etc. - it feels rather exemplary in this regard and I only wish more authors of cardiac simulation studies took such an approach. The training speed of the network is encouraging and the technique is accessible to anyone with a reasonably strong GPU, not needing specialized equipment.

Weaknesses:

Below are several points that I consider to be weaknesses and/or uncertainties of the work:

1. The scope for acceleration of single cell simulations is not vast, as it is easy to simulate tens of thousands of cells per day on a workstation computer, using simulation conditions similar to those of the authors. While this covers a large part of what is needed in the field, I agree with the authors that there are applications where the presented technology is helpful. In such cases, e.g., in uncertainly quantification, it will enable studies that would be difficult to

carry out previously. In addition, any application involving long-term pre-pacing of a large number of cells will benefit greatly from the reported tool.

An area which is definitely in need of acceleration is simulations of whole ventricles or hearts, but it is not clear how much potential for speedup would the presented technology bring there. I can imagine interesting applications of rapid emulation in such a setting, some of which could be hybrid in nature (e.g. using simulation for the region around the wavefront of propagating electrical waves, while emulating the rest of the tissue, which is behaving more regularly/predictable, and is likely to be emulated well), but this is definitely beyond of the scope of this article.

2. The exact speed-up achieved by the NN emulation is somewhat context-dependent. In particular, the reported speedup critically depends on the number of beats in the simulation. The emulator learns to directly estimate the state of the cell after X beats (where X is decided by the operator of training). The speedup appears to be relatively marginal when a single beat is simulated versus emulated - but when 1000 beats are simulated, this takes 1000fold more time for simulation, but unchanged time for emulation.

While the initial submission did not communicate the practical speedup entirely clearly, this was addressed well by the authors in the revised version.

3. It appears that the accuracy of emulation drops off relatively sharply with increasing real-world applicability/relevance of the tasks it is applied to. That said, the authors are to be commended on declaring this transparently, rather than withholding such analyses. I particularly enjoyed the discussion of the not always amazing results of the inverse problem on the experimental data. The point on low parameter identifiability is an important one, and serves as a warning against overconfidence in our ability to infer cellular parameters from action potentials alone. On the other hand, I'm not that sure the difference between small tissue preps and single cells which authors propose as another source of the discrepancy will be that vast beyond the AP peak potential (probably much of the tissue prep is affected by the pacing electrode?), but that is a subjective view only. The influence of coupling could be checked if the simulated data were generated from 2D tissue samples/fibres, e.g. using the Myokit software.

In summary, I believe the range of tasks where the emulator provides a major advance is relatively narrow, particularly given the relatively limited need for further speedup compared to simulations. However, this does not make the study uninteresting in the slightest - on the contrary, it explores something that many of us are thinking about, and it is likely to stimulate further development in the direction of computationally efficient emulation of relatively complex simulations.

<https://doi.org/10.7554/eLife.91911.2.sa1>

Reviewer #3 (Public Review):

Summary:

1. Grandits and colleagues were trying to develop a new tool to accelerate pharmacological studies by using neural networks to emulate the human ventricular cardiomyocyte action potential (AP). The AP is a complex electrical signal that governs the heartbeat, and it is important to accurately model the effects of drugs on the AP to assess their safety and efficacy. Traditional biophysical simulations of the AP are computationally expensive and time-consuming. The authors hypothesized that neural network emulators could be trained to predict the AP with high accuracy and that these emulators could also be used to quickly and accurately predict the effects of drugs on the AP.

Strengths:

2. One of the study's major strengths is that the authors use a large and high-quality dataset to train their neural network emulator. The dataset includes a wide range of APs, including normal and abnormal APs exhibiting EADs. This ensures that the emulator is robust and can be used to predict the AP for a variety of different conditions.

Another major strength of the study is that the authors demonstrate that their neural network emulator can be used to accelerate pharmacological studies. For example, they use the emulator to predict the effects of a set of known arrhythmogenic drugs on the AP. The emulator is able to predict the effects of these drugs, even though it had not been trained on these drugs specifically.

Weaknesses:

One weakness of the study is that it is important to validate neural network emulators against experimental data to ensure that they are accurate and reliable. The authors do this to some extent, but further validation would be beneficial. In particular for the inverse problem, where the estimation of pharmacological parameters is very challenging and led to particularly large inaccuracies.

Additional context:

4. The work by Grandits et al. has the potential to revolutionize the way that pharmacological studies are conducted. Neural network emulation has the promise to reduce the time and cost of drug development and to improve the safety and efficacy of new drugs. The methods and data presented in the paper are useful to the community because they provide a starting point for other researchers to develop and improve neural network emulators for the human ventricular cardiomyocyte AP. The authors have made their code and data publicly available, which will facilitate further research in this area.

5. It is important to note that neural network emulation is still a relatively new approach, and there are some challenges that need to be addressed before it can be widely adopted in the pharmaceutical industry. For example, neural network emulators need to be trained on large and high-quality datasets. Additionally, it is important to validate neural network emulators against experimental data to ensure that they are accurate and reliable. Despite these challenges, the potential benefits of neural network emulation for pharmacological studies are significant. As neural network emulation technology continues to develop, it is likely to become a valuable tool for drug discovery and development.

<https://doi.org/10.7554/eLife.91911.2.sa0>

Author Response

The following is the authors' response to the original reviews.

eLife assessment

This valuable study reports on the potential of neural networks to emulate simulations of human ventricular cardiomyocyte action potentials for various ion channel parameters with the advantage of saving simulation time in certain conditions. The evidence supporting the claims of the authors is solid, although the inclusion of open analysis of drop-off accuracy and validation of the neural network emulators against experimental data would have strengthened the study. The work will be of interest to scientists working in cardiac simulation and quantitative pharmacology.

Thank you for the kind assessment. It is important for us to point out that, while limited, experimental validation was performed in this study and is thoroughly described in the work.

Reviewer 1 - Comments

This manuscript describes a method to solve the inverse problem of finding the initial cardiac activations to produce a desired ECG. This is an important question. The techniques presented are novel and clearly demonstrate that they work in the given situation. The paper is well-organized and logical.

Strengths:

This is a well-designed study, which explores an area that many in the cardiac simulation community will be interested in. The article is well written and I particularly commend the authors on transparency of methods description, code sharing, etc. - it feels rather exemplary in this regard and I only wish more authors of cardiac simulation studies took such an approach. The training speed of the network is encouraging and the technique is accessible to anyone with a reasonably strong GPU, not needing specialized equipment.

Weaknesses:

Below are several points that I consider to be weaknesses and/or uncertainties of the work:

C I-(a) I am not convinced by the authors' premise that there is a great need for further acceleration of cellular cardiac simulations - it is easy to simulate tens of thousands of cells per day on a workstation computer, using simulation conditions similar to those of the authors. I do not really see an unsolved task in the field that would require further speedup of single-cell simulations. At the same time, simulations offer multiple advantages, such as the possibility to dissect mechanisms of the model behaviour, and the capability to test its behaviour in a wide array of protocols - whereas a NN is trained for a single purpose/protocol, and does not enable a deep investigation of mechanisms. Therefore, I am not sure the cost/benefit ratio is that strong for single-cell emulation currently.

An area that is definitely in need of acceleration is simulations of whole ventricles or hearts, but it is not clear how much potential for speedup the presented technology would bring there. I can imagine interesting applications of rapid emulation in such a setting, some of which could be hybrid in nature (e.g. using simulation for the region around the wavefront of propagating electrical waves, while emulating the rest of the tissue, which is behaving more regularly/predictable, and is likely to be emulated well), but this is definitely beyond of the scope of this article.

Thank you for this point of view. Simulating a population of few thousand cells is completely feasible on single desktop machines and for fixed, known parameters, emulation may not fill ones need. Yet we still foresee a great untapped potential for rapid evaluations of ionic models, such as for the gradient-based inverse problem, presented in the paper. Such inverse optimization requires several thousand evaluations per cell and thus finding maximum conductances for the presented experimental data set (13 cell pairs control/drug → 26 APs) purely through simulations would require roughly a day of simulation time even in a very conservative estimation (3.5 seconds per simulation, 1000 simulations per optimization). Additionally, the emulator provides local sensitivity information between the AP and maximum conductances in the form of the gradient, which enables a whole new array of efficient optimization algorithms [Beck, 2017]. To further emphasize these points, we added

the number of emulations and runtime of each conducted experiment in the specific section and a paragraph in the discussion that addresses this point:

"Cardiomyocyte EP models are already very quick to evaluate in the scale of seconds (see Section 2.3.1), but the achieved runtime of emulations allows to solve time consuming simulation protocols markedly more efficient. One such scenario is the presented inverse maximum conductance estimation problem (see Section 3.1.2 and Section 3.1.3), where for estimating maximum conductances of a single AP, we need to emulate the steady state AP at least several hundred times as part of an optimization procedure. Further applications include the probabilistic use of cardiomyocyte EP models with uncertainty quantification [Chang et al., 2017, Johnstone et al., 2016] where thousands of samples of parameters are potentially necessary to compute a distribution of the steady-state properties of subsequent APs, and the creation of cell populations [Muszkiewicz et al., 2016, Gemmell et al., 2016, Britton et al., 2013]." (Section 4.2)

We believe that rapid emulations are valuable for several use-cases, where thousands of evaluations are necessary. These include the shown inverse problem, but similarly arise in uncertainty quantification, or cardiomyocyte population creation. Similarly, new use-cases may arise as such efficient tools become available. Additionally, we provided the number of evaluations along with the runtimes for each of the conducted experiments, showing how essential these speedups are to realize these experiments in reasonable timeframes. Utilizing these emulations in organ-level electrophysiological models is a possibility, but the potential problems in such scenarios are much more varied and depend on a number of factors, making it hard to pin-point the achievable speed-up using ionic emulations.

C I-(b) The authors run a cell simulation for 1000 beats, training the NN emulator to mimic the last beat. It is reported that the simulation of a single cell takes 293 seconds, while emulation takes only milliseconds, implying a massive speedup. However, I consider the claimed speedup achieved by emulation to be highly context-dependent, and somewhat too flattering to the presented method of emulation. Two specific points below:

First, it appears that a not overly efficient (fixed-step) numerical solver scheme is used for the simulation. On my (comparable, also a Threadripper) CPU, using the same model ("ToR-ORd-dyncl"), but a variable step solver ode15s in Matlab, a simulation of a cell for 1000 beats takes ca. 50 seconds, rather than 293 of the authors. This can be further sped up by parallelization when more cells than available cores are simulated: on 32 cores, this translates into ca. 2 seconds amortized time per cell simulation (I suspect that the NN-based approach cannot be parallelized in a similar way?). By amortization, I mean that if 32 models can be simulated at once, a simulation of X cells will not take X50 seconds, but (X/32)50. (with only minor overhead, as this task scales well across cores).

Second, and this is perhaps more important - the reported speed-up critically depends on the number of beats in the simulation - if I am reading the article correctly, the runtime compares a simulation of 1000 beats versus the emulation of a single beat. If I run a simulation of a single beat across multiple simulated cells (on a 32-core machine), the amortized runtime is around 20 ms per cell, which is only marginally slower than the NN emulation. On the other hand, if the model was simulated for aeons, comparing this to a fixed runtime of the NN, one can get an arbitrarily high speedup.

Therefore, I'd probably emphasize the concrete speedup less in an abstract and I'd provide some background on the speedup calculation such as above, so that the readers understand the context-dependence. That said, I do think that a simulation for anywhere between 250 and 1000 beats is among the most reasonable points of comparison (long enough for reasonable stability, but not too long to beat an already stable horse; pun

with stables was actually completely unintended, but here it is...). I.e., the speedup observed is still valuable and valid, albeit in (I believe) a somewhat limited sense.

We agree that the speedup comparison only focused on a very specific case and needs to be more thoroughly discussed and benchmarked. One of the main strengths of the emulator is to cut the time of prepacing to steady state, which is known to be a potential bottleneck for the speed of the single-cell simulations. The time it takes to reach the steady state in the simulator is heavily dependant on the actual maximum conductance configuration and the speed-up is thus heavily reliant on a per-case basis. The differences in architecture of the simulator and emulator further makes direct comparisons very difficult. In the revised version we now go into more detail regarding the runtime calculations and also compare it to an adaptive time stepping simulation (Myokit [Clerx et al., 2016]) in a new subsection:

"The simulation of a single AP (see Section 2.1) sampled at a resolution of 20kHz took 293s on one core of a AMD Ryzen Threadripper 2990WX (clock rate: 3.0GHz) in CARPentry. Adaptive timestep solver of variable order, such as implemented in Myokit [Clerx et al., 2016], can significantly lower the simulation time (30s for our setup) by using small step sizes close to the depolarization (phase 0) and increasing the time step in all other phases. The emulation of a steady state AP sampled at a resolution of 20kHz for $t \in [-10, 1000]$ ms took 18.7ms on a AMD Ryzen 7 3800X (clock rate: 3.9GHz) and 1.2ms on a Nvidia A100 (Nvidia Corporation, USA), including synchronization and data copy overhead between CPU and GPU.

"The amount of required beats to reach the steady state of the cell in the simulator has a major impact on the runtime and is not known a-priori. On the other hand, both simulator and emulator runtime linearly depends on the time resolution, but since the output of the emulator is learned, the time resolution can be chosen at arbitrarily without affecting the AP at the sampled times. This makes direct performance comparisons between the two methodologies difficult. To still be able to quantify the speed-up, we ran Myokit using 100 beats to reach steady state, taking 3.2s of simulation time. In this scenario, we witnessed a speed-up of 171 and $2 \cdot 10^3$ of our emulator on CPU and GPU respectively (again including synchronization and data copy overhead between CPU and GPU in the latter case). Note that both methods are similarly expected to have a linear parallelization speedup across multiple cells.

For the inverse problem, we parallelized the problem for multiple cells and keep the problem on the GPU to minimize the overhead, achieving emulations (including backpropagation) that run in 120 μ s per AP at an average temporal resolution of 2kHz. We consider this the peak performance which will be necessary for the inverse problem in Section 3.1.2." (Section 2.3.1)

Note that the mentioned parallelization across multiple machines/hardware applies equally to the emulator and simulator (linear speed-up), though the utilization for single cells is most likely different (single vs. multi-cell parallelization).

C I-(c) It appears that the accuracy of emulation drops off relatively sharply with increasing real-world applicability/relevance of the tasks it is applied to. That said, the authors are to be commended on declaring this transparently, rather than withholding such analyses. I particularly enjoyed the discussion of the not-always amazing results of the inverse problem on the experimental data. The point on low parameter identifiability is an important one and serves as a warning against overconfidence in our ability to infer cellular parameters from action potentials alone. On the other hand, I'm not that sure the difference between small tissue preps and single cells which authors propose as another source of the discrepancy will be that vast beyond the AP peak potential (probably much of the tissue prep is affected by the pacing electrode?), but that is a subjective view only. The influence of coupling could be checked if the simulated data were generated from 2D tissue samples/fibres, e.g. using the Myokit software.

Given the points above (particularly the uncertain need for further speedup compared to running single-cell simulations), I am not sure that the technology generated will be that broadly adopted in the near future.

However, this does not make the study uninteresting in the slightest - on the contrary, it explores something that many of us are thinking about, and it is likely to stimulate further development in the direction of computationally efficient emulation of relatively complex simulations.

We agree that the parameter identifiability is an important point of discussion. While the provided experimental data gave us great insights already, we still believe that given the differences in the setup, we can not draw conclusions about the source of inaccuracies with absolute certainty. The suggested experiment to test the influence of coupling is of interest for future works and has been integrated into the discussion. Further details are given in the response to the recommendation R III- (t)

Reviewer 2 - Comments

Summary:

This study provided a neural network emulator of the human ventricular cardiomyocyte action potential. The inputs are the corresponding maximum conductances and the output is the action potential (AP). It used the forward and inverse problems to evaluate the model. The forward problem was solved for synthetic data, while the inverse problem was solved for both synthetic and experimental data. The NN emulator tool enables the acceleration of simulations, maintains high accuracy in modeling APs, effectively handles experimental data, and enhances the overall efficiency of pharmacological studies. This, in turn, has the potential to advance drug development and safety assessment in the field of cardiac electrophysiology.

Strengths:

- 1. Low computational cost: The NN emulator demonstrated a massive speed-up of more than 10,000 times compared to the simulator. This substantial increase in computational speed has the potential to expedite research and drug development processes*
- 1. High accuracy in the forward problem: The NN emulator exhibited high accuracy in solving the forward problem when tested with synthetic data. It accurately predicted normal APs and, to a large extent, abnormal APs with early afterdepolarizations (EADs). High accuracy is a notable advantage over existing emulation methods, as it ensures reliable modeling and prediction of AP behavior*

C II-(a) Input space constraints: The emulator relies on maximum conductances as inputs, which explain a significant portion of the AP variability between cardiomyocytes. Expanding the input space to include channel kinetics parameters might be challenging when solving the inverse problem with only AP data available.

Thank you for this comment. We consider this limitation a major drawback, as discussed in Section 4.3. Identifiability is already an issue when only considering the most important maximum conductances. Further extending the problem to include kinetics will most likely only increase the difficulty of the inverse problem. For the forward problem though, it might be of interest to people studying ionic models to further analyze the effects of channel kinetics.

C II-(b) Simplified drug-target interaction: In reality, drug interactions can be time-, voltage-, and channel state-dependent, requiring more complex models with multiple parameters compared to the oversimplified model that represents the drug-target interactions by scaling the maximum conductance at control. The complex model could also pose challenges when solving the inverse problem using only AP data.

Thank you pointing out this limitation. We slightly adapted Section 4.3 to further highlight some of these limitations. Note however that the experimental drugs used have been shown to be influenced by this drug interaction in varying degrees [Li et al., 2017] (e.g. dofetilide vs. cisapride). However, the discrepancy in identifiability was mostly channel-based (0%-100%), whereas the variation in identifiability between drugs was much lower (39%-66%).

C II-(c) Limited data variety: The inverse problem was solved using AP data obtained from a single stimulation protocol, potentially limiting the accuracy of parameter estimates. Including AP data from various stimulation protocols and incorporating pacing cycle length as an additional input could improve parameter identifiability and the accuracy of predictions.

The proposed emulator architecture currently only considers the discussed maximum conductances as input and thus can only compensate when using different stimulation protocols. However, the architecture itself does not prohibit including any of these as parameters for future variants of the emulator. We potentially foresee future works extending on the architecture with modified datasets to include other parameters of importance, such as channel kinetics, stimulation protocols and pacing cycle lengths. These will however vary between the actual use-cases one is interested in.

C II-(d) Larger inaccuracies in the inverse problem using experimental data: The reasons for this result are not quite clear. Hypotheses suggest that it may be attributed to the low parameter identifiability or the training data set were collected in small tissue preparation.

The low parameter identifiability on some channels (e.g. GK1) poses a problem, for which we state multiple potential reasons. As of yet, no final conclusion can be drawn, warranting further research in this area.

Reviewer 3 - Comments

Summary:

Grandits and colleagues were trying to develop a new tool to accelerate pharmacological studies by using neural networks to emulate the human ventricular cardiomyocyte action potential (AP). The AP is a complex electrical signal that governs the heartbeat, and it is important to accurately model the effects of drugs on the AP to assess their safety and efficacy. Traditional biophysical simulations of the AP are computationally expensive and time-consuming. The authors hypothesized that neural network emulators could be trained to predict the AP with high accuracy and that these emulators could also be used to quickly and accurately predict the effects of drugs on the AP.

Strengths:

One of the study's major strengths is that the authors use a large and high-quality dataset to train their neural network emulator. The dataset includes a wide range of APs, including normal and abnormal APs exhibiting EADs. This ensures that the emulator is robust and can be used to predict the AP for a variety of different conditions.

Another major strength of the study is that the authors demonstrate that their neural network emulator can be used to accelerate pharmacological studies. For example, they use the emulator to predict the effects of a set of known arrhythmogenic drugs on the AP. The emulator is able to predict the effects of these drugs, even though it had not been trained on these drugs specifically.

C III-(a) One weakness of the study is that it is important to validate neural network emulators against experimental data to ensure that they are accurate and reliable. The authors do this to some extent, but further validation would be beneficial. In particular for the inverse problem, where the estimation of pharmacological parameters was very challenging and led to particularly large inaccuracies.

Thank you for this recommendation. Further experimental validation of the emulator in the context of the inverse problem would be definitely beneficial. Still, an important observation is that the identifiability varies greatly between channels. While the inverse problem is an essential reason for utilizing the emulator, it is also empirically validated for the pure forward problem and synthetic inverse problem, together with the (limited) experimental validation. The sources of problems arising in estimating the maximum conductances of the experimental tissue preparations are important to discuss in future works, as we now further emphasize in the discussion. See also the response to the recommendations R III-(t).

Reviewer 1 - Recommendations

R I-(a) Could further detail on the software used for the emulation be provided? E.g. based on section 2.2.2, it sounds like a CPU, as well as GPU-based emulation, is possible, which is neat.

Indeed as suspected, the emulator can run on both CPUs and GPUs and features automatic parallelization (per-cell, but also multi-cell), which is enabled by the engineering feats of PyTorch [Paszke et al., 2019]. This is now outlined in a bit more detail in Sec. 2 and 5.

"The trained emulator is provided as a Python package, heavily utilizing PyTorch [Paszke et al., 2019] for the neural network execution, allowing it to be executed on both CPUs and NVidia GPUs." (Section 5)

R I-(b) I believe that a potential use of NN emulation could be also in helping save time on prepacing models to stability - using the NN for "rough" prepacing (e.g. 1000 beats), and then running a simulation from that point for a smaller amount of time (e.g. 50 beats). One could monitor the stability of states, so if the prepacing was inaccurate, one could quickly tell that these models develop their state vector substantially, and they should be simulated for longer for full accuracy - but if the model was stable within the 50 simulated beats, it could be kept as it is. In this way, the speedup of the NN and accuracy and insightfulness of the simulation could be combined. However, as I mentioned in the public review, I'm not sure there is a great need for further speedup of single-cell simulations. Such a hybrid scheme as described above might be perhaps used to accelerate genetic algorithms used to develop new models, where it's true that hundreds of thousands to millions of cells are eventually simulated, and a speedup there could be practical. However one would have to have a separate NN trained for each protocol in the fitness function that is to be accelerated, and this would have to be retrained for each explored model architecture. I'm not sure if the extra effort would be worth it - but maybe yes to some people.

Thank you for this valuable suggestion. As pointed out in C I-(a), one goal of this study was to reduce the timeconsuming task of prepacing. Still, in its current form the emulator could not

be utilized for prepacing simulators, as only the AP is computed by the emulator. For initializing a simulation at the N-th beat, one would additionally need all computed channel state variables. However, a simple adaptation of the emulator architecture would allow to also output the mentioned state variables.

R I-(c) Re: "Several emulator architectures were tried on the training and validation data sets and the final choice was hand-picked as a good trade-off between high accuracy and low computational cost" - is it that the emulator architecture was chosen early in the development, and the analyses presented in the paper were all done with one previously selected architecture? Or is it that the analyses were attempted with all considered architectures, and the well-performing one was chosen? In the latter case, this could flatter the performance artificially and a test set evaluation would be worth carrying out.

We apologize for the unclear description of the architectural validation. The validation was in fact carried out with 20% of the training data (data set #1), which is however completely disjoint with the test set (#2, #3, #4, formerly data set #1 and #2) on which the evaluation was presented. To further clarify the four different data sets used in the study, we now dedicated an additional section to describing each set and where it was used (see also our response below R I-(d)), and summarize them in Table 1, which we also added at R II-(a). The cited statement was slightly reworked.

"Several emulator architectures were tried on the training and validation data sets and the final choice was hand-picked as a good trade-off between high accuracy on the validation set (#1) and low computational runtime cost." (Section 2.2.2)

R I-(d) When using synthetic data for the forward and inverse problem, with the various simulated drugs, is it that split of the data into training/validation test set was done by the drug simulated (i.e., putting 80 drugs and the underlying models in the training set, and 20 into test set)? Or were the data all mixed together, and 20% (including drugs in the test set) were used for validation? I'm slightly concerned by the potential of "soft" data leaks between training/validation sets if the latter holds. Presumably, the real-world use case, especially for the inverse problem, will be to test drugs that were not seen in any form in the training process. I'm also not sure whether it's okay to reuse cell models (sets of max conductances) between training and validation tests - wouldn't it be better if these were also entirely distinct? Could you please comment on this?

We completely agree with the main points of apprehension that training, validation and test sets all serve a distinct purpose and should not be arbitrarily mixed. However, this is only a result of the sub-optimal description of our datasets, which we heavily revised in Section 2.2.1 (Data, formerly 2.3.1). We now present the data using four distinct numbers: The initial training/validation data, now called data set #1 (formerly no number), is split 80%/20% into training and validation sets (for architectural choices) respectively. The presented evaluations in Section 2.3 (Evaluation) are purely performed on data set #2 (normal APs, formerly #1), #3 (EADs, formerly #2) and #4 (experimental).

R I-(e) For the forward problem on EADs, I'm not sure if the 72% accuracy is that great (although I do agree that the traces in Fig 12-left also typically show substantial ICaL reactivation, but this definitely should be present, given the IKr and ICaL changes). I would suggest that you also consider the following design for the EAD investigation: include models with less severe upregulation of ICaL and downregulation of IKr, getting a population of models where a part manifests EADs and a part does not. Then you could run the emulator on the input data of this population and be able to quantify true, falsepositive, negative detections. I think this is closer to a real-world use case where we have drug parameters and a cell population, and we want to quickly assess the

arrhythmic risk, with some drugs being likely entirely nonrisky, some entirely risky, and some between (although I still am not convinced it's that much of an issue to just simulate this in a couple of thousands of cells).

Thank you for pointing out this alternative to address the EAD identification task. Even though the values chosen in Table 2 seem excessively large, we still only witnessed EADs in 171 of the 950 samples. Especially border cases, which are close to exhibiting EADs are hardest to estimate for the NN emulator. As suggested, we now include the study with the full 950 samples (non-EAD & EAD) and classify the emulator AP into one of the labels for each sample. The mentioned 72.5% now represent the sensitivity, whereas our accuracy in such a scenario becomes 90.8% (total ratio of correct classifications):

"The data set #3 was used second and Appendix C shows all emulated APs, both containing the EAD and non-EAD cases. The emulation of all 950 APs took 0.76s on the GPU specified in Section 2.2.3 We show the emulation of all maximum conductances and the classification of the emulation. The comparison with the actual EAD classification (based on the criterion outlined in Appendix A) results in true-positive (EAD both in the simulation and emulation), false-negative (EAD in the simulation, but not in the emulation), false-positive (EAD in the emulation, but not in the simulation) and true-negative (no EAD both in the emulation and simulation). The emulations achieved 72.5% sensitivity (EAD cases correctly classified) and 94.9% specificity (non-EAD cases correctly classified), with an overall accuracy of 90.8% (total samples correctly classified). A substantial amount of wrongly classified APs showcase a notable proximity to the threshold of manifesting EADs. Figure 7 illustrates the distribution of RMSEs in the EAD APs between emulated and ground truth drugged APs. The average RMSE over all EAD APs was 14.5mV with 37.1mV being the maximum. Largest mismatches were located in phase 3 of the AP, in particular in emulated APs that did not fully repolarize." (Section 3.1.1)

R I-(f) Figure 1 - I think a large number of readers will understand the mathematical notation describing inputs/outputs; that said, there may be a substantial number of readers who may find that hard to read (e.g. lab-based researchers, or simulation-based researchers not familiar with machine learning). At the same time, this is a very important part of the paper to explain what is done where, so I wonder whether using words to describe the inputs/outputs would not be more practical and easier to understand (e.g. "drug-based conductance scaling factor" instead of "s" ?). It's just an idea - it needs to be tried to see if it wouldn't make the figure too cluttered.

We agree that the mathematical notation may be confusing to some readers. As a compromise between using verbose wording and mathematical notation, we introduced a legend in the lower right corner of the figure that shortly describes the notation in order to help with interpreting the figure.

R I-(g) "APs with a transmembrane potential difference of more than 10% of the amplitude between $t = 0$ and 1000 ms were excluded" - I'm not sure I understand what exactly you mean here - could you clarify?

With this criterion, we try to discard data that is far away from fully repolarizing within the given time frame, which applies to 116 APs in data set #1 and 50 APs in data set #3. We added a small side note into the text:

"APs with a transmembrane potential difference of more than 10% of the amplitude between $t = 0$ and 1000ms (indicative of an AP that is far away from full repolarization) were excluded." (Section 2.2.1)

R I-(h) Speculation (for the future) - it looks like a tool like this could be equally well used to predict current traces, as well as action potentials. I wonder, would there be a likely benefit in feeding back the currents-traces predictions on the input of the AP predictor to provide additional information? Then again, this might be already encoded within the network - not sure.

Although not possible with the chosen architecture (see also R I-(b)), it is worth thinking about an implementation in future works and to study differences to the current emulator.

Entirely minor points:

R I-(i) "principle component analysis" → principal component analysis

Fixed

R I-(j) The paper will be probably typeset by elife anyway, but the figures are often quite far from their sections, with Results figures even overflowing into Discussion. This can be often fixed by using the !htb parameters (\begin{figure}[!htb]), or potentially by using "\usepackage[section]{placeins}" and then "\FloatBarrier" at the start and end of each section (or subsection) - this prevents floating objects from passing such barriers.

Thank you for these helpful suggestions. We tried reducing the spacing between the figures and their references in the text, hopefully improving the reader's experience.

R I-(k) Alternans seems to be defined in Appendix A (as well as repo-/depolarization abnormalities), but is not really investigated. Or are you defining these just for the purpose of explaining what sorts of data were also included in the data?

We defined alternans since this was an exclusion criterion for generating simulation data.

Reviewer 2 - Recommendations

R II-(a) Justification for methods selection: Explain the rationale behind important choices, such as the selection of specific parameters and algorithms.

Thank you for this recommendation, we tried to increase transparency of our choices by introducing a separate data section that summarizes all data sets and their use cases in Section 2.2.1 and also collect many of the explanations there. Additionally we added an overview table (Table 1) of the utilized data.

Author response table 1.

Table 1: Summary of the data used in this study, along with their usage and the number of valid samples. Note that each AP is counted individually, also in cases of control/drug pairs.

ID	Description	Usage	Origin	Samples
#1	Training/validation data	Training and validating the emulator, choosing the best architecture (Section 2.2.2)	Simulation	39,884
#2	Synthetic drug data, normal APs	Testing forward and inverse performance for normal APs (Section 3.1.1 and Section 3.1.2)	Simulation	10 ⁴
#3	Synthetic drug data, including EAD APs	Testing forward performance of abnormal (EAD) APs (Section 3.1.1)	Simulation	950
#4	Experimental cardiomyocytes	Testing and comparing the inverse performance with data published by the CiPA initiative [Li et al., 2017, Chang et al., 2017] (Section 3.1.3)	[Orvos et al., 2019]	

R II-(b) Interpretation of the evaluation results: After presenting the evaluation results, consider interpretations or insights into what the results mean for the performance of the emulator. Explain whether the emulator achieved the desired accuracy or compare it with other existing methods. In the revised version, we tried to further expand the discussion on possible applications of our emulator (Section 4.2). See also our response to C I-(a). To the best of our knowledge, there are currently no out-of-the-box methods available for directly comparing all experiments we considered in our work.

Reviewer 3 - Recommendations

R III-(a) In the introduction (Page 3) and then also in the 2.1 paragraph authors speak about the "limit cycle": Do you mean steady state conditions? In that case, it is more common to use steady state.

When speaking about the limit cycle, we refer to what is also sometimes called the steady state, depending on the field of research and/or personal preference. We now mention both terms at the first occurrence, but stick with the limit cycle terminology which can also be found in other works, see e.g. [Endresen and Skarland, 2000].

R III-(b) On page 3, while comparing NN with GP emulators, I still don't understand the key reason why NN can solve the discontinuous functions with more precision than GP.

The potential problems in modeling sharp continuities using GPs is further explained in the referenced work [Ghosh et al., 2018] and further references therein:

"Statistical emulators such as Gaussian processes are frequently used to reduce the computational cost of uncertainty quantification, but discontinuities render a standard Gaussian process emulation approach unsuitable as these emulators assume a smooth and continuous response to changes in parameter values [...] Applying GPs to model discontinuous functions is largely an open problem. Although many advances (see the discussion about non-stationarity in [Shahriari et al., 2016] and the references in there) have been made towards solving this problem, a common solution has not yet emerged. In the recent GP literature there are two specific streams of work that have been proposed for modelling non-stationary response surfaces including those with discontinuities. The first approach is based on designing nonstationary processes [Snoek et al., 2014] whereas the other approach attempts to divide the input space into separate regions and build separate GP models for each of the segmented regions. [...]"([Ghosh et al., 2018])

We integrated a short segment of this explanation into Section 1.

R III-(c) Why do authors prefer to use CARPentry and not directly openCARP? The use of CARPentry is purely a practical choice since the simulation pipeline was already set up. As we now point out however in Sec. 2.1 (Simulator), simulations can also be performed using any openly available ionic simulation tool, such as Myokit [Clerx et al., 2016], OpenCOR [Garny and Hunter, 2015] and openCARP [Plank et al., 2021]. We emphasized this in the text.

"Note, that the simulations can also be performed using open-source software such as Myokit [Clerx et al., 2016], OpenCOR [Garny and Hunter, 2015] and openCARP [Plank et al., 2021]." (Section 2.1)

R III-(d) In paragraph 2.1:

(a) In this sentence: "Various solver and sampling time steps were applied to generate APs and the biomarkers used in this study (see Appendix A)" this reviewer suggests putting the Appendix reference near "biomarkers". In addition, a figure that shows the test of various solver vs. sampling time steps could be interesting and can be added to the Appendix as well.

(b) Why did the authors set the relative difference below 5% for all biomarkers? Please give a reference to that choice. Instead, why choose 2% for the time step?

1. We adjusted the reference to be closer to "biomarkers". While we agree that further details on the influence of the sampling step would be of interest to some of the readers, we feel that it is far beyond the scope of this paper.

1. There is no specific reference we can provide for the choice. Our goal was to reach 5% relative difference, which we surpassed by the chosen time steps of 0.01 ms (solver) and 0.05 ms (sampling), leading to only 2% difference. We rephrased the sentence in question to make this clear.

"We considered the time steps with only 2% relative difference for all AP biomarkers (solver: 0.01ms; sampling: 0.05ms) to offer a sufficiently good approximation." (Section 2.1)

R III-(e) In the caption of Figure 1 authors should include the reference for AP experimental data (are they from Orvos et al. 2019 as reported in the Experimental Data section?)

We added the missing reference as requested. As correctly assumed, they are from [Orvos et al., 2019].

R III-(f) Why do authors not use experimental data in the emulator development/training?

For the supervised training of our NN emulator, we need to provide the maximum conductances of our chosen channels for each AP. While it would be beneficial to also include experimental data in the training to diversify the training data, the exact maximum conductances in our the considered retrospective experiments are not known. In the case such data would be available with low measurement uncertainty, it would be possible to include.

R III-(g) What is TP used in the Appendix B? I could not find the acronymous explanation.

We are sorry for the oversight, TP refers to the time-to-peak and is now described in Appendix A.

R III-(h) Are there any reasons for only using ST and no S1? Maybe are the same?

The global sensitivity analysis is further outlined in Appendix B, also showing S1 (first-order effects) and ST (variance of all interactions) together (Figure 11) [Herman and Usher, 2017] and their differences (e.g. in TP) Since S1 only captures first-order effects, it may fail to capture higher-order interactions between the maximum conductances, thus we favored ST.

R III-(i) In Training Section Page 8. It is not clear why it is necessary to resample data. Can you motivate?

The resampling part is motivated by exactly capturing the swift depolarization dynamics, whereas the output from CARPentry is uniformly sampled. This is now further highlighted in the text.

"Then, the data were non-uniformly resampled from the original uniformly simulated APs, to emphasize the depolarization slope with a high accuracy while lowering the number of repolarization samples. For this purpose, we resampled the APs [...]" (Section 2.2.1)

R III-(j) For the training of the neuronal network, the authors used the ADAM algorithm: have you tested any other algorithm?

For training neural networks, ADAM has become the current de-facto standard and is certainly a robust choice for training our emulator. While there may exist slightly faster, or better-suited training algorithms, we witnessed (qualitative) convergence in the training (Equation (2)). We thus strongly believe that the training algorithm is not a limiting factor in our study.

R III-(k) What is the amount of the drugs tested? Is the same dose reported in the description of the second data set or the values are only referring to experimental data? Moreover, it is not clear if in the description of experimental data, the authors are referring to newly acquired data (since they described in detail the protocol) or if they are obtained from Orvos et al. 2019 work.

In all scenarios, we tested 5 different drugs (cisapride, dofetilide, sotalol, terfenadine, verapamil). We revised our previous presentation of the data available, and now try to give a concise overview over the utilized data (Section 2.2.1 and table 1) and drug comparison with the CiPA distributions (Table 5, former 4). Note that in the latter case, the available expected channel scaling factors by the CiPA distributions vary, but are now clearly shown in Table 5.

R III-(l) In Figure 4, I will avoid the use of "control" in the legend since it is commonly associated with basal conditions and not with the drug administration.

The terminology "control" in this context is in line with works from the CiPA initiative, e.g. [Li et al., 2017] and refers to the state of cell conditions before the drug wash-in. We added a minor note the first time we use the term control in the introduction to emphasize that we refer to the state of the cell before administering any drugs

"To compute the drugged AP for given pharmacological parameters is a forward problem, while the corresponding inverse problem is to find pharmacological parameters for given

control (before drug administration) and drugged AP." (Section 1)

R III-(m) In Table 1 when you referred to Britton et al. 2017 work, I suggest adding also 10.1371/journal.pcbi.1002061.

We added the suggested article as a reference.

R III-(n) For the minimization problem, only data set #1 has been used. Have you tested data set #2?

In the current scenario, we only tested the inverse problem for data set #2 (former #1). The main purpose for data set #3 (former #2), was to test the possibility to emulate EAD APs. Given the overall lower performance in comparison to data set #2 (former #1), we also expect deteriorated results in comparison to the existing inverse synthetic problem.

R III-(o) In Figure 6 you should have the same x-axis (we could not see any points in the large time scale for many biomarkers). Why dVmMax is not uniform distributed compared to the others? Can you comment on that?

As suggested, we re-adjusted the x-range to show the center of distributions. Additionally, we denoted in each subplot the number of outliers which lie outside of the shown range. The error distribution on dVmMax exhibits a slightly off-center, left-tailed normal distribution, which we now describe a bit more in the revised text:

"While the mismatches in phase 3 were simply a result of imperfect emulation, the mismatches in phase 0 were a result of the difficulty in matching the depolarization time exactly. [...] Likewise, the difficulty in exactly matching the depolarization time leads to elevated errors and more outliers in the biomarkers influenced by the depolarization phase (TP and dVmMax)," (Section 3.1.1)

R III-(p) Page 14. Can the authors better clarify "the average RMSE over all APs 13.6mV": is it the mean for all histograms in Figure 7? (In Figure 5 is more evident the average RMSE).

The average RMSE uses the same definition for Figures 5 and 7: It is the average over all the RMSEs for each pair of traces (simulated/emulated), though the amount of samples is much lower for the EAD data set and not normal distributed.

R III-(q) In Table 4, the information on which drugs are considered should be added. For each channel, we added the names of the drugs for which respective data from the CiPA initiative were available.

R III-(r) Pag. 18, second paragraph, there is a repetition of "and".

Fixed

R III-(s) The pair's combination of scaling factors for simulating synthetic drugs reported in Table 2, can be associated with some effects of real drugs? In this case, I suggest including the information or justifying the choice.

The scaling factors in Table 2 are used to create data set #3 (former #2), and is meant to provide several APs which expose EADs. This is described in more detail in the new data section, Section 2.2.1:

"Data set #3: The motivation for creating data set #3 was to test the emulator on data of abnormal APs showing the repolarization abnormality EAD. This is considered a particularly relevant AP abnormality in pharmacological studies because of their role in the genesis of drug-induced ventricular arrhythmia's [Weiss et al., 2010]. Drug data were created using ten synthetic drugs with the hERG channel and the Cav1.2 channel as targets. To this end, ten samples with pharmacological parameters for GKr and PCa (Table 2) were generated and the synthetic drugs were applied to the entire synthetic cardiomyocyte population by scaling GKr and PCa with the corresponding pharmacological parameter. Of the 1000 APs simulated, we discarded APs with a transmembrane potential difference of more than 10% of the amplitude between $t = 0$ and 1000ms (checked for the last AP), indicative of an AP that does not repolarize within 1000ms. This left us with 950 APs, 171 of which exhibit EAD (see Appendix C)." (Section 2.2.1)

R III-(t) A general comment on the work is that the authors claim that their study highlights the potential of NN emulators as a powerful tool for increased efficiency in future quantitative systems pharmacology studies, but they wrote "Larger inaccuracies were found in the inverse problem solutions on experimental data highlight inaccuracies in estimating the pharmacological parameters": so, I was wondering how they can claim the robustness of NN use as a tool for more efficient computation in pharmacological studies.

The discussed robustness directly refers to efficiently emulating steady-state/limit cycle APs from a set of maximum conductances (forward problem, Section 3.1.1). We extensively evaluated the algorithm and feel that given the low emulation RMSE of APs (< 1 mV), the statement is warranted. The inverse estimation, enabled through this rapid evaluation, performs well on synthetic data, but shows difficulties for experimental data. Note however that at this point there are multiple potential sources for these problems as highlighted in the Evaluation section (Section 4.1) and Table 5 (former 4) highlights the difference in accuracy of estimating per-channel maximum conductances, revealing a potentially large discrepancy. The emulator also offers future possibilities to incorporate additional informations in the forms of either priors, or more detailed measurements (e.g. calcium transients) and can be potentially improved to a point where also the inverse problem can be satisfactorily solved in experimental preparations, though further analysis will be required.

References [Beck, 2017] Beck, A. (2017). First-order methods in optimization. SIAM.

[Britton et al., 2013] Britton, O. J., Bueno-Orovio, A., Ammel, K. V., Lu, H. R., Towart, R., Gallacher, D. J., and Rodriguez, B. (2013). Experimentally calibrated population of models predicts and explains intersubject variability in cardiac cellular electrophysiology. *Proceedings of the National Academy of Sciences*, 110(23).

[Chang et al., 2017] Chang, K. C., Dutta, S., Mirams, G. R., Beattie, K. A., Sheng, J., Tran, P. N., Wu, M., Wu, W. W., Colatsky, T., Strauss, D. G., and Li, Z. (2017). Uncertainty quantification reveals the importance of data variability and experimental design considerations for in silico proarrhythmia risk assessment. *Frontiers in Physiology*, 8.

[Clerx et al., 2016] Clerx, M., Collins, P., de Lange, E., and Volders, P. G. A. (2016). Myokit: A simple interface to cardiac cellular electrophysiology. *Progress in Biophysics and Molecular Biology*, 120(1):100–114.

[Endresen and Skarland, 2000] Endresen, L. and Skarland, N. (2000). Limit cycle oscillations in pacemaker cells. *IEEE Transactions on Biomedical Engineering*, 47(8):1134–1137.

[Garny and Hunter, 2015] Garny, A. and Hunter, P. J. (2015). OpenCOR: a modular and interoperable approach to computational biology. *Frontiers in Physiology*, 6.

[Gemmell et al., 2016] Gemmell, P., Burrage, K., Rodríguez, B., and Quinn, T. A. (2016). Rabbit-specific computational modelling of ventricular cell electrophysiology: Using populations of models to explore variability in the response to ischemia. *Progress in Biophysics and Molecular Biology*, 121(2):169–184.

[Ghosh et al., 2018] Ghosh, S., Gavaghan, D. J., and Mirams, G. R. (2018). Gaussian process emulation for discontinuous response surfaces with applications for cardiac electrophysiology models.

[Herman and Usher, 2017] Herman, J. and Usher, W. (2017). SALib: An open-source python library for sensitivity analysis. *J. Open Source Softw.*, 2(9):97.

[Johnstone et al., 2016] Johnstone, R. H., Chang, E. T., Bardenet, R., de Boer, T. P., Gavaghan, D. J., Pathmanathan, P., Clayton, R. H., and Mirams, G. R. (2016). Uncertainty and variability in models of the cardiac action potential: Can we build trustworthy models? *Journal of Molecular and Cellular Cardiology*, 96:49–62.

[Li et al., 2017] Li, Z., Dutta, S., Sheng, J., Tran, P. N., Wu, W., Chang, K., Mdluli, T., Strauss, D. G., and Colatsky, T. (2017). Improving the in silico assessment of proarrhythmia risk by combining hERG (human etherà-go-go-related gene) channel–drug binding kinetics and multichannel pharmacology. *Circulation: Arrhythmia and Electrophysiology*, 10(2).

[Muszkiewicz et al., 2016] Muszkiewicz, A., Britton, O. J., Gemmell, P., Passini, E., Sánchez, C., Zhou, X., Carusi, A., Quinn, T. A., Burrage, K., Bueno-Orovio, A., and Rodríguez, B. (2016). 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(1):115–127.

[Orvos et al., 2019] Orvos, P., Kohajda, Z., Szlovák, J., Gazdag, P., Arpádfy-Lovas, T., Tóth, D., Geramipour, A., Tálosi, L., Jost, N., Varró, A., and Virág, L. (2019). Evaluation of possible proarrhythmic potency: Comparison of the effect of dofetilide, cisapride, sotalol, terfenadine, and verapamil on hERG and native iKr currents and on cardiac action potential. *Toxicological Sciences*, 168(2):365–380.

[Paszke et al., 2019] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

[Plank et al., 2021] Plank, G., Loewe, A., Neic, A., Augustin, C., Huang, Y.-L., Gsell, M. A., Karabelas, E., Nothstein, M., Prassl, A. J., Sánchez, J., Seemann, G., and Vigmond, E. J. (2021). The openCARP simulation environment for cardiac electrophysiology. *Computer Methods and Programs in Biomedicine*, 208:106223.

[Shahriari et al., 2016] Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., and de Freitas, N. (2016). Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proceedings of the IEEE*, 104(1):148–175. Conference Name: Proceedings of the IEEE.

[Snoek et al., 2014] Snoek, J., Swersky, K., Zemel, R., and Adams, R. (2014). Input Warping for Bayesian Optimization of Non-Stationary Functions. In *Proceedings of the 31st International Conference on Machine Learning*, pages 1674–1682. PMLR. ISSN: 1938-7228.

[Weiss et al., 2010] Weiss, J. N., Garfinkel, A., Karagueuzian, H. S., Chen, P.-S., and Qu, Z. (2010). Early afterdepolarizations and cardiac arrhythmias. *Heart Rhythm*, 7(12):1891–1899.