

Scientific Challenge of EMBEC 2024!

Introduction

Welcome to the IFMBE Scientific Challenge Competition of the 9th European Medical and Biological Engineering and Computing Conference (EMBEC 2024). In this edition of the IFMBE Scientific Challenge, the goal is to predict responses of genetically modified HEK (human embryonic kidney cells) to pulsed electric fields. The motivation includes finding new relationships between the shape and size of the cells and their influence on the response of intracellular Ca^{2+} concentrations as a response to stimulation with pulsed electric fields.

Context

Electroporation is a phenomenon in which the permeability of the cell membrane to ions and small molecules is temporarily increased after exposure to high intensity electric field pulses.

Electroporation is used in various fields of medicine. The electrical pulses used in electroporation-based treatments have been shown to affect the excitability of muscle and nerve cells. However, understanding the interplay between electroporation and electrophysiological response of excitable cells is complex because both ion channel gating and electroporation depend on dynamic changes in transmembrane voltage. In the context of cell excitation, calcium ion influx is also particularly important for muscle cells because it is directly involved in the molecular machinery responsible for muscle cell contraction.

Dataset – calcium concentration in spiking and non-spiking HEK cells

In this challenge, participants are given access to a dataset derived from microscopic images from 18 experiments in which intracellular calcium was monitored in genetically modified HEK (human embryonic kidney) cells that were exposed to pulsed electric fields of increasing magnitude. The HEK cells are genetically modified to stably express $\text{Na}_v1.5$ sodium channels and conditionally express $\text{K}_{ir}2.1$ potassium (K^+) channels. As such, the cells can be grown as an excitable variant (expressing both $\text{Na}_v1.5$ and $\text{K}_{ir}2.1$) and non-excitable variant (expressing $\text{Na}_v1.5$ only). $\text{K}_{ir}2.1$ expression is induced by growing the cells in the presence of the antibiotic doxycycline. In each experiment, the same sample of cells (excitable or non-excitable variant) was exposed to a sequence of nine 100- μs -long electric pulses resulting in different strength of electric field (0, 126, 150, 176, 200, 250, 300, 350, 360, and 400 V/cm) in order of increasing magnitude. For each applied pulse, the intracellular calcium was monitored using a calcium-sensitive fluorescent dye for 5 s before and 35 s after the pulse exposure. Fluorescence microscopy images (Figure 1A) show calcium concentration in cells before exposure to an electric pulse. For each of the 9 exposures, the cells in the captured images were segmented (see Figure 1 for an example) and the average fluorescence signal was extracted from individual cells to obtain the time course of the intracellular calcium. Segmentations were performed separately for each of the 9 exposures, because cells can change shape and move between exposures (the time between exposure to each pulse in the sequence was 2 min). Different classes of cell response can be identified from the extracted time courses of the intracellular calcium.

Since the responses to 9 exposures of the same cells cannot be considered entirely independent, we also provide the numbering of the cells, which was kept consistent throughout each experiment (meaning that within each experiment, the same number belongs to the same cell).

For convenience, we also include the original extracted fluorescence signal of individual cell for each experimental point, as well as the defined classes of response (see Figure 2 for examples). **Please**

note however, that these fluorescence signals are **not to be used** to achieve the goal of this challenge defined below; they are provided just to illustrate the different responses and the variability of these responses within the same class of response in terms of fluorescence changes after the exposure in time domain.

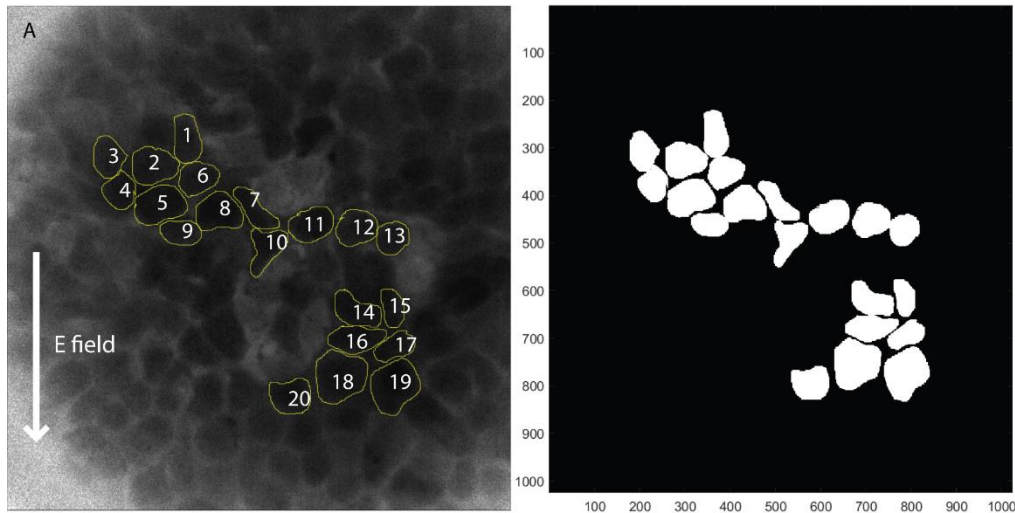


Figure 1: A) example of fluorescence image showing calcium ion concentration in cells. The image shows experiment 1 at 0 V/cm. The labelled cells are included in the public dataset. B) All segmentations of cells of the experiment shown in A. The dataset however is split into one binary mask per cell. The electric field in all exposures of all experiments is always oriented in the up-down direction, as indicated in A.

The entire dataset is split into the public and evaluation dataset. Data for the evaluation dataset is a subset of the entire dataset in which cells from all 18 experiments are represented almost equally, and result in a similar distribution of classes, i.e. the test sample is representative of the total dataset.

The public dataset consists of:

1. 2758 binary masks representing cells from 18 experiments (each experiment consisting of 9 exposures of the cells to varying electric field). An example of 20 segmented cells from one such exposure is shown in Figure 1.
2. Experimental parameters for the above mentioned cells
3. Class of each response

Binary segmentation masks

The segmentations are provided in the `publicSegmentation.mat` file. The file contains a cell array `publicSegmentation`, which contains the subscript indices of each nonzero pixel of every cell.

Experimental parameters

The **experimental parameters** are provided in the `publicExperimentInformation.csv` file which contains 6 columns with the following data:

Column	Description	Range
1	Presence of K+ channels This is a true/false variable that indicates if incubation with doxycycline was performed. 1 indicates K+ channels are present 0 indicates, that they are not.	[0,1]
2	Electric field strength [V/cm] The electric field intensity of the pulse. The 9 levels are listed in the description above above.	[0,400]
3	Number of nonzero pixels in cell segmentation binary mask Number of pixels in each binary mask. While this feature can be easily extracted from the segmentations, we are providing it here as one possibility of what to extract.	[2176,24651]
4	Experiment number This is simply the experiment number. Provided because responses of cells from a single experiment cannot be considered entirely independent.	[1, 18]
5	Cell number – index of segmented cell in a single experiment Number of the segmentation mask label. The number of segmented cells varies between experiments.	[1, 26]
6	Initial average fluorescence level of the cell before pulse delivery Additional information on the starting fluorescence of the signal. This corresponds to the starting point of the graphs (at time=0) shown in figure 2.	[21, 75.4]

Classes of response

The class of each response is provided in the `publicClasses.csv` file.

The classes are:

1. No response
2. Single small short response
3. Single prominent peak
4. Sustained response
5. Multiple peaks

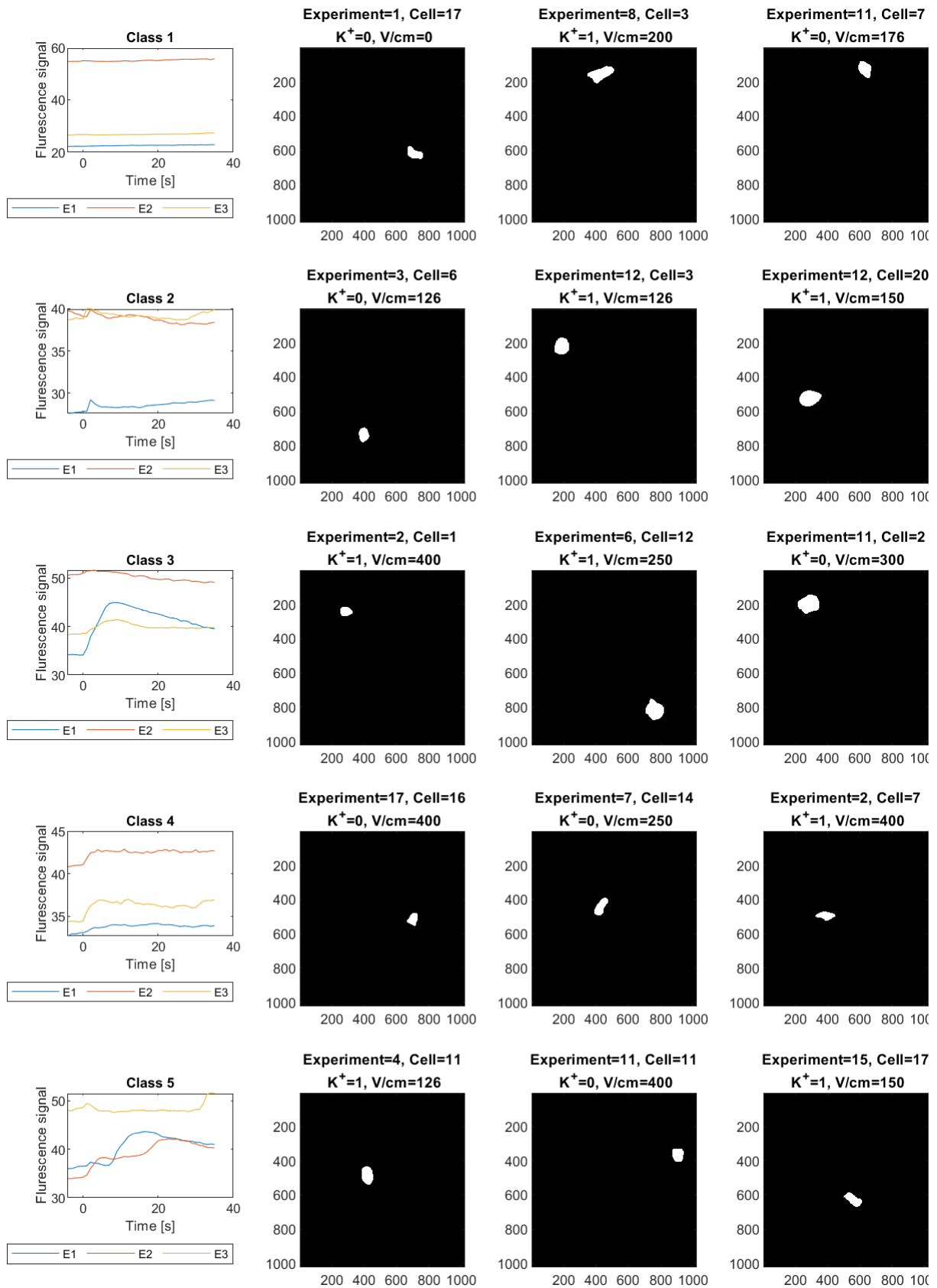


Figure 2: Examples of all possible classes of response. Each row contains three randomly selected cells from the public database with a certain class of response. In the left column are the fluorescence responses, of each respective class. Note that the curves show the time course of fluorescence for 5 seconds before and 35 seconds after application of the electric pulse. On the right

are the three binary masks of the cells whose time courses of response are shown in the leftmost column.

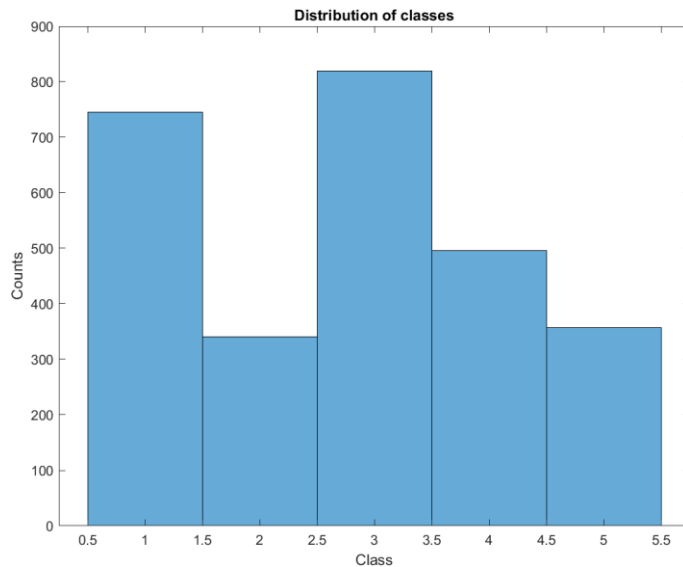


Figure 3: Distribution of classes in the public dataset.

Figure 2 shows the examples of responses of individual cells to exposure to electric pulse of given strength. Three examples are shown for each class of response (left-most column) along with the experimental parameters for each of the response and the binary masks of the corresponding cells. Please note once again that the fluorescence signals are **will not be provided in the evaluation set** to achieve the goal of this challenge; they are provided just to illustrate the different responses and the variability of these responses within the same class of response.

Figure 3 illustrates absolute frequencies (number of responses) for the five classes of response in the public data set.

Evaluation dataset

The evaluation dataset contains a subset of the entire dataset, and will be used to score the entries, but will be not made available to the participants. However, the data structure is identical to the public dataset above, and the distribution of classes is similar to the distribution of the public dataset.

Supplementary fluorescence signals

The supplementary fluorescence signals (the examples of which are shown in the left-most column of Figure 2) are stored in file publicFluorescenceSignals.csv. This is a 40x2758 matrix, where each column represents the time course of the fluorescence signal for a single cell. Again: these signals **will not be used in the evaluation set** for scoring the entries, therefore your solution of the challenge must also not use these signals; they are provided to illustrate the actual responses of individual cells to exposure of electric field.

Main goal

For the scientific challenge purposes, a set of 5 distinct classes of response of HEK cells to exposure to electric field is considered. The main goal is to develop a model, that is able to predict the response of each individual cell after exposure to an electric field pulse as one of the five different classes {1,2,3,4,5} based on the **segmentation (all information, that can be derived from the segmented cells)** and **experimental parameters alone**.

Since the induced transmembrane voltage is dependent on the size of the cell, and its orientation relative to the externally applied electric field (see also Figure 1 for the direction of the externally applied electric field, which was kept uniform across all experiments and exposures) [1]–[3], we encourage the participants to consider the provided segmentations and extract features that could be used to improve the prediction of the cellular response.

The Scientific Challenge

Quick Start

To compete in the Scientific Challenge of EMBEC 2024 follow these instructions:

1. Register for the Scientific Challenge using this link: <https://www.embec2024.org/challenge-registration-form/>
2. Download the public data set that you can use to develop your solution and the example source code (MATLAB or Python, respectively `exampleCode.m` and `exampleCode.py`) to be adapted for your entry.
3. Develop your algorithm based on available public data
4. Modify the example source code file `exampleCode.m` or `exampleCode.py` according to your algorithm.
5. Submit your modified compressed entry file for scoring through the Challenge website. The compressed file has to output the required data in the prescribed format in order to be scored.

Improperly formatted entries will not be scored (or counted against the submission limit).

Rules and Deadlines

In this challenge, we adopt modified rules of the Physionet Competition. The challenge is structured in a single phase in which the participants may submit up to 10 entries for scoring. Entries that cannot be scored (because of missing components, improper formatting, or excessive run time) are not counted against the entry limits.

The challenge starts on 15 November 2023 and will end on 13 February 2024 (3 months). Note that this deadline has been extended to 22 February.

Please note that, to be eligible for the award, you must also submit a full paper to EMBEC 2024. The paper should describe your approach to the Scientific Challenge of EMBEC 2024 (Deadline for submission of full papers only for Challenge participants is extended to February 25), at least, one team member must attend EMBEC 2024 and present the paper at the Scientific Challenge Special Session.

Moreover, we expect the submission of an overview paper with the best solutions to a prestigious Scientific Journal, co-authored by participants of the challenge and considering the publications and discussions at EMBEC 2024.

Eligibility

All official entries must be received no later than 23:59 UTC, February 22, 2024. In the interest of fairness to all participants, late entries will not be accepted or scored.

To be eligible for the award you must do all of the following:

1| Register

you should include the username, email address and team name.

2| Submit at least one entry that can be successfully scored

Submit at least one entry that can be successfully evaluated by the organizers before the set deadline (23:59 UTC, February 22, 2024). All submissions must be written in Matlab or Python 3.x and be self-contained (i.e. if a specific open-source and non-standard toolbox or library is required, it must be provided). All commercially available Mathworks toolboxes for MATLAB R2023a are valid). Please refer to the detailed information on data formats, input and output formats contained in the sample files.

3| Submit a full paper

to describe your work for the Scientific Challenge of EMBEC 2024 no later than on February 25, 2024 via the regular submission system. In your submission, indicate the global score for at least one successfully evaluated entry. You will be notified by email from EMBEC 2024 by March 3, 2024 whether your submission has been accepted (first decision).

4| Attend EMBEC 2024 (June 9-13, 2024)

and present your work there (at least one team member must attend EMBEC 2024).

Please do not submit analysis of this year's Scientific Challenge data to other conferences or journals until after EMBEC 2024 has taken place, so that all competitors are able to discuss the results in a single forum. We expect the submission of an overview paper with the best solutions of the challenge to a prestigious Scientific Journal, co-authored by participants of the challenge and taking into account the publications and discussions at EMBEC 2024.

Input/output data formats for scoring

For your entry to be scored successfully, you have to develop an algorithm that is able to read the segmentations stored in file `evaluationSegmentations.mat`, and experiment information

evaluationExperimentInformation.csv, and store the predicted classes as a file predictedClass.csv.

The segmentations stored in the evaluationSegmentations.mat file are stored as a cell array of subscript indices of nonzero element in each binary mask, assuming the image size of the original experiment 1024x1024 pixels. The indices are Matlab-based subscript indices, to use them as Python indices, you should therefore subtract 1 from each index.

Performance evaluation

The overall score for each entry is computed based on the predictedClass.txt file that is generated by your code. In this Scientific Challenge, we adopt the F1score metric as the basis for evaluation:

$$F1score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Where

$$Precision = \frac{TP}{TP + FP}$$

and

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

Where TP is true positive, FP is false positive, and FN is false negative.

The F1score is computed through a multi-class fomulation: each class against the others. Finally, the global score is computed as the average value of the F1scores for obtained for each of the five classes.

$$SCORE = \frac{1}{5} \sum_{i=1}^5 F1score_i$$

Score example

Let us assume an evaluation result for 20 cells. Your algorithm predicts the following result:

Cpredicted=[1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 5, 5].

The actual classes of the cells are as follows:

Cactual=[1, 1, 1, 1, 2, 1, 2, 2, 2, 3, 4, 3, 3, 4, 5, 3, 4, 4, 5, 5].

From the prediction and the actual class vectors, it is possible to build a confusion matrix:

		True class				
		C1	C2	C3	C4	C5
Predicted class	C1	4	1			
	C2	1	3	1	1	
	C3			2	1	1
	C4			1	2	
	C5					2

The individual F1score can be computed for each class as shown in the following table:

	TP	FP	FN	Precision	Recall	F1score
Class 1	4	1	1	0.800	0.800	0.800
Class 2	3	3	1	0.500	0.750	0.600
Class 3	2	2	2	0.500	0.500	0.500
Class 4	2	1	2	0.667	0.500	0.571
Class 5	2	0	1	1.000	0.667	0.800

Global score= 0.654

If all required files are included in your entry, and they are successfully executed, your submission will be evaluated and you will receive your preliminary score once the test is completed. You will receive an automatic email informing you of the results. The evaluation is carried out weekly. If you receive an error message instead, please read it carefully and resolve the problem(s) before resubmitting your entry.

Award

The scientific challenge will award a total of €1.000 (one thousand Euros) in scientific prizes.

Final provisions

The Challenge organizers reserve the right to disqualify entries if they violate the spirit of the challenge or submit malicious code.

The Challenge organizers reserve the right to modify the rules of the challenge if any, errors, inconsistencies, or important omissions are found.

Current version of this document: 2. Published on 16 February 2024.

Appendix 1: Technical details on experimental methods

The genetically modified HEK cell line was developed and generously provided by Adam E. Cohen's group at Harvard University [4], [5]. The cells are commercially available through ATCC (CRL-3479).

Intracellular Ca^{2+} was monitored using the fluorescent Ca^{2+} indicator Fura-2 acetoxymethyl (AM) ester (ThermoFisher Scientific). Changes in intracellular Ca^{2+} concentration influence the fluorescence spectrum of Fura-2 which can be detected in ratiometric measurements [6]. Cells grown in Lab-Tek chambers were stained with 2 μM Fura-2 AM in culture medium at 37°C for 30 min, 3x washed with Tyrode buffer (TB) and transferred to low K^+ TB containing 2 mM CaCl_2 .

Each sample of cells was exposed to a sequence of nine 100- μs -long electric pulses with interpulse delay of 2 min. Each pulse in the sequence had a higher amplitude. The electric field to which the cells were exposed was estimated as the ratio between the applied voltage and interelectrode

distance. Pulses were delivered by the Electro cell B10 electroporator (BetaTech, France) using a pair of parallel Pt/Ir wire electrodes.

Cells were observed under an inverted epifluorescent microscope (Zeiss Axiovert 200, Zeiss, Oberkochen, Germany) using a 63× objective and Prime sCMOS camera (Photometrics, Tucson, AZ, USA). Fura-2 was used in ratiometric measurements. Cells were illuminated with Xe light source and a monochromator (VisiChrome Polychromator, Visitron, Puchheim, Germany) at two excitation wavelengths (340 nm and 380 nm) and an exposure time of 200 ms for both wavelengths. The emission was detected at 510 nm wavelength using an appropriate filter set (Chroma 71500, Chroma Technology Corporation, Bellows Falls, VT, USA). Images were acquired with the software VisiView (Visitron) in time-lapse acquisition mode (39 s total duration of image acquisition, 40 images, one image every 1 s). Pulses were delivered manually (with no trigger) at the 5th image (at around 5 s). In the VisiView software, background was subtracted from each image and ratio image Fura-2 340/380 was calculated in each pixel. A higher intracellular Ca^{2+} concentration results in a higher Fura-2 340/380 ratio. The average ratio of the whole image was determined using ImageJ to obtain the Fura-2 ratio $R_{340/380}(t)$ as a function of time for each applied pulse.

References

- [1] T. Kotnik, L. Rems, M. Tarek, and D. Miklavčič, "Membrane Electroporation and Electroporation: Mechanisms and Models," *Annu. Rev. Biophys.*, vol. 48, no. 1, pp. 63–91, May 2019, doi: 10.1146/annurev-biophys-052118-115451.
- [2] J. Dermol-Černe, T. Batista Napotnik, M. Reberšek, and D. Miklavčič, "Short microsecond pulses achieve homogeneous electroporation of elongated biological cells irrespective of their orientation in electric field," *Sci Rep*, vol. 10, no. 1, p. 9149, Dec. 2020, doi: 10.1038/s41598-020-65830-3.
- [3] M. Scuderi *et al.*, "Characterization of Experimentally Observed Complex Interplay between Pulse Duration, Electrical Field Strength, and Cell Orientation on Electroporation Outcome Using a Time-Dependent Nonlinear Numerical Model," *Biomolecules*, vol. 13, no. 5, p. 727, Apr. 2023, doi: 10.3390/biom13050727.
- [4] H. M. McNamara, H. Zhang, C. A. Werley, and A. E. Cohen, "Optically Controlled Oscillators in an Engineered Bioelectric Tissue," *Phys. Rev. X*, vol. 6, no. 3, p. 031001, Jul. 2016, doi: 10.1103/PhysRevX.6.031001.
- [5] H. Tian *et al.*, "Video-based pooled screening yields improved far-red genetically encoded voltage indicators," *Nat Methods*, vol. 20, no. 7, pp. 1082–1094, Jul. 2023, doi: 10.1038/s41592-022-01743-5.
- [6] T. Batista Napotnik and D. Miklavčič, "In vitro electroporation detection methods – An overview," *Bioelectrochemistry*, vol. 120, pp. 166–182, Apr. 2018, doi: 10.1016/j.bioelechem.2017.12.005.