PNAS
www.pnas.org

Supplementary Information for

Prediction of arrhythmia susceptibility through mathematical modeling and machine learning.

Meera Varshneya, Xueyan Mei, Eric A. Sobie

Corresponding Author:
Eric A. Sobie
Email: eric.sobie@mssm.edu

This PDF file includes:

Supplementary Methods

The code used to run the simulations in this manuscript have been uploaded to the following repository https://github.com/meeravarshneya1234/ArrhythmiaPredictionProject.git.

Mechanistic Mathematical Model Methods

Stimulation Protocol

The mathematical model was stimulated with an amplitude of -32.2 uA/uF for 2 ms and paced at 1 Hz(1). It was run to steady-state, by pacing the model for 2000 beats and then evaluating whether the action potential duration (APD) varied by +/- 1% for 500 consecutive beats. Simulations were performed using MATLAB v2019b with a 64-bit Intel processor and Windows 10. Model equations were evaluated using the MATLAB function *ode15s* with a relative tolerance of 10^{-3} and absolute tolerance of 10^{-6}.

Cellular Waveform Features

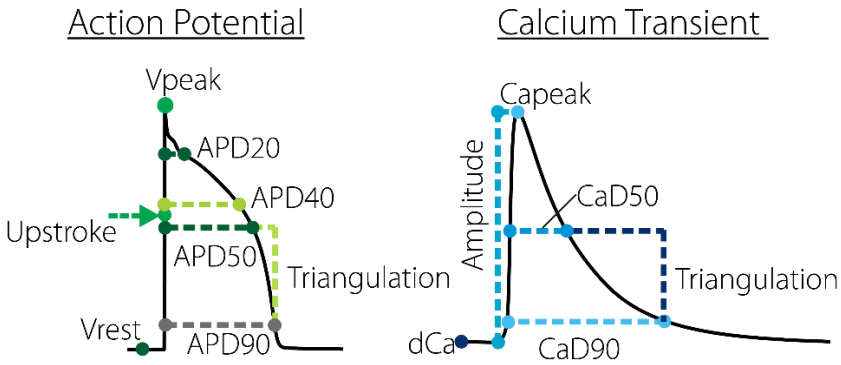
We computed the following metrics for each cell in the population:

Action Potential (AP)

1) peak membrane voltage (Vpeak)
2) upstroke velocity (Upstroke)
3) resting membrane potential (Vrest)
4) AP duration at 20% repolarization (APD20)
5) AP duration at 40% repolarization (APD40)
6) AP duration at 50% repolarization (APD50)
7) AP duration at 90% repolarization (APD90)
8) Difference between APD50 and APD90 (Triangulation)

Calcium Transient (CaT)

1) diastolic calcium (dCa)
2) CaT Amplitude
3) CaT Peak (Capeak)
4) CaT duration at 50% return to baseline (CaD50)
5) CaT duration at 90% return to baseline (CaD90)
6) Difference between CaD90 and CaD50 (Triangulation)



Machine Learning Methods

Supervised Machine Learning

We trained eight machine learning (ML) algorithms which include Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), Random Forest (RF), Naïve Bayes (NB), Gradient Boosting (GB), XGBoost (XGB), Logistic Regression (LR), and K-Nearest Neighbors (KNN). We utilized ML package *scikit-learn* version 0.21.3 in Python 3.7.4(2). To train each algorithm, we split the population into 90% training and 10% testing stratified by the target class. We normalized the input features using *MinMaxScaler* for MLP and StandardScaler for the remaining. The *MinMaxScaler* normalized the features to a range of 0 and 1, while the *StandardScaler* normalized by removing the mean and scaling to the unit variance. To fine-tune the hyperparameters for each classifier we applied the function *GridSearchCV* on the training set. This allowed us to effectively loop through a series of different parameter sets using 3-fold cross validation. The best parameters were then used to assess the performance of the algorithms on the test set.

The SVM classifier works by computing a hyperplane that maximizes the distance between positive and negative classes and minimizes misclassification. To build this model, we tuned three hyperparameters- the kernel (linear and radial basis function (rbf)), kernel coefficient (λ), and regularization (C). Often, features are linearly inseparable in a low dimensional space, but can become linearly distinguishable when mapped to a higher dimensional space using the rbf kernel. We tuned λ and C to find a decision boundary that finds a balance between maximizing the margin and minimizing misclassification. On the other hand, a MLP classifier is a feedforward artificial neural network that transforms the input using non-linear activation functions and projects it into a space where the data can be linearly separated. By tuning the number of layers, number of hidden nodes in each layer, the activation function, and the learning rate of the neural network, we found the optimal set of points parameters needed to have a high prediction and prevent overfitting.

To evaluate the performance of each classifier, we used the following metrics: receiver operator characteristic (ROC) and the calculated area under the ROC (auROC), accuracy, positive and negative predictive value, and specificity and sensitivity. We used the Youden index to select the optimal threshold to determine the predicted class of being arrhythmia(3). The probability greater than the selected threshold is labeled as arrhythmia by the ML classifiers. Accuracy, positive and negative predictive value and specificity and sensitivity are determined by this selected cutoff. Since our dataset was balanced, we saw similar trends in each metric and have thus reported only the ROC and auROC in the main text and placed the remaining on our github repository.

Supplementary Figures

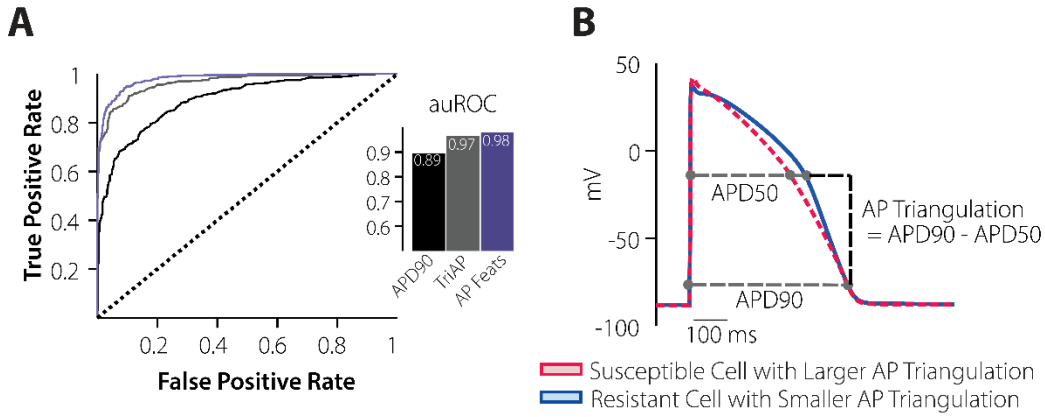


Fig. S1. Machine learning feature analysis reveals that current injection trigger performs exceptionally well due to the addition of Action Potential (AP) Triangulation.
A.) Receiver operator curves (ROC) and area under the curve (auROCs) comparing the performance of predicting susceptibility to current injection with only APD90 (black), only AP Triangulation (grey), and all AP Waveform features (purple). B.) Examples of two cells from the population that have the same APD90 but different levels of risk. The triangulation of the AP is more informative than simply looking at APD90.

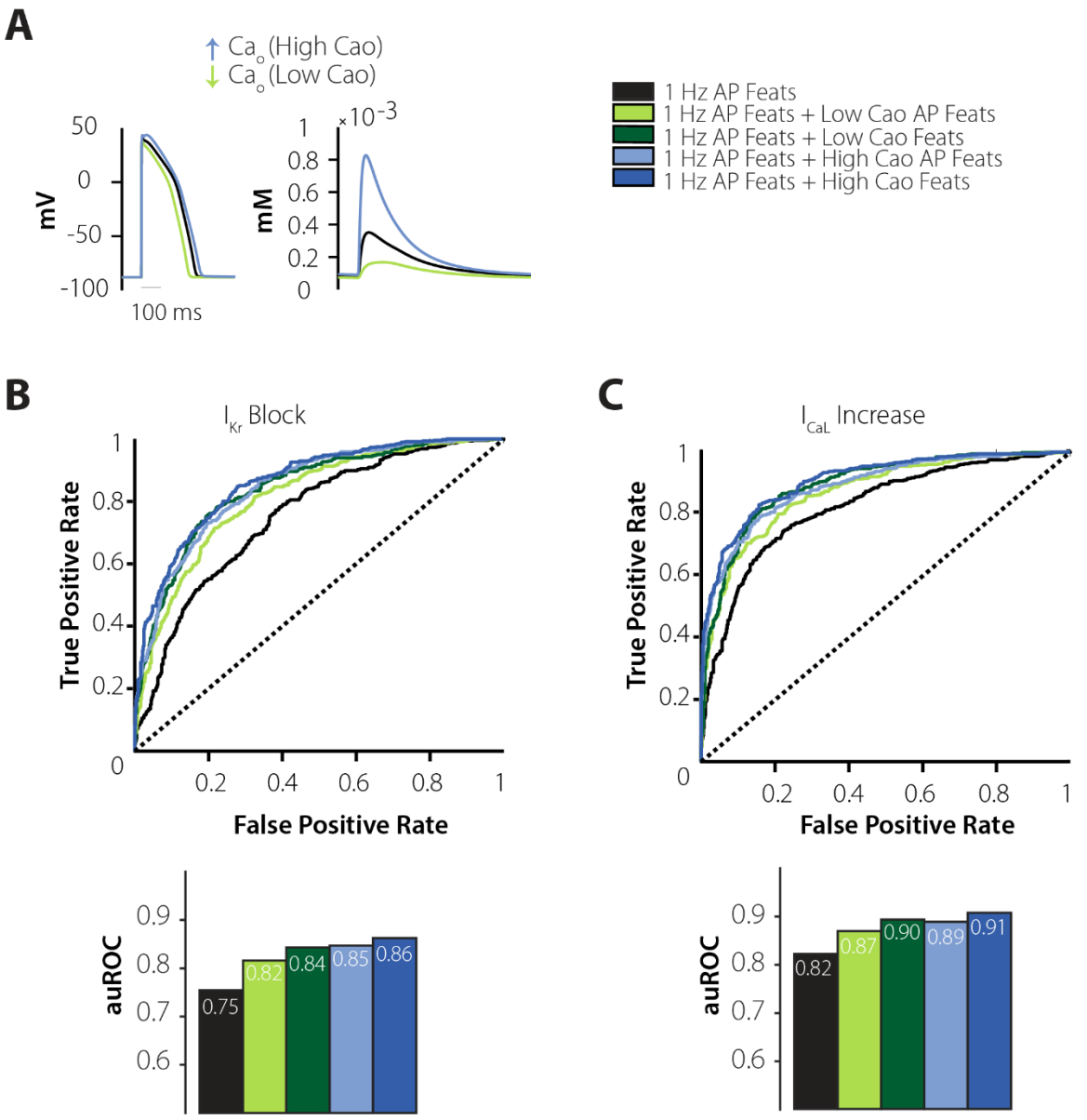


Fig. S2. Addition of calcium transients under varying calcemic conditions does not improve prediction of cellular response to an increase in I_{CaL}.

A.) Action potentials and calcium transients of cells under hyper (blue) and hypocalcemic (green) conditions. B-C.) Predictive performance does not change when adding calcium transient features under either hyper (dark green vs light green) or hypocalcemic conditions (dark blue vs light blue). Action potential waveform features are sufficient to achieve high performance (see Figure 5).

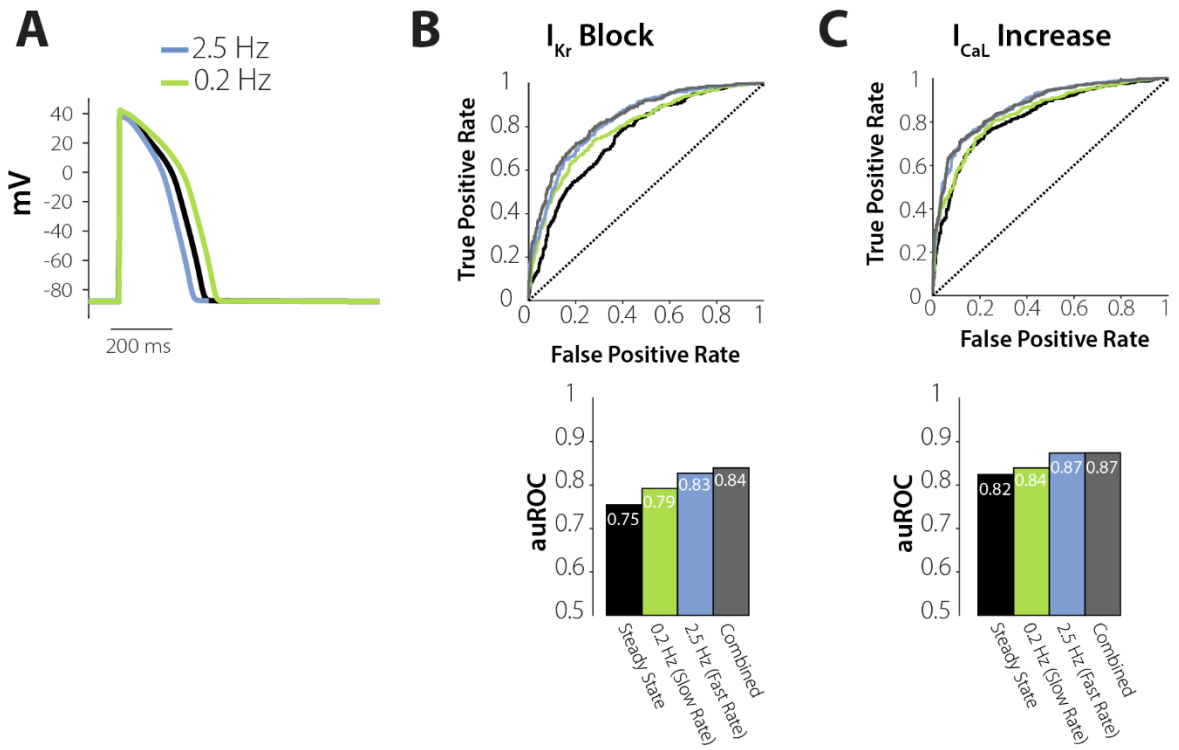


Fig. S3. Addition of features under different pacing protocols does not classify cells better than hypocalcemic and hypercalcemic conditions.

A.) Action potentials and calcium transients of cells paced at 2.5Hz (blue) and 0.2 Hz (green) conditions. B-C.) Predictive performance for both triggers (responses to I_{Kr} Block and Increase I_{CaL}) is weaker when provided features from the different pacing protocols. Features from the hypercalcemic experiment (Figure 5) provide more value than either the slow or fast paced features for both triggers.

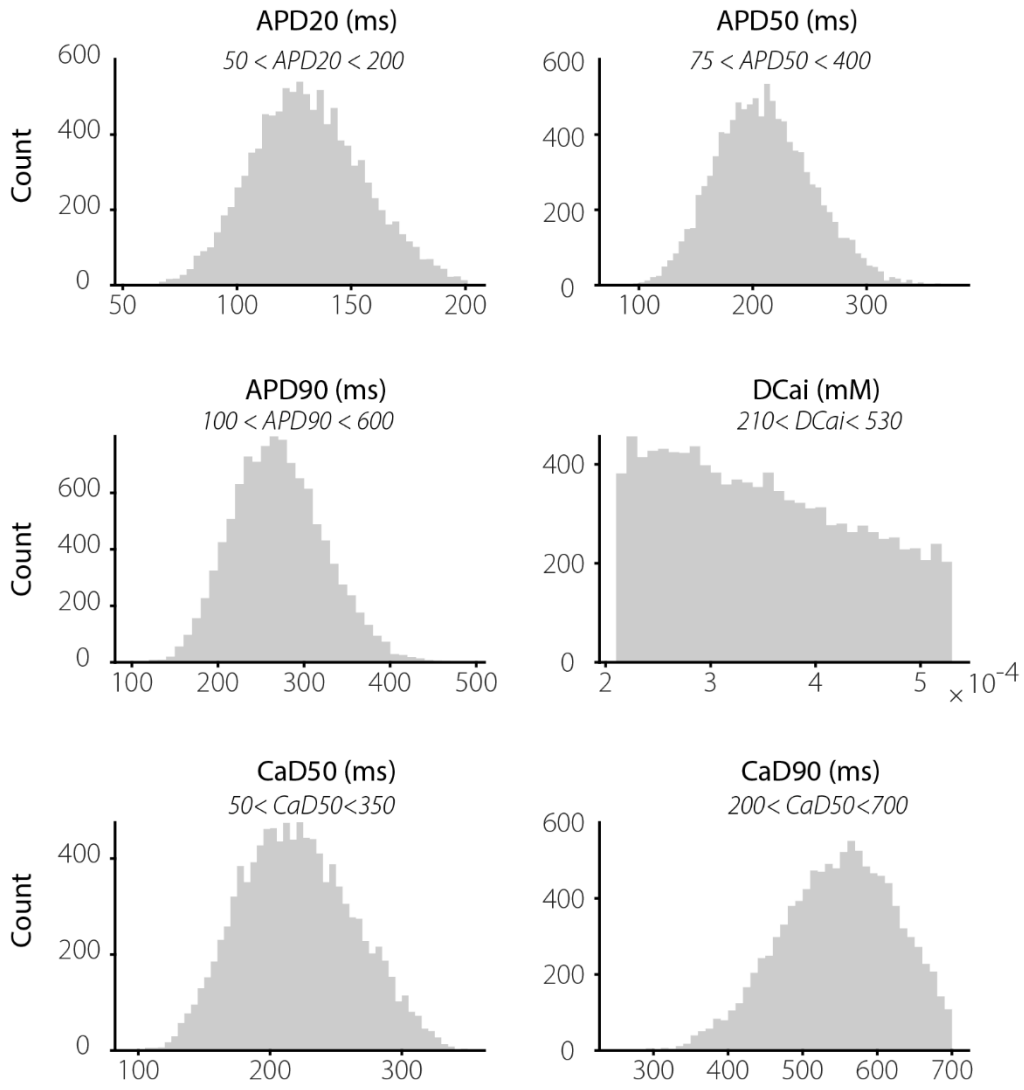


Fig. S4. Distribution of metrics used to experimentally calibrate the population

Range of metrics describing experimental recordings of non-diseased ventricular myocytes, as reported by Coppini et al(4) and implemented by Passini et al(5). Any cells that did not come within the experimental bounds of all 6 metrics were automatically discarded from the analysis.

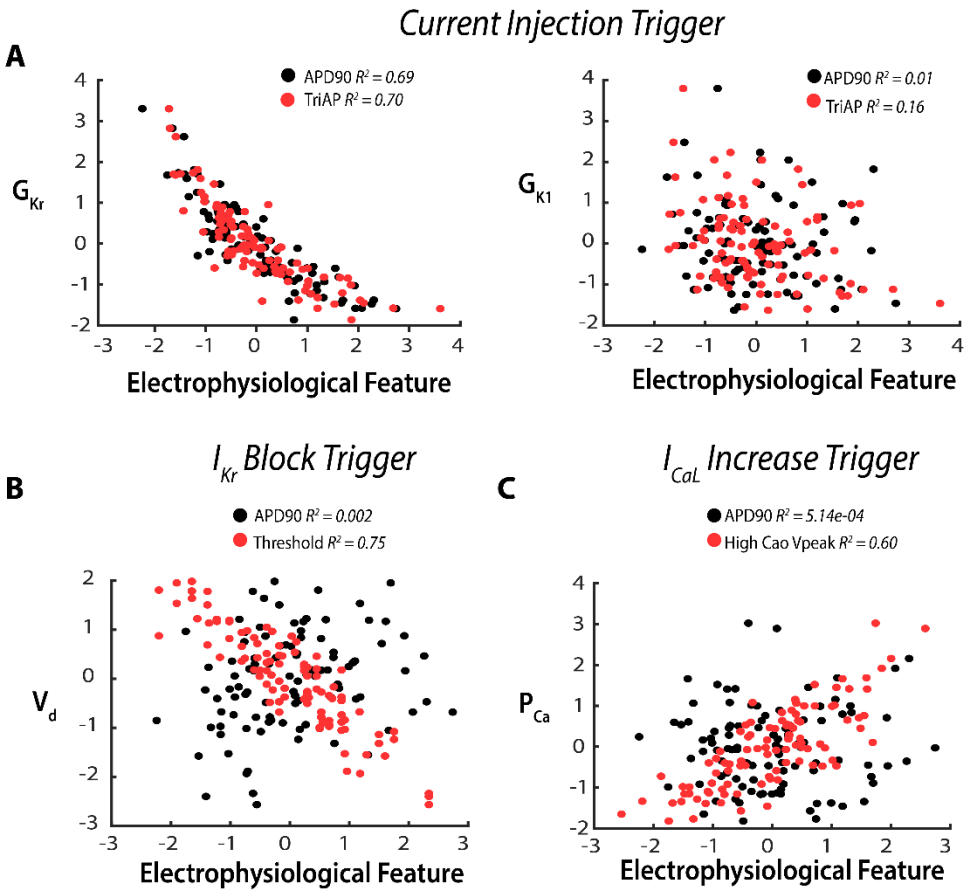


Fig. S5. Electrophysiological features correlate with important biological parameters.

Scatterplots show parameters that are important in determining susceptibility to particular triggers on the ordinate versus particular electrophysiological features on the abscissa, as indicated. Each variable is expressed as a z-score: i.e. the number of standard deviations away from the population mean.

In these plots, the correlations, or lack thereof, provide insight into why particular experimental measurements either improve ML performance or fail to do so. **(A)** For current injection (top plots), arrhythmia susceptibility is controlled almost entirely by rapid delayed rectifier conductance G_{Kr} and inward rectifier conductance G_{K1}. Both APD90 and AP triangulation correlate with G_{Kr} (left), but AP triangulation correlates more strongly with G_{K1}, explaining why AP triangulation is the most informative feature.

(B) For I_{Kr} Block (bottom left), the most important model parameter is V_d, the voltage dependence of activation of I_{CaL}. This parameter does not correlate with APD90, but does correlate strongly with the excitation threshold (under I_{Na} block conditions), which demonstrates why the excitation threshold improves ML performance so dramatically.

(C) For I_{CaL} Increase (bottom right), multiple parameters determine arrhythmia susceptibility, including the maximum I_{CaL} permeability (P_{Ca}). This parameter does not correlate with APD90 or with peak AP voltage under normal conditions (V_{peak}), but it does correlate with V_{peak} measured under hypercalcemic conditions (red symbols). This result helps to explain why the hypercalcemia experiment improves ML performance, and why multiple experimental protocols must be conducted to achieve excellent susceptibility prediction for this trigger.

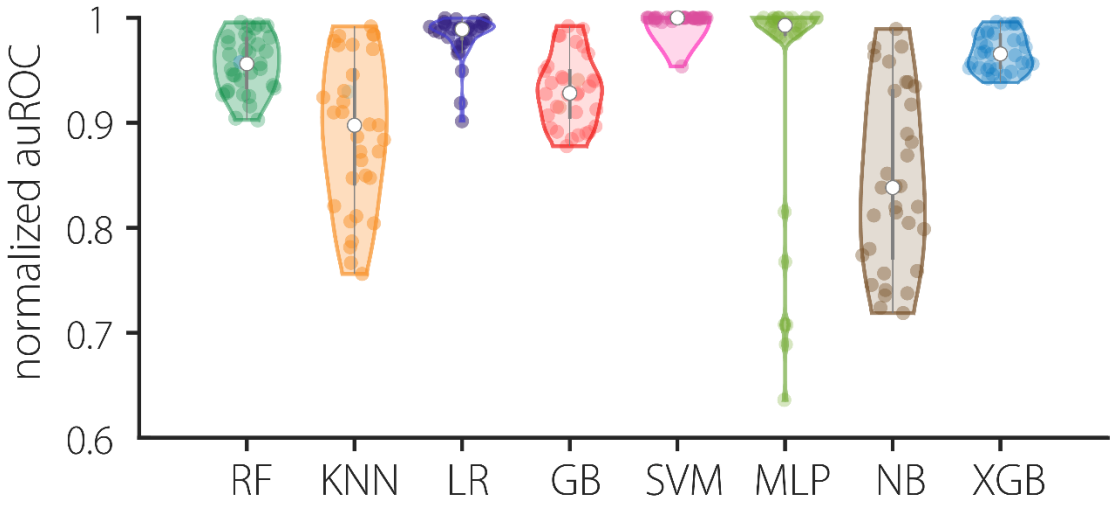


Fig. S6. Machine learning performance depends on algorithm choice.
For each machine learning task (n=33) tested throughout this paper, we compared how the eight different classifiers performed. To visualize this, we calculated the normalized area under the receiver operator curve (auROC) where we divided each auROC by the best classifier's auROC. Hence, a normalized auROC closer to 1 indicates the classifier performed similar to the best classifier. Here it is evident that SVM performed consistently well with MLP and LR in close second and third, respectively. In contrast, classifiers such as KNN and NB performed well for certain tasks and poorly for others.
RF=Random Forest; KNN=K-Nearest Neighbors; LR= Logistic Regression; GB=Gradient Boosting; SVM=Support Vector Machine; MLP= Multi-Layer Perceptron; NB=Naïve Bayes; XGB=XGBoost

Supplementary Tables

Table S1. Channel Conductance Parameters varied in O'Hara Model

Parameter	Definition	Baseline value[1]
$\overline{G}_{Na,fast}$	Maximal Na^+ conductance	75 mS/μF
$\overline{G}_{Na,late}$	Maximal late Na^+ conductance	0.0075 mS/μF
$\overline{G}_{to}$	Maximal transient outward K^+ conductance	0.02 mS/μF
$\overline{G}_{Kr}$	Rapid delayed rectifier K^+ conductance scaling factor [1]	0.046 mS/μF
$\overline{G}_{Ks}$	Slow delayed rectifier K^+ conductance scaling factor [2]	0.0034 mS/μF
$\overline{G}_{K1}$	Inward rectifier K^+ conductance scaling factor [1]	0.1908 mS/μF
$\overline{G}_{NaCa}$	Maximal Na^+-Ca^{2+} exchange current	0.0008 μA/μF
$\overline{G}_{Kb}$	Maximal conductance of background K^+	0.003 mS/μF
P_{Ca}	L-type Ca^{2+} current permeability	0.0001 cm/s
$\overline{I}_{NaK}$	Scales the Na^+-K^+ ATPase current	30
P_{Nab}	Background Na^+ current permeability	3.75e-10 cm/s
P_{Cab}	Background Ca^{2+} current permeability	2.5e-8 cm/s
$\overline{G}_{pCa}$	Maximal sarcolemmal Ca2+ pump current	0.0005 mS/uF
$SERCA_{total}$	SR Ca^{2+} release scaling factor [4]	1
RyR_{total}	SR Ca^{2+} uptake (SERCA) scaling factor[4]	1
$Trans_{total}$	NSR to JSR Ca^{2+} translocation[4]	1
$Leak_{total}$	Ca^{2+} Leak from the NSR[4]	1

Notes

[1]The scaling factors for I_{K1} and I_{Kr} are not formally maximal conductance, since each is multiplied by $\sqrt{K_o}/5.4$ and can therefore be greater than this value. Changing this factor scales the current at all values of extracellular [K^+] while maintaining the dependence on this variable.

[2]The scaling factor for I_{Ks} is multiplied by a function of intracellular [Ca^{2+}]. This value is therefore not precisely the current's maximal conductance.

[3]$\overline{I}_{NaK}$ was not labeled in the original paper and have been given these names to keep to terminology consistent.

[4]Parameters controlling the magnitude of these channels are an introduced unitless multiplier with baseline value equals to 1.00.

Table S2. Kinetic Parameters varied in O'Hara Model

Parameter	Definition	Baseline value
pm	Fast Na^+ current activation time constant	1
ph	Fast Na^+ current inactivation time constant	1
pj	Fast Na^+ current recovery from inactivation time constant	1
php	Fast Na^+ current CaMK modulated inactivation time constant	1
pjp	Fast Na^+ current CaMK modulated recovery from inactivation time constant	1
pmL	Late Na^+ current activation time constant	1
phL	Late Na^+ current inactivation time constant	1
phLp	Late Na^+ current CaMK modulated inactivation time constant	1
pa	Transient outward K^+ current activation time constant	1
pif	Transient outward K^+ current fast inactivation time constant	1
pis	Transient outward K^+ current slow inactivation time constant	1
pap	Transient outward K^+ current CaMK modulated activation time constant	1
pipf	Transient outward K^+ current CaMK modulated fast inactivation time constant	1
pips	Transient outward K^+ current CaMK modulated slow inactivation time constant	1
pd	L-type Ca^{2+} current activation time constant	1
pf	L-type Ca^{2+} current voltage-dependent inactivation time constant	1
pfcaf	L-type Ca^{2+} current calcium-dependent fast inactivation time constant	1
pfcas	L-type Ca^{2+} current calcium-dependent slow inactivation time constant	1
pjca	L-type Ca^{2+} current recovery from inactivation time constant	1
pfpf	L-type Ca^{2+} current CaMK modulated voltage dependent inactivation time constant	1
pfcapf	L-type Ca^{2+} current CaMK modulated calcium dependent inactivation time constant	1
pxrf	Rapid delayed rectifier K^+ current fast activation/deactivation time constant	1
pxrs	Rapid delayed rectifier K^+ current slow activation/deactivation time constant	1
pxs1	Slow delayed rectifier K^+ current activation time constant	1
pxs2	Slow delayed rectifier K^+ current deactivation time constant	1
pxk1	Inward rectifier K^+ current inactivation time constant	1
Vm	Fast Na^+ current activation voltage dependence	0
Vh	Fast Na^+ current inactivation voltage dependence	0
Vj	Fast Na^+ current recovery from inactivation voltage dependence	0
Vhp	Fast Na^+ current CaMK modulated inactivation voltage dependence	0

Vjp	Fast Na^+ current CaMK modulated recovery from inactivation voltage dependence	0
VmL	Late Na^+ current activation voltage dependence	0
VhL	Late Na^+ current inactivation voltage dependence	0
VhLp	Late Na^+ current CaMK modulated inactivation voltage dependence	0
Va	Transient outward K^+ current activation voltage dependence	0
Vi	Transient outward K^+ current fast inactivation voltage dependence	0
Vap	Transient outward K^+ current CaMK modulated activation voltage dependence	0
Vip	Transient outward K^+ current CaMK modulated inactivation voltage dependence	0
Vd	L-type Ca^{2+} current activation voltage dependence	0
Vf	L-type Ca^{2+} current voltage-dependent inactivation voltage dependence	0
Vfca	L-type Ca^{2+} current calcium-dependent inactivation voltage dependence	0
Vjca	L-type Ca^{2+} current recovery from inactivation voltage dependence	0
Vfp	L-type Ca^{2+} current CaMK modulated voltage dependent inactivation voltage dependence	0
Vfcap	L-type Ca^{2+} current CaMK modulated calcium dependent inactivation voltage dependence	0
Vxr	Rapid delayed rectifier K^+ current activation/deactivation voltage dependence	0
Vxs1	Slow delayed rectifier K^+ current activation voltage dependence	0
Vxs2	Slow delayed rectifier K^+ current activation voltage dependence	0
Vncx	Sodium-Calcium Exchanger voltage dependence	0
Vnak	Sodium-Potassium Pump voltage dependence	0

SI References

1. T. O'Hara, L. Virag, A. Varro, Y. Rudy, Simulation of the undiseased human cardiac ventricular action potential: model formulation and experimental validation. *PLoS Comput Biol* **7**, e1002061 (2011).
2. F. V. Pedregosa, G.; Gramfort, A.; Michel,V.; Thirion,B.; Grisel,O.; Blondel,M.; Prettenhofer,P.; Weiss,R.; Dubourg,V.; Vanderplas,J.; Passos,A.; Cournapeau,D.; Brucher,M.; Perrot,M.; Duchesnay, E., Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825-2830 (2011).
3. W. J. Youden, Index for rating diagnostic tests. *Cancer* **3**, 32-35 (1950).
4. R. Coppini *et al.*, Late sodium current inhibition reverses electromechanical dysfunction in human hypertrophic cardiomyopathy. *Circulation* **127**, 575-584 (2013).
5. E. Passini *et al.*, Mechanisms of pro-arrhythmic abnormalities in ventricular repolarisation and anti-arrhythmic therapies in human hypertrophic cardiomyopathy. *J Mol Cell Cardiol* **96**, 72-81 (2016).