Optimal Biomarkers Design for Drug Safety Evaluation Using Microelectrode Array Measurements

Eliott Tixier, Fabien Raphel, Damiano Lombardi, Jean-Frédéric Gerbeau

Optimal Biomarkers Design for Drug Safety Evaluation Using Micro-Electrode Array Measurements

Eliott Tixier [2,1], Fabien Raphel [1,2], Damiano Lombardi [1,2], Jean-Frédéric Gerbeau[1,2]

[1]Inria Paris, 75012 Paris, France

[2]Sorbonne Universités, UPMC Univ Paris 6, UMR 7598 LJLL, 75005 Paris, France

Abstract

The Micro-Electrode Array device enables high-throughput electrophysiology measurements that are less labour-intensive than patch-clamp based techniques. Combined with human-induced pluripotent stem cells (hiPSC), it represents a new and promising paradigm for automated and accurate in-vitro drug safety evaluation. In this article, the following question is addressed: which features of the MEA signals should be measured to better classify the effects of drugs? A framework for the classification of drugs using MEA measurements is proposed. It relies on an in silico electrophysiology model of the MEA, a feature selection algorithm and automatic classification tools. An in silico model of the MEA is developed and is used to generate synthetic measurements. An algorithm that extracts MEA measurements features designed to perform well in a classification context is described. These features are called numerical biomarkers. A state-of-the-art machine learning program is used to carry out the classification of drugs using MEA measurements. We show that the numerical biomarkers outperform the classical ones in different classification scenarios. We show that using both synthetic and experimental MEA measurements improves the robustness of the numerical biomarkers and that the classification scores are increased.

Keywords:

cardiac electrophysiology, numerical simulations, bidomain model, microelectrode array, classification, drug safety evaluation

Introduction

One of the main goals of safety pharmacology studies is to anticipate the effect of drugs on cardiomyocytes. Among other adverse effects, it focuses on predicting arrhythmic behaviors which may lead to torsades de pointes (TdP). The most common risk factors under consideration are QT prolongation and hERG block. However these risk factors are now considered insufficient and the guidelines need to be improved. Several advances in technology and computational modeling may favor the emergence of new methods for more efficient drug safety evaluation. On the hardware side, the Micro-Electrode Array (MEA) technology [21, 14] enables high-throughput electrophysiology measurements that are less labour-intensive than patch-clamp based techniques. On the biological side, the use of human-induced pluripotent stem cells (hiPSC) has

developed [20] and their recent large-scale production makes it a viable human model replacement. The combined use of the MEA technology and hiPSC represents a new and promising paradigm for automated and accurate in-vitro drug safety evaluation [8, 6]. In parallel of these technological breakthroughs, several efforts have been recently made towards promoting the use of computational tools in drug safety evaluation [9, 13]. In this context, a framework for drug safety evaluation using *in silico* models and experimental measurements using a MEA device is hereby presented.

The framework aims at predicting the effect of a drug onto the cardiomyocytes ionic channels activities from the knowledge of MEA experimental recordings. More precisely, the goal is to determine which ionic channels are affected by a given drug. The approach is based on an *in silico* model of the MEA and the cardiomyocytes tissue, a feature selection algorithm and a classification model. The *in silico* model is based on a simple ionic model [5] for the cardiomyocytes electrical activity and on the bidomain equations [23] for the spatial propagation of the electrical potentials. The ionic model counts three different currents (fast inward, slow inward, slow outward), each being associated with an ionic species (respectively sodium, calcium, potassium). The activity of each current is controlled by a scaling parameter that is referred to as conductance in the following. In the present work, the drugs considered are assumed to affect one of these currents. Thus, the inactivation of a current caused by a drug is modeled by a diminution of the corresponding conductance in the ionic model. The conductances and some other parameters of the model can be varied in order to replicate the variability observed in the experimental measurements. The *in silico* model is used to generate what is later referred to as synthetic MEA measurements. The experimental data set itself consists of MEA electrode recordings which come in the form of time series. Each recording is done in control conditions (no drug) and with different drug concentrations levels. The experimental data is also labelled, meaning the affected ionic channels are known for each drug.

As explained above, the MEA measurements, whether synthetic or experimental, come in the form of time series. For classification purposes, it is more efficient to extract features from these time series. Some features, also called biomarkers, are already widely used in the community such as the field potential duration which may be associated with the QT segment in ECGs. These common features are referred to as classical biomarkers. We propose a way to automatically extract features from the MEA measurements that are designed to perform well in a classification context. These new features, referred to as numerical biomarkers in the following, are defined as linear combinations of dictionary entries whose weights are solution of a sparse optimization problem. The weights are computed using MEA features coming from experimental measurements, synthetic ones or a composite set of both. To predict the effect of drugs, the idea developed in this work is to train a classification model, or classifier, to associate MEA measurements with a type of affected ionic channel, or label. Then, the classifier is tested with new MEA measurements for which it predicts labels. Provided that the true labels are known, it is possible to measure the precision of the classification and therefore evaluate a given classifier. In the present work, a state-of-the-art machine learning classification tool, Support Vector Classification (SVC), is used.

The paper is organized as follows. First, the methods are described. The *in silico* model is presented and the generation of synthetic data is explained. The algorithm that computes the numerical biomarkers is described and the classification tools are presented. Second, the performance of the numerical biomarkers and of the classification tools are studied in different drug classification scenarios. The numerical biomarkers are compared to the classical ones using two different classification strategies. Finally, numerical biomarkers computed with experimental data only and with a composite set of experimental and synthetic data are compared.

Methods

Equations

Bidomain equations and ionic model

Let Ω be the domain representing a MEA's well. The thickness of the layer of cells being small compared to the size of the well, the problem is assumed to be two-dimensional. The activation is assumed to be triggered by a current I_{app} that is applied in an arbitrary region of the well. We denote by A_{m}, C_{m}, z_{thick} the surface area of membrane per unit volume of tissue, the membrane capacitance, and the thickness of the cell layer, respectively. The intra and extra-cellular conductivity tensors σ_{i} and σ_{e} are assumed to be scalar. The propagation of the transmembrane potential V_{m} and the extracellular potential ϕ_{e} are modeled in Ω with the bidomain model [23]:

$$\begin{cases} A_{\text{m}}C_{\text{m}}\dfrac{\partial V_{\text{m}}}{\partial t} + A_{\text{m}}I_{\text{ion}}(V_m, w) - \nabla\cdot(\sigma_{\text{i}}\nabla V_{\text{m}}) - \nabla\cdot(\sigma_{\text{i}}\nabla\phi_{\text{e}}) = A_{\text{m}}I_{\text{app}}, \\ -\nabla\cdot((\sigma_{\text{i}}+\sigma_{\text{e}})\nabla\phi_{\text{e}}) - \nabla\cdot(\sigma_{\text{i}}\nabla V_{\text{m}}) = \frac{1}{z_{thick}}\displaystyle\sum_{e_k}\frac{I^k_{\text{el}}}{|e_k|}\chi_{e_k}. \end{cases} \tag{1}$$

In the second equation, I^k_{el} is the electric current which goes through the electrode located at e_k, $|e_k|$ is the electrode surface and χ_{e_k} is the characteristic function of e_k (which takes the value 1 on the electrode and 0 elsewhere). An imperfect model for the electrode is used to compute I^k_{el} and described in the Supplementary Material. Let $\boldsymbol{n}$ be the outward normal to the boundary of the domain Ω. Equations (1) are completed with the following boundary conditions: $\sigma_{\text{i}}\nabla\phi_{\text{i}}\cdot\boldsymbol{n} = 0$ (where $\phi_{\text{i}} = V_{\text{m}} + \phi_{\text{e}}$), and either $\phi_e = 0$ on the region connected to the ground or $\sigma_{\text{e}}\nabla\phi_{\text{e}}\cdot\boldsymbol{n} = 0$ elsewhere.

The transmembrane ionic current I_{ion} is described with the Minimal Ventricular (MV) model [5] which includes three currents: fast inward (fi), slow inward (si) and outward (so) currents. The reader is referred to the original publications for more details. Schematically, I_{ion} depends on V_{m} and on gating variables $w = (w_j)_{1\leq j\leq 3}$, solution of a system of three nonlinear ordinary differential equations. A conductance coefficient g_s, with $s = fi, so$ or si, controls the activity of the idealized channels associated with each of three currents of the model.

The partial differential equations are discretized in space by means of P1 finite elements, and in time by using backward differentiation formula (BDF) schemes with adaptive time steps and order provided by the Sundials library [11]. The quantity of interest is the extra-cellular potentiel, also referred to as field potential (FP). It is a function of time and recorded at the electrodes locations.

Synthetic measurements In the present work, the computational model is used to generate synthetic MEA measurements. For a given set of conductances, the model is evaluated and the electrodes FPs are recorded. The conductances are chosen as to represent meaningful scenarios, as explained later in the Results section. To mimic experimental measurements, a zero-mean Gaussian noise of standard deviation 10 μV is added to the FPs (see Figure 3). A heterogeneity model of some ionic parameters is also considered to replicate the variability exhibited by the experimental measurements. This model is described later in this section.

Steady-state regime Because the initial conditions of the ionic model do not correspond to those of a steady-state regime, several beats may need to be simulated before reaching a regime where there is negligible

beat-to-beat variations. A numerical experiment was carried out to determine when this regime is reached. Figure 2 shows super-imposed consecutive simulated FPs and the normalized beat-to-beat variations in the FP. When considering noisy synthetic measurements as described above, the steady-state is assumed to be reached when the beat-to-beat variations is comparable to variations induced by noise only. The beat-to-beat variability observed after this beat may be imputed to the coarseness of the mesh, the time discretization errors and the fluctuations of the ionic model itself. In the present work, the steady-state is assumed to be reached at the second beat. Therefore, the simulations are run for two cardiac cycles and the second beat is recorded to be used as a synthetic measurement.

Drug modeling

We chose to model the action of drugs on the ion channels by the conductance-block formulation of the pore block model [15, 24, 4]. This simple approach, which relies on a small number of parameters, was shown in [1] to be able to reproduce the expected effects of several drugs on MEA signals. The conductance of a given channel s is given by:

$$g_s = g_{control,s}\left[1 + \left(\frac{[D]}{IC_{50}}\right)^n\right]^{-1}, \tag{2}$$

where $g_{control,s}$ is the drug-free maximal conductance, $[D]$ is the drug concentration, IC_{50} is the value of the drug concentration at which the peak current is reduced of 50%, n is the Hill coefficient. In this work, n will be assumed to be equal to 1.

Heterogeneity modeling

A typical experimental MEA FP measurement exhibits both a depolarization spike and a repolarization wave (see Figure 3). Using the computational model described above, the repolarization wave is usually too small compared to what is observed in experiments. As noted in [1], the repolarization wave provided by this model is larger when the domain includes cells with different APDs. In [1], the cell heterogeneity was defined on a checkerboard arbitrarily chosen in the MEA's well. We propose here a different approach, based on a probabilistic description of the heterogeneity. The tissue is supposed to be a continuous mixture characterized by a space dependent coefficient:

$$\theta_w(x,y) = (1 - c(x,y))\theta_w^{(A)} + c(x,y)\theta_w^{(B)}, \tag{3}$$

where c is a random process with values in $[0,1]$, $\theta_w^{(A)}$ and $\theta_w^{(B)}$ are coefficients of the MV model characterizing two kinds of cells, called "type A" and "type B". In our simulations, we took $\theta_w^{(A)} = 0.1$, $\theta_w^{(B)} = 0.8$. The AP corresponding to different homogeneous realizations of c is shown in Figure 4. We make the hypothesis that the spatial variations of c are structured by a normal correlation function f_c:

$$f_c\left[\begin{pmatrix}x\\y\end{pmatrix}, \begin{pmatrix}x'\\y'\end{pmatrix}\right] = \exp\left[-\frac{(x-x')^2 + (y-y')^2}{2l_c^2}\right], \tag{4}$$

where l_c is the correlation length, set to $l_c = 0.25$ mm in the present work. To discretize the random process c, we compute the correlation matrix on the finite element mesh used for the discretization of the bidomain

equations. The correlation matrix $\mathbf{C} = [C_{i,j}] \in \mathbb{R}^{N_{\text{mesh}} \times N_{\text{mesh}}}$ reads:

$$C_{i,j} = f_c \left[\begin{pmatrix} \hat{x}_i \\ \hat{y}_i \end{pmatrix}, \begin{pmatrix} \hat{x}_j \\ \hat{y}_j \end{pmatrix} \right], \tag{5}$$

where N_{mesh} is the total number of mesh nodes and $(\hat{x}_i, \hat{y}_i)$ are the coordinates of the i^{th} node. The eigenpairs of $\mathbf{C}$ are denoted by (λ_i, Φ_i), and ordered by decreasing order of the eigenvalues λ_i. By a convenient abuse of notation, we denote by $(\hat{x}, \hat{y}) \to \Phi_i(\hat{x}, \hat{y})$ the function of the finite element space associated with the eigenmode Φ_i. Finally, the discretized heterogeneity field is approximated by the following truncated expansion:

$$c(\hat{x}, \hat{y}, \boldsymbol{\xi}) = \sum_{i=1}^{n_c} \xi_i \Phi_i(\hat{x}, \hat{y}) \tag{6}$$

where $\boldsymbol{\xi} = (\xi_i)_{i=1\dots n_c}$ is a random vector and n_c a truncation index chosen so that the truncation explains at least 99% of the variance. In other words, n_c is the smallest index n such that the following criterion is verified:

$$\frac{\sum_{i=1}^{n} \lambda_i}{\sum_{i=1}^{N_{\text{mesh}}} \lambda_i} > 0.99 \ . \tag{7}$$

Heterogeneity fields can now be generated simply by sampling the random variable $\boldsymbol{\xi}$. In the present work, $N_h = 128$ heterogeneity fields were generated by sampling $\boldsymbol{\xi}$ from an uncorrelated uniform distribution over $[-1, 1]^{n_c}$, and each sample is rescaled to range between 0 and 1. An example of heterogeneity field is presented in Figure 5.

Biomarkers

Biomarkers may be defined as quantities extracted from a signal that convey information about hidden quantities of interest. In our case, the biomarkers are features extracted from the MEA FP which would ideally provide information about the conductances of interest: g_{fi}, g_{so}, g_{si}. In this section, we present different choices of biomarkers to be used in a classification context.

"Classical" biomarkers

The MEA FP can be split into two regions of interest: the depolarization and the repolarization. The depolarization observed at one electrode corresponds to the local depolarization of the cardiomyocytes. The depolarization amplitude (DA) may be qualitatively linked to the AP upstroke velocity. This biomarker is commonly associated with the activity of the fast sodium channel (g_{fi} for the MV model). The repolarization amplitude (RA) may be qualitatively linked to some extent to the AP repolarization slope and to a bigger extent to spatial heterogeneities in AP durations. Once the depolarization and repolarization have been detected, it is possible to measure the FP duration (FPD), simply as the difference between the repolarization and depolarization times. Both biomarkers RA and FPD are associated with the activity of the potassium and calcium currents (g_{so} and g_{si} in the MV model). In Figure 3, a sample of FP with the corresponding classical biomarkers is shown. As explained above, each (real or numerical) experiment is performed both in a drug-block condition and in control condition. Because of the significant variability of measurements in MEA, it is important to consider the variations observed in the FP in drug block conditions with respect to the control conditions to isolate the effect of the drug from other sources of variability: tissue variability, stimulation protocol, *etc.* Therefore, as proposed in [18], the features of interest are the biomarkers in drug

block condition divided by the biomarkers in control conditions. For instance, the depolarization amplitude is actually the following ratio:

$$\mathrm{DA}_{\mathrm{ratio}} = \frac{\mathrm{DA}_{\mathrm{drug}}}{\mathrm{DA}_{\mathrm{control}}} \tag{8}$$

For the sake of clarity in the notation, the subscript "ratio" is omitted in the following and any biomarker actually refers to a ratio with the control value. For each MEA measurement, the FP is recorded at each of the nine electrodes. Again, the important variability in the measurements motivates the use of robust features. Since the behavior of the FP may greatly vary from one electrode to another, the median of the biomarkers over all electrodes is in practice a good choice of features. In the following, the set of biomarkers $\{\tilde{\mathrm{DA}}, \tilde{\mathrm{RA}}, \tilde{\mathrm{FPD}}\}$ is referred to as the classical biomarkers, where the $\tilde{}$ operator denotes the median over all nine electrodes.

Numerical biomarkers

The rationale behind the choice of biomarkers described above is only qualitative and oftentimes does not represent the best set of features in a classification context. Here, we adopt a more automatic strategy to select the best set of biomarkers for a given experimental scenario, as recently proposed in [10]. First, the set of features to be extracted from a given FP is enriched to build a dictionary of features. It is indeed possible to extract additional quantities from the FP other than DA, RA and FPD. We propose to compute also, for each electrode of the MEA, the following features: the area under curve of the repolarization wave (AUCr), the repolarization center (RC), the repolarization width (RW) and the FP notch (FPN). The details on how to compute these additional biomarkers are described in the Supplementary Material. Ratios of these quantities are also added to the dictionary of features: RA/DA, DA/RA, RA/FPD, FPD/RA, DA/FPD, FPD/DA, RA/RW, RW/RA. Each feature is actually a ratio with its control counterpart as described in (8). To include the information of all nine electrodes, the median (denoted by the $\tilde{}$ operator), mean (denoted by the $<>$ operator) and maximum values (denoted by a max subscript) over the electrodes are retained in the dictionary. This amounts to a total of $N_b = 38$ dictionary entries.

The purpose of the method described below is to associate each conductance with a numerical biomarker obtained by weighting the entries of the dictionary of features. The weights of such a combination are solution of an optimization problem. First, let us introduce some notation.

We denote by y_1 (resp. y_2, y_3) the numerical biomarker (to be determined) associated with g_{fi} (resp. g_{so}, g_{si}). From now on, the conductances (g_{fi}, g_{so}, g_{si}) are denoted by $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$. Each dictionary entry is considered as a function of $\boldsymbol{\theta}$. The numerical biomarkers are sought as a linear combination of the dictionary entries:

$$y_h(\boldsymbol{\theta}) = \sum_{j=1}^{N_b} w_j^{(h)} b_j(\boldsymbol{\theta}), \quad 1 \leq h \leq 3, \tag{9}$$

where the weights $\boldsymbol{w}^{(h)} = (w_j^{(h)}) \in \mathbb{R}^{N_b}$ are the unknowns of the problem. These weights are sought so that $y_h(\boldsymbol{\theta})$ is maximally correlated with θ_h and minimally correlated with θ_k, $\forall k \neq h$. This may be stated as follows:

$$\forall h \in \{1, \ldots, 3\}, \quad \begin{cases} \max\limits_{y_h} & \mathrm{cov}\left(y_h(\boldsymbol{\theta}), \theta_h\right) & \text{(10a)} \\ \min\limits_{y_h} & \left|\mathrm{cov}\left(y_h(\boldsymbol{\theta}), \theta_k\right)\right|, \quad \forall k \neq h & \text{(10b)} \\ \text{s.t.} & \mathrm{var}\left(y_h(\boldsymbol{\theta})\right) = 1 & \text{(10c)} \end{cases}$$

where $\mathrm{cov}(\cdot,\cdot)$ and $\mathrm{var}(\cdot)$ are respectively the covariance and variance operators. In the following, we assume that each component of $\boldsymbol{\theta}$ is a zero-mean unit-variance random variable. This is achieved in practice by a simple rescaling of the conductances samples. We also adopt the following notation:

$$\tilde{b}_j(\boldsymbol{\theta}) = b_j(\boldsymbol{\theta}) - \mathbb{E}\left[b_j(\boldsymbol{\theta})\right], \tag{11}$$

where $\mathbb{E}\left[\cdot\right]$ is the expectation operator. The problem may now be recast into an optimization problem where the cost function to be minimized reads:

$$\mathcal{J}(\boldsymbol{w}^{(h)}) = \mathcal{J}_C(\boldsymbol{w}^{(h)}) + \mathcal{J}_N(\boldsymbol{w}^{(h)}) + \mathcal{J}_P(\boldsymbol{w}^{(h)}), \tag{12}$$

where

$$\mathcal{J}_C(\boldsymbol{w}^{(h)}) = \frac{1}{2}\|\mathbf{C}\boldsymbol{w}^{(h)} - \mathbf{e}^{(h)}\|^2 \quad \text{where} \quad C_{kj} := \mathbb{E}(\theta_k \tilde{b}_j),\ e_k^{(h)} := \delta_{kh}, \tag{13a}$$

$$\mathcal{J}_N(\boldsymbol{w}^{(h)}) = \frac{\xi}{2}\left(\boldsymbol{w}^{(h)T}\mathbf{G}\boldsymbol{w}^{(h)} - 1\right)^2 \quad \text{where} \quad G_{ij} := \mathbb{E}(\tilde{b}_i \tilde{b}_j), \tag{13b}$$

$$\mathcal{J}_P(\boldsymbol{w}^{(h)}) = \frac{\lambda_h}{N_g}\|\boldsymbol{w}^{(h)}\|_1. \tag{13c}$$

Let us now explain each term of (13). $\mathcal{J}_C(\boldsymbol{w}^{(h)})$ corresponds to (10a) and (10b). It measures the discrepancy to the situation where $\mathrm{cov}\left(y_h(\boldsymbol{\theta}), \theta_h\right) = 1$ and $\mathrm{cov}\left(y_h(\boldsymbol{\theta}), \theta_k\right) = 0$, $\forall k \neq h$.
$\mathcal{J}_N(\boldsymbol{w}^{(h)})$ is a relaxation of the constraint in (10c). ξ is a regularization parameter that is set to 1 in practice. $\mathcal{J}_P(\boldsymbol{w}^{(h)})$ is a regularization term by penalization of the 1–norm of $\boldsymbol{w}^{(h)}$. ℓ_1 penalized cost functions tend to promote sparse solutions [22]. Sparse solutions for $\boldsymbol{w}^{(h)}$ are interesting in that they offer a more interpretable decomposition onto the dictionary entries (since most weights are zero) than what an ℓ_2 penalization would yield.

We now discretize the problem by considering N samples of the parameters $\boldsymbol{\theta}$ drawn over a parameter space $\Theta \subset \mathbb{R}^3$. The expectation operator is approximated using a quasi-Monte-Carlo quadrature rule and the cost function in (12) is minimized using a Nesterov accelerated gradient descent [16]. The Monte-Carlo samples may come from synthetic or experimental measurements. For synthetic measurements, the conductances are known, but this is not the case for experimental measurements. In that case, an approximation of these conductances is computed using Equation (2). Note that the solution weights depend strongly on the choice of samples used for the Monte-Carlo approximations.

An example of the obtained weights is shown in Figure 7. Interestingly, the classical biomarkers are still among the most weighted features. The correlation between the conductances of interest and the numerical biomarkers is compared to the correlation with the classical biomarkers in Figure 8. The correlation between two quantities u and v is defined as follows:

$$\mathrm{cor}(u, v) = \frac{\mathrm{cov}(u, v)}{\sqrt{\mathrm{var}(u)\mathrm{var}(v)}}. \tag{14}$$

As expected, each numerical biomarker is well correlated with its associated conductance whereas uncorrelated with the others. This is not the case for the classical biomarkers. The results in the next section show that such a choice of features improves the classification performance.

Experimental data set

The MEA considered in the present work is a 6-well MEA with nine electrodes per well. Its geometry as well as the corresponding finite element mesh is shown in Figure 1. The MEA measurements come in the form of FP recordings corresponding to the different electrodes of the different wells of the MEA.The MEA used is a 6-well MEA with nine electrodes per well. These recordings come in the form of time series where several cardiac cycles, or beats, are recorded. We extracted several beats on each electrode from each well of the MEA. Data were provided by Janssen Pharmaceutica NV using MC_Rack (Multi Channel Systems GmbH) and post-processed by NOTOCORD Systems (NOTOCORD-FPS 3.0 software). Cells cultures were developped by the CDI company (iCell Cardiomyocytes). As explained earlier the recordings were made in control conditions (no drug) and with different drugs at different concentrations levels. The drugs used for the present study are summarized in Table 2. Note that the diltiazem was recorded in two different wells (A and B) to compensate the scarcity of calcium-antagonist drugs. The experimental process consists in adding five times a compound at increasing concentrations in a given well. Thus, including the control condition record, we finally obtain field potentials for six contexts in each well. Equation (2) was used to obtain an approximation of the conductances values associated with the experimental measurements which are needed for the numerical biomarkers calculations. The Hill coefficients and IC_{50} values are given in the Supplementary Material of [12] and in [15]. Concerning the dictionary of features, a few adjustments need to be made in some cases. Indeed, it appears that at some high concentration levels of mexiletine, there is simply no action potential (because the sodium channels are too blocked) and therefore the field potential is a flat line. To take this into account, the values of dictionary entries are set to the ones at the last concentration where an action potential was observed. In addition, all features where DA is in the numerator position in a ratio are set to zero for this concentration.

Classification

Support Vector Classification

Support vector classification [3] (SVC) is an adaptation of the support vector machine (SVM) method in a classification setting. Classification generally consists in attributing labels to inputs. The available data set, comprising both inputs and labels, is generally split into a training set used to build the classifier and a validation set to test the classifier. The inputs are often multi-dimensional and in our case correspond to the biomarkers, whether classical or numerical. The labels are integers that represent the classes to which the inputs are assigned. These classes are mutually exclusive, meaning one sample can only belong to a single class. SVC belongs to the so-called supervised methods since the labels are known, at least for the training set. The main idea behind SVC is to maximize the margin between the inputs and the decision boundary [3]. In the linear case, the decision boundary is a hyperplane of the input space. In general however, this is not sufficient to properly separate the samples according to their classes. A common way to obtain more complex boundary decisions is to use a so-called "kernel trick" [19] which is based on a mapping from the input space to a higher-dimensional space where the existence of a separating hyperplane is more likely. In the present case, the labels are "sodium antagonist", "calcium antagonist" and "potassium antagonist", respectively associated with labels 0, 1 and 2. Among various possible choices of kernels, a Gaussian kernel is employed in this work.

We used a Python implementation of SVC through the Scikit-learn [17] machine learning library which itself uses the LIBSVM library [7]. For a given training set, a so-called classifier is built. The classifier is then

called to predict the labels of the validation set samples. The predictions are finally compared to the true labels. There exist several metrics to quantify the prediction quality. Two different metrics are considered here: the Cohen's kappa and the receiver operating characteristic area under curve (AUC). The Cohen's kappa is a single scalar designed to measure the performance of multi-class classifiers. Its value ranges from -1 (worst possible classifier) to 1 (perfect classifier), 0 corresponding to a coin-flip classifier. The AUC is defined for each class and measures how a classifier performs with respect to a given class. Its value ranges from 0 (worst) to 1 (best), 0.5 being a coin-flip. Because the classification is repeated several times with different data set splittings, the classification metrics are summarized using their means and standard deviations (see Table 3 for instance). The "averaged AUC" corresponds to the average of all AUCs (one AUC per class).

Both metrics are described in detail in the Supplementary Material. We now present two different strategies to employ SVC in the context of drug classification.

3-versus-3 classification Since there are three distinct classes in the experimental set, those three classes need to be included in the training set, preferably in equal proportions. The strategy of 3-versus-3 (3v3) classification consists in dividing the experimental set into a training set and validation set that both include samples from the three classes. Each class is divided into two sub-classes. This is naturally done for the sodium and potassium antagonist classes since they are each comprised of data from two different drugs. For the calcium antagonist class, the diltiazem data is artificially split into two drugs "diltiazem A" and "diltiazem B" (see Table 2). Each sub-class is associated with an identification number (ID) from 0 to 5. Therefore, there are 8 possible choices for the training and validation set combinations as summarized in Table 1.

One-versus-All classification The One-versus-All (OvA) classification strategy consists in training one classifier for each class. For each class j, the training set labels are modified to take the value 1 for samples in class j and 0 otherwise and a classifier is trained on this relabeled training set. In other words, the classifier for class j is only trained to recognize whether or not a sample belongs to class j. For the validation step, the classifiers do not predict a class label but a probability for a given sample to be in their respective class. Each sample of the validation step goes through each of the three classifiers and the predicted class corresponds to the classifier returning the highest probability. The splitting between training and validation sets is done in the same way as in the 3-vs-3 classification strategy.

Results

Comparison between classical and numerical biomarkers

Here the performance of the numerical biomarkers in a classification context is compared to that of the classical biomarkers for two different classification strategies. The data set is composed of 880 experiments, each counting one control measurement and 5 measurements at different drug concentration levels. For each experiment, the conductances values and FP features are computed as explained in the Methods section and the labels are defined according to Table 2.

3v3 classification

The performance of the numerical biomarkers compared to the classical ones is evaluated using the 3v3 classification strategy. The classification procedure is carried out for each different splitting of the data set as summarized in Table 1. First, the classification inputs are the 3 classical biomarkers for each drug concentration level:

$$\left\{\tilde{\mathrm{DA}}_{c1}, \tilde{\mathrm{RA}}_{c1}, \tilde{\mathrm{FPD}}_{c1}, \ldots, \tilde{\mathrm{DA}}_{c5}, \tilde{\mathrm{RA}}_{c5}, \tilde{\mathrm{FPD}}_{c5}\right\}, \tag{15}$$

where c_k is the k-th concentration level. The inputs are therefore of dimension 15.
Then, the classification inputs are the numerical biomarkers for each concentration, computed as explained in the Methods section using the classification training set as samples for the Monte-Carlo approximations. The inputs now read:

$$\left\{y_{1,c1}, y_{2,c1}, y_{3,c_1}, \ldots, y_{1,c5}, y_{2,c5}, y_{3,c_5}\right\}. \tag{16}$$

Note that for each splitting of the data set, new weights for the numerical biomarkers are computed. The classification procedure is carried out in both cases and the results are summarized in Table 3. Regardless of the chosen classification score, the results are better using the numerical biomarkers as inputs.

OvA classification

The same procedure as in the 3v3 case is applied to the OvA strategy. The classification procedure is carried out with both classical and numerical biomarkers as inputs and the results are summarized in Table 4. Again, the classification results are better using the numerical biomarkers as input, regardless of the classification score considered. Furthermore, the results are overall better when using the OvA approach rather than the 3v3 one.

Using combined experimental and synthetic measurements for the numerical biomarkers computation

Having established that numerical biomarkers outperform classical ones in two different classification scenarios, we now investigate the addition of synthetic measurements for the computation of the numerical biomarkers weights. To enrich the set of experimental samples used to compute the numerical biomarkers, a set of synthetic measurements is built. First, conductances samples are chosen to mimic the effect of drugs as shown in Figure 6. Depending on the most affected conductance, these samples are associated to a synthetic sodium (resp. calcium and potassium) antagonist drug called "synth A" (resp. B and C). 775 samples per drug are chosen which amounts to 155 experiments per drug. and their repartition is summarized in Table 2. This approximately corresponds to a 50%/50% split between experimental and synthetic measurements. For each conductances sample, the computational model described in the Methods section is evaluated and the dictionary features are computed from the simulated FPs. For each experiment, the computational model is also evaluated in the control conditions, *i.e.* with $g_{fi} = g_{si} = g_{so} = 1$ in order to compute the ratios as defined in (8). The features are incorporated in the experimental set to create a composite set. This composite set is then used to compute the numerical biomarkers weights. The same data set splitting procedure as described before is carried out. Note that the synthetic measurements are only used for the numerical biomarkers computation and are not included neither in the training set nor in the validation set. Again, two classification strategies are explored.

Classification results

The classification is carried out using both 3v3 and OvA approaches. The results are summarized in Tables 5 and 6. The addition of synthetic measurements to compute the numerical biomarkers improves the classifier predictions in both cases. In the 3v3 case, the improvement is more important than in the OvA case.

Discussion

In this study, a framework for an automatic classification of drugs from MEA measurements has been presented. The framework relies on an *in silico* model of a MEA device, on a feature selection algorithm and on state-of-the-art machine learning tools. The *in silico* model is a PDE model (the bidomain equations) coupled with a ionic model that describes the transmembrane current of the cardiomyocytes. The ionic model is a phenomenological model consisting of a set of coupled non-linear ODEs. The feature selection algorithm proposes a way to compute a so-called numerical biomarker for each conductance of interest, designed to perform better in a classification context than classical biomarkers. The numerical biomarkers are linear combinations of the entries of a dictionary of features which is given. The calculation of the weights involves Monte-Carlo approximations which use experimental or synthetic (or both) conductances and FP samples. It has been applied to drug classification problems using experimental MEA recordings. The classification was carried out using the Scikit-Learn Python library [17] which includes several classification tools. In the present work a Support Vector Classification was used. The data used for the classification consist in FP features extracted from experimental measurements and their associated labels corresponding to the type of drug that is considered.

The purpose of the present work is twofold. First, it intends to establish that the classically used biomarkers may be improved, at least in a classification context, by using numerical biomarkers instead. Second, it intends to show that the classification performance may benefit from the addition of synthetic measurements in the calculation of the numerical biomarkers. More generally, the authors intend to show that numerical simulations are useful to cardiac electrophysiology in general, beyond the sole scope of drug classification.

First, a comparison between classical and numerical biomarkers was carried out. The comparison consists in classifying drugs from experimental measurements using two different strategies: 3v3 and OvA. For each strategy, the classification is performed using classical or numerical biomarkers as inputs. As expected, the classification results in both cases are improved when using the numerical biomarkers. The latter were indeed designed to be maximally correlated to their associated conductance and minimally correlated to the others. As a consequence, they are more revealing of the underlying conductances than the classical biomarkers. In the 3v3 case, the mean AUCs are increased when using the numerical biomarkers as inputs and the standard deviations remain similar. The mean Cohen's kappa also significantly increases. However, its standard deviation is also higher but not enough to compromise the improvement of its mean. The same observations can be made for the OvA case except for the g_{fi} AUC which is reduced when using numerical biomarkers.

Second, the use of combined experimental synthetic measurements to compute numerical biomarkers is investigated. The numerical biomarkers are computed using Monte-Carlo approximations that require conductances and FP features samples. In the previous case, these samples are experimental. The idea is to improve the robustness of the numerical biomarkers by incorporating synthetic measurements that span better the parameters (*i.e.* conductances) space. This approach is meant to compensate the scarcity of experimental data and more generally the fact that experiments do not cover every possible drug block scenario. Conductances samples were drawn and the computational model was evaluated to generate noisy FPs. From

these FPs, the entries of the dictionary of features were computed. The numerical biomarkers weights are then computed using a composite set of experimental and synthetic samples. These numerical biomarkers are compared to the ones computed using only experimental data. The same two classification strategies as before are used to compare both approaches. In the 3v3 case, the improvement is not significant, both for the AUCs and the Cohen's kappa. This relatively low improvement needs to be mitigated by the fact that the classification scores were already high when using experimentally derived numerical biomarkers, especially for the AUCs. The improvements for the OvA case are however important. These results suggest that, for the classification scenarios envisioned here, the addition of synthetic measurements is always beneficial. Note also that the OvA strategy clearly outperforms 3v3 when using composite numerical biomarkers whereas it is not clear which strategy is the best when using numerical biomarkers computed from experimental measurements only.

The use of FP features in a classification context is now discussed. In classification problems, and in machine learning in general, a large number of inputs tend to provoke an over-fitting of the model. This means that the classifier tends to have satisfactory training scores but generalizes poorly on a validation test. This is in part solved by the regularization used but the number of inputs still remains important. When dealing with experimentally recorded FPs, the different signals are often not perfectly synchronized, making timestep-wise comparisons meaningless. Furthermore, an important variability of the signal amplitudes is observed in practice, making even perfectly synchronized signals difficult to compare. Using features extracted from the FP that are do not depend on time shifts and amplitude variations are therefore more robust in a classification context.

The limitations of the proposed approach are now discussed. First, the conductances values associated with the experimental measurements are not knwown and are therefore approximated using Equation (2). This approximation is however subject to several sources of uncertainty such as the IC_{50} whose value for a given drug may vary according to the source considered [12, 15]. The uncertainties also come from the Hill's equation which is an imperfect model. Knowing the exact values for the conductances is however not critical since those values are only needed to derive the numerical biomarkers and are not directly used dusing the classification procedure. Another limitation comes from the computational model used in the present work. The sources of error are multiple: ionic model error, space and time dicretizations, conductivities errors, etc. These errors are not critical either since the computational model is only used to compute the numerical biomarkers weights. This study shows that, despite the modeling errors, adding synthetic measurements simulated by the computational model leads to better classification results. Other limitations come from the classification strategies. Both classification strategies are non-exhaustive in that they do not explore every possible way of splitting the data set. Furthermore, the classification metrics used to compare the different approaches are not flawless. In some cases comparing AUCs for instance is not the best way to compare classifiers [2]. Other metrics exist, such as the mean squared error, but were not investigated in this work. Finally, the numerical biomarkers derived in the present work are not optimal in the sense that their correlation with their associated conductances is not one, as seen in Figure 8.

We now discuss some perspectives that could lead to interesting future works. Other classification methods than SVC exist, such as neural networks or random forests for instance. It would be interesting to assess whether the findings of this work are still valid when considering other classification tools. It would also be interesting to evaluate which classification tool generally performs best in the present drug classification context. Other perspectives concern the numerical biomarkers computed using a composite set of synthetic and experimental measurements. In the present work, the composite set is roughly composed of half synthetic

and half experimental measurements. However, other proportions could be investigated and an optimal proportion with respect to the classification score could be found. In the present work, only sodium, potassium and calcium antagonists drugs are considered but other types of drugs exist. Drugs that affect other ionic channels or even simultaneously several of them could be investigated. In parallel, more sophisticated ionic models including more current types would need to be used to model these new drugs. This would of course come at the expense or more computationally intensive simulations. Finally, training classifiers with only synthetic measurements instead of experimental ones could be considered. This would be very useful when experimental data are insufficient or even not available. The classifiers could also be trained with a composite set of synthetic and experimental data just like it is done in this work for the computation of numerical biomarkers.

References

[1] E. Abbate, M. Boulakia, Y. Coudière, J-F. Gerbeau, P. Zitoun, and N. Zemzemi. In silico assessment of the effects of various compounds in MEA/hiPSC-CM assays: Modelling and numerical simulations. Research report, Inria, https://hal.inria.fr/hal-01562673, July 2017.

[2] Niall M. Adams and David J. Hand. Improving the practice of classifier performance assessment. *Neural computation*, 12(2):305–311, 2000.

[3] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM, 1992.

[4] D. Bottino, R. C. Penland, A. Stamps, M. Traebert, B. Dumotier, A. Georgieva, G. Helmlinger, and G. S. Lett. Preclinical cardiac safety assessment of pharmaceutical compounds using an integrated systems-based computer model of the heart. *Progress in biophysics and molecular biology*, 90(1):414–443, 2006.

[5] Alfonso Bueno-Orovio, Elizabeth M Cherry, and Flavio H Fenton. Minimal model for human ventricular action potentials in tissue. *Journal of theoretical biology*, 253(3):544–560, 2008.

[6] I. Cavero, J-M. Guillon, V. Ballet, M. Clements, J-F. Gerbeau, and H. Holzgrefe. Comprehensive in vitro proarrhythmia assay (cipa): Pending issues for successful validation and implementation. *Journal of pharmacological and toxicological methods*, 81:21–36, 2016.

[7] Chih-Chung Chang and Chih-Jen Lin. Libsvm: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, 2(3):27, 2011.

[8] Mike Clements and Nick Thomas. High-throughput multi-parameter profiling of electrophysiological drug effects in human embryonic stem cell derived cardiomyocytes using multi-electrode arrays. *Toxicological Sciences*, 140(2):445–461, 2014.

[9] Mark R Davies, Ken Wang, Gary R Mirams, Antonello Caruso, Denis Noble, Antje Walz, Thierry Lavé, Franz Schuler, Thomas Singer, and Liudmila Polonchuk. Recent developments in using mechanistic cardiac modelling for drug safety evaluation. *Drug Discovery Today*, 2016.

[10] J-F. Gerbeau, D. Lombardi, and E. Tixier. How to choose biomarkers in view of parameter estimation. submitted for publication, 2017.

[11] Alan C Hindmarsh, Peter N Brown, Keith E Grant, Steven L Lee, Radu Serban, Dan E Shumaker, and Carol S Woodward. Sundials: Suite of nonlinear and differential/algebraic equation solvers. *ACM Transactions on Mathematical Software (TOMS)*, 31(3):363–396, 2005.

[12] James Kramer, Carlos A Obejero-Paz, Glenn Myatt, Yuri A Kuryshev, Andrew Bruening-Wright, Joseph S Verducci, and Arthur M Brown. Mice models: superior to the herg model in predicting torsade de pointes. *Scientific reports*, 3, 2013.

[13] M Cummins Lancaster and EA Sobie. Improved prediction of drug-induced torsades de pointes through simulations of dynamics and machine learning algorithms. *Clinical Pharmacology & Therapeutics*, 100(4):371–379, 2016.

[14] Thomas Meyer, Karl-Heinz Boven, Elke Günther, and Michael Fejtl. Micro-electrode arrays in cardiac safety pharmacology. *Drug Safety*, 27(11):763–772, 2004.

[15] G. Mirams, Y. Cui, A. Sher, M. Fink, J. Cooper, B. Heath, N. McMahon, D. Gavaghan, and D. Noble. Simulation of multiple ion channel block provides improved early prediction of compounds clinical torsadogenic risk. *Cardiovascular research*, 91(1):53–61, 2011.

[16] Brendan O'Donoghue and Emmanuel Candes. Adaptive restart for accelerated gradient schemes. *Foundations of computational mathematics*, 15(3):715–732, 2015.

[17] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[18] F. Raphel, M. Boulakia, C. Zemzemi, Y. Coudière, J-M. Guillon, P. Zitoun, and J-F. Gerbeau. Identification of ion currents components generating field potential recorded in MEA from hiPSC-CM. Preprint available at https://hal.inria.fr/hal-01570341, 2017.

[19] Bernhard Schölkopf and Alexander J Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.

[20] Clay W Scott, Matthew F Peters, and Yvonne P Dragan. Human induced pluripotent stem cells and their use in drug discovery for toxicity testing. *Toxicology letters*, 219(1):49–58, 2013.

[21] Multichannel Systems. Microelectrode array (mea) manual. `http://www.multichannelsystems.com/sites/multichannelsystems.com/files/documents/manuals/MEA_Manual.pdf`, -.

[22] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[23] L. Tung. *A bi-domain model for describing ischemic myocardial D–C potentials*. PhD thesis, MIT, 1978.

[24] N. Zemzemi, M. Bernabeu, J. Saiz, J. Cooper, P. Pathmanathan, G. Mirams, J. Pitt-Francis, and B. Rodriguez. Computational assessment of drug-induced effects on the electrocardiogram: from ion channel to body surface potentials. *British journal of pharmacology*, 168(3):718–733, 2013.

Conflict of Interest Statement

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

Author Contributions

All authors contributed equally to this work.

Funding

ET was funded by a doctoral grant from the French Ministry of Research and Higher Education. FR was funded by Instem. This work was partially supported by the Agency for Interaction in Mathematics with Business and Society (AMIES).

Acknowledgments

We would like to thank Janssen Pharmaceutica NV for providing us with the raw experimental data and Philippe Zitoun for many fruitful discussions.

Tables

Splitting index	0	1	2	3	4	5	6	7
training set IDs	$\{0,2,4\}$	$\{0,2,5\}$	$\{0,3,4\}$	$\{0,3,5\}$	$\{1,2,4\}$	$\{1,2,5\}$	$\{1,3,4\}$	$\{1,3,5\}$
validation set IDs	$\{1,3,5\}$	$\{1,3,4\}$	$\{1,2,5\}$	$\{1,2,4\}$	$\{0,3,5\}$	$\{0,3,4\}$	$\{0,2,5\}$	$\{0,2,4\}$

Table 1: Different possible splittings of the experimental data set.

Drug name	Blocked ionic channel	Associated conductance	ID	SVC class label	# experiments
Mexiletine	sodium	g_{fi}	0	0	160
Flecainide	sodium	g_{fi}	1	0	120
Diltiazem A	calcium	g_{si}	2	1	160
Diltiazem B	calcium	g_{si}	3	1	160
Moxifloxacin	potassium	g_{so}	4	2	120
Dofetilide	potassium	g_{so}	5	2	160
synth. A	sodium	g_{fi}	6	0	155
synth. B	calcium	g_{si}	7	1	155
synth. C	potassium	g_{so}	8	2	155

Table 2: Repartition of the available (experimental and synthetic) data set.

	classical biomarkers		numerical biomarkers	
Score	mean	std.	mean	std.
Cohen's kappa	0.18	0.15	0.51	0.27
g_{fi} AUC	0.65	0.10	0.86	0.11
g_{si} AUC	0.92	0.09	1.00	0.00
g_{so} AUC	0.36	0.11	0.82	0.10
averaged AUC	0.65	-	0.89	-

Table 3: Comparison between classical and numerical biomarkers with the 3v3 classification strategy.

	classical biomarkers		numerical biomarkers	
Score	mean	std.	mean	std.
Cohen's kappa	0.18	0.14	0.53	0.25
g_{fi} AUC	0.66	0.11	0.59	0.41
g_{si} AUC	0.92	0.11	0.96	0.06
g_{so} AUC	0.43	0.11	0.85	0.13
averaged AUC	0.67	-	0.80	-

Table 4: Comparison between classical and numerical biomarkers. Classification scores in the one-vs-all scenario.

	experiments only		experiments + synthetic	
Score	mean	std.	mean	std.
Cohen's kappa	0.51	0.27	0.54	0.06
g_{fi} AUC	0.86	0.11	0.87	0.13
g_{si} AUC	1.00	0.00	1.00	0.00
g_{so} AUC	0.82	0.10	0.84	0.08
averaged AUC	0.89	-	0.91	-

Table 5: Comparison between numerical biomarkers computed from experiments only and combined experiments and synthetic measurements. 3v3 classification strategy.

	experiments only		experiments + synthetic	
Score	mean	std.	mean	std.
Cohen's kappa	0.53	0.25	0.69	0.28
g_{fi} AUC	0.59	0.41	0.87	0.32
g_{si} AUC	0.96	0.06	1.00	0.01
g_{so} AUC	0.85	0.13	0.89	0.15
averaged AUC	0.80	-	0.92	-

Table 6: Comparison between numerical biomarkers computed from experiments only and combined experiments and synthetic measurements. OvA classification strategy

Figure captions

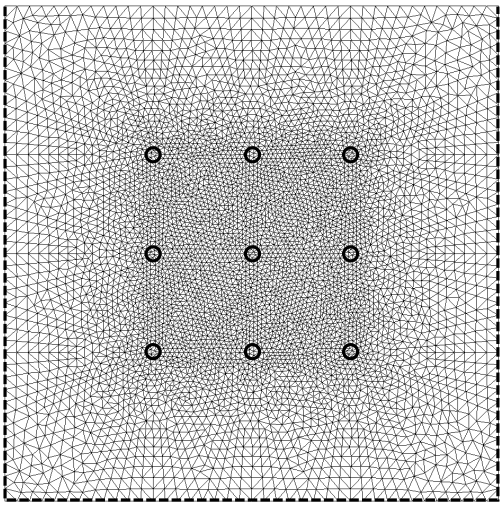

Figure 1: MEA geometry and its corresponding finite element mesh. The circles represent the locations of the nine electrodes.

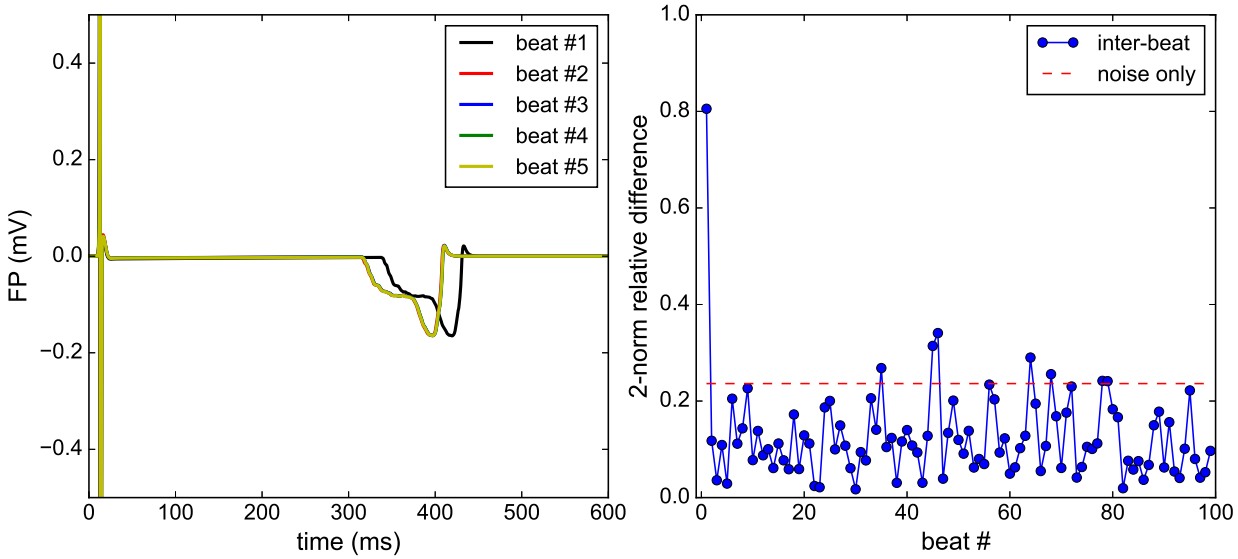

Figure 2: Steady-state analysis: the Bidomain equations are solved for 100 consecutive beats. Qualitatively, a satisfactory steady state is reached at the second beat (left). The beat-to-beat relative difference of the FP is monitored (right) and is to be compared to the relative difference between two identical solutions, each polluted by an independent noise (right).

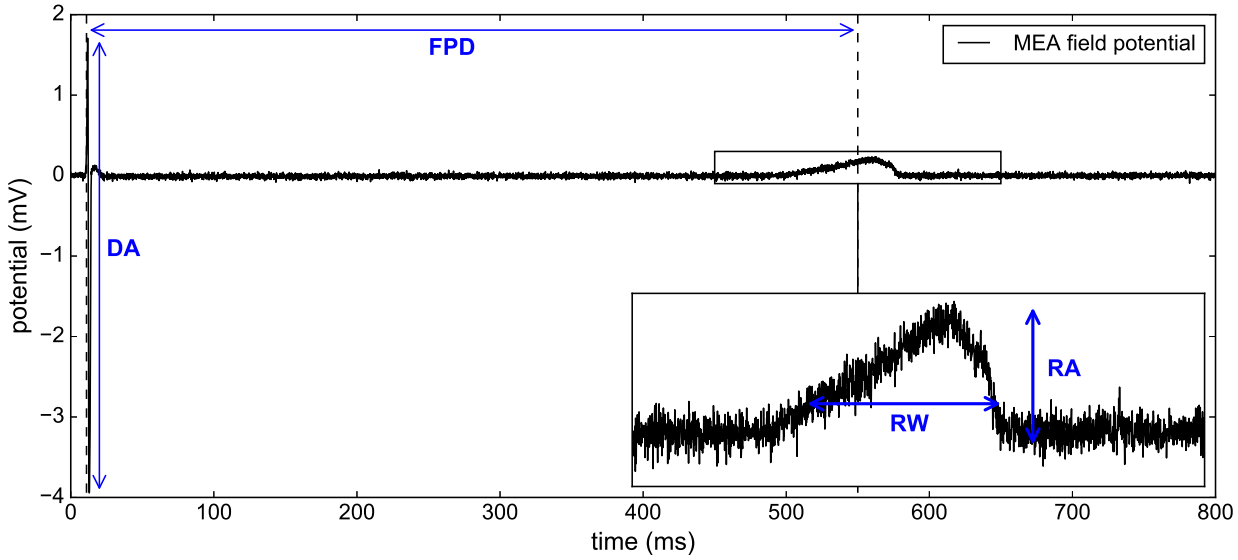

Figure 3: Synthetic MEA field potential and some associated biomarkers.

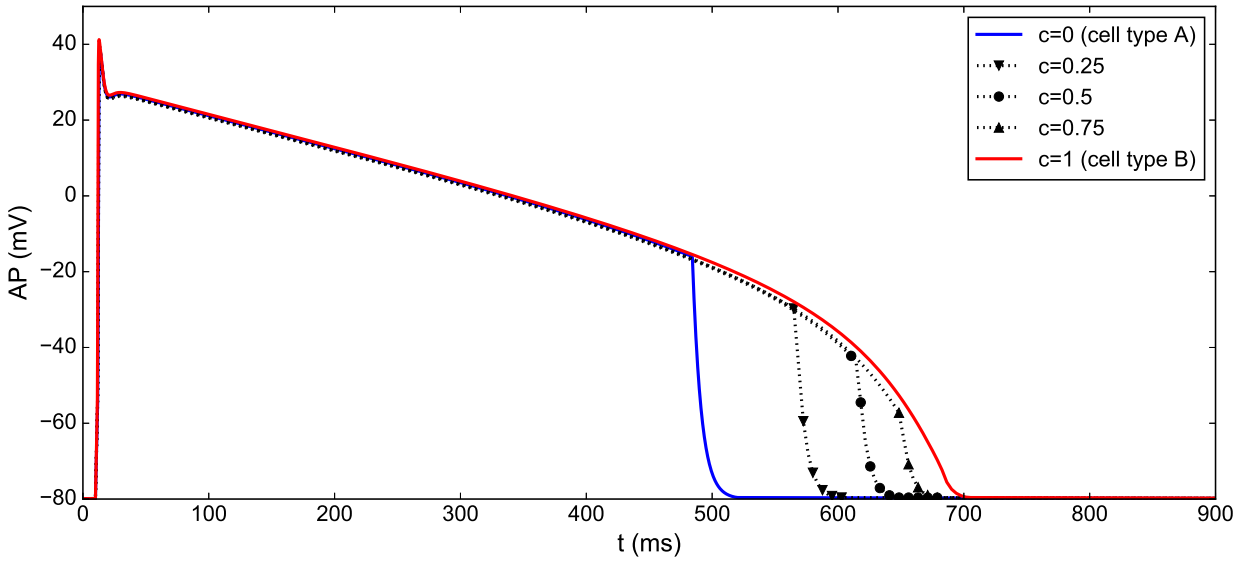

Figure 4: Heterogeneity modeling: different APs obtained by simulating the MV model with different values of the heterogeneity parameter c.

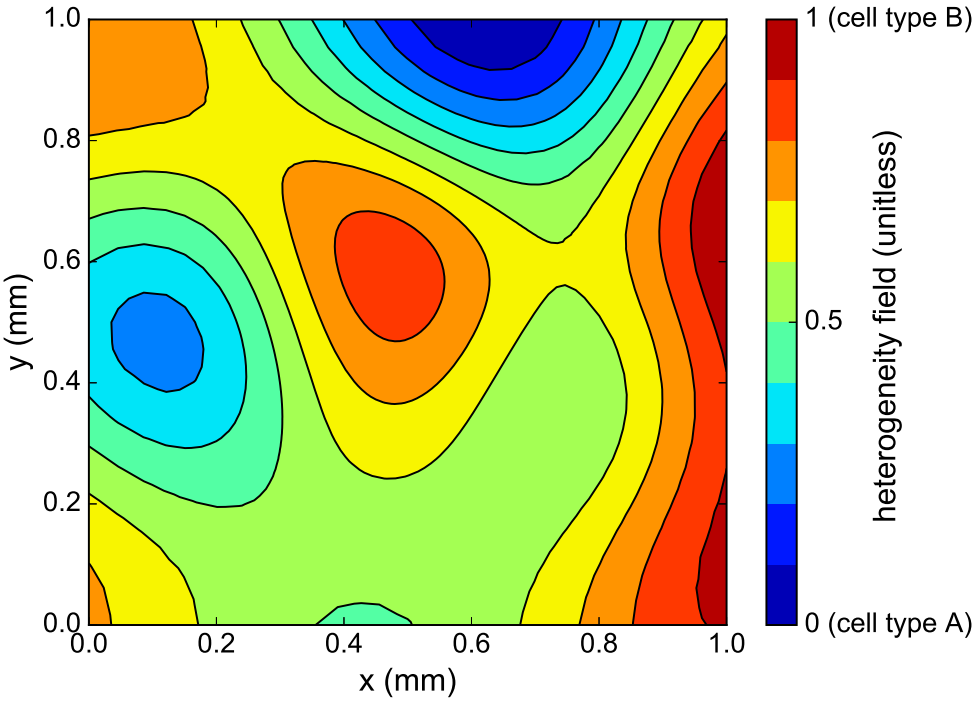

Figure 5: One sample of cell heterogeneity field generated using the correlation matrix method.

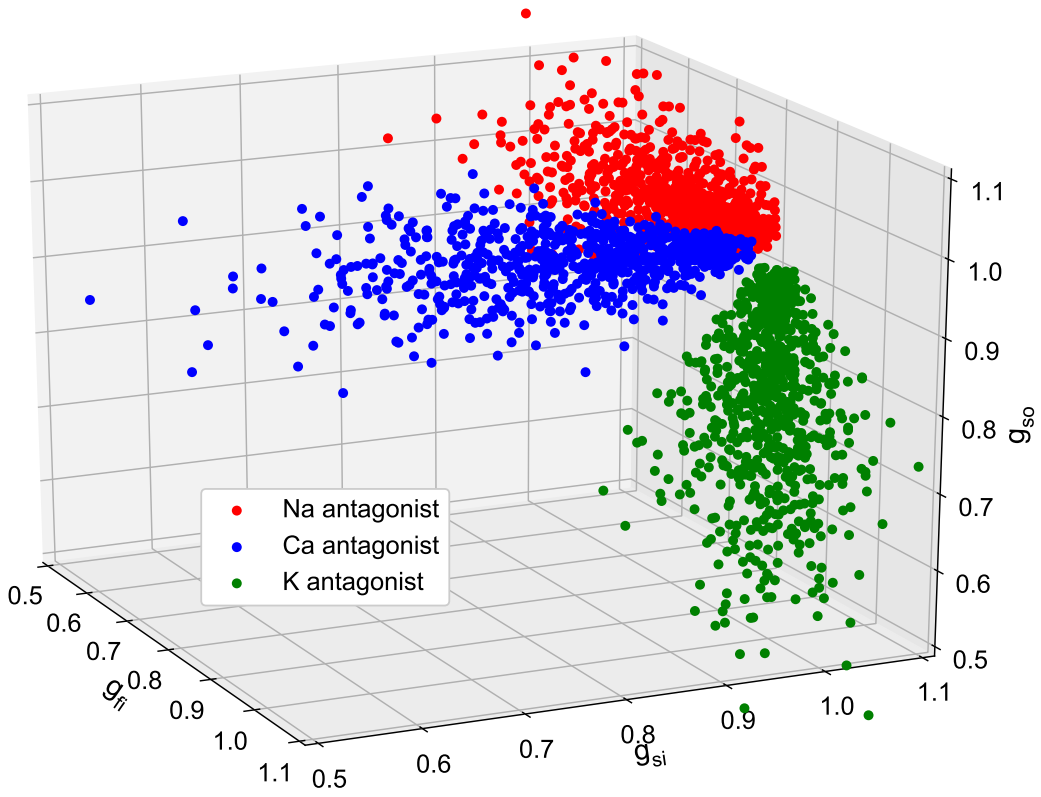

Figure 6: Conductances samples used to build the simulation set.

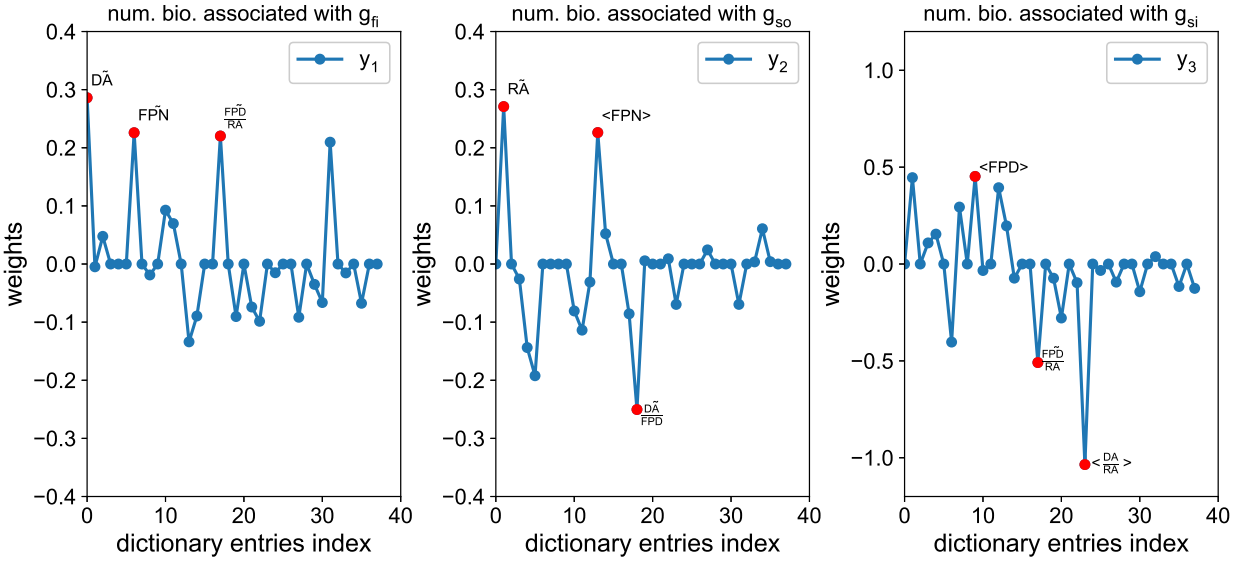

Figure 7: Example of numerical biomarkers weights.Caption here.

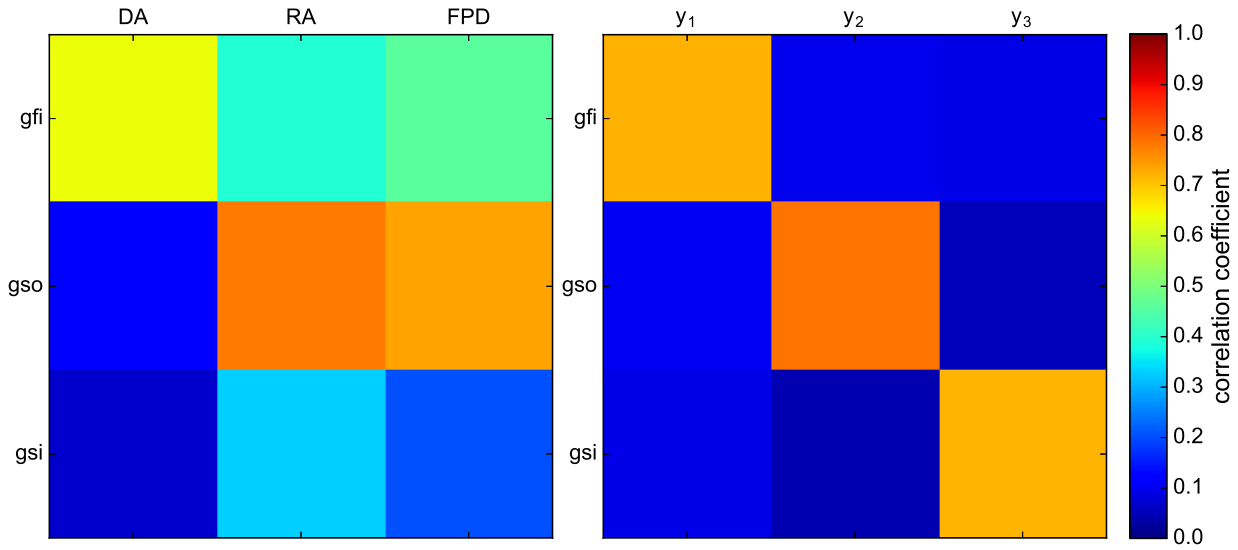

Figure 8: Correlation matrix of the conductances of interest with the "classical" biomarkers (left) and with the numerical biomarkers (right)