

Linking the sampling frequency with multiscale entropy to classify mitoBK patch-clamp data

Lukasz Machura^{a,b,*}, Agata Wawrzekiewicz-Jałowicka^c, Piotr Bednarczyk^d, Paulina Trybek^{a,b}

^a Chelkowski Institute of Physics, University of Silesia in Katowice, 41-500 Chorzow, Poland

^b Silesian Center for Education and Interdisciplinary Research, University of Silesia, 41-500 Chorzow, Poland

^c Department of Physical Chemistry and Technology of Polymers, Silesian University of Technology, 44-100 Gliwice, Poland

^d Department of Physics and Biophysics, Institute of Biology, Warsaw, University of Life Sciences – SGGW, 02-787 Warszawa, Poland

ARTICLE INFO

Keywords:

Sampling frequency

PSD

Channel noise

Sample entropy

Ion channels

Classification

ABSTRACT

We analyze the activity of large-conductance voltage- and Ca^{2+} -activated potassium channels located in the inner mitochondrial membrane (mitoBK) from human dermal fibroblast cells. The ion current activity registered via the patch-clamp technique was taken into consideration. At the preliminary stage, we performed an in-depth analysis of the signal power spectrum to find an optimal sampling frequency and study the impact of different sampling on changes in the information hidden in the signal. We found the optimal 10 kHz sampling frequency for the fibroblast's mitoBK currents sequences. Interestingly, as the signal sampling rate increases, we can observe a decrease in entropy values. The application of Multiscale Entropy analysis enabled a practical classification of single-channel current traces at various membrane potentials. Using the machine learning techniques such as K-Nearest Neighbors and Support Vector Machine, optimized by the Stochastic Gradient Descent algorithm with Sample Entropy values as inputs, allowed us to assess the more outstanding accuracy scores for the chosen classifiers at membrane the hyperpolarization than at its depolarization.

1. Introduction

There is no overall consensus on the definition of complexity. In an effort to understand the complex phenomena hidden behind the experiments, every now and then, measures for a description of changing system complexity from data are proposed [1]. Ludwig Boltzmann explained Entropy as the measure of the number of possible microscopic states of a system that relates to the macroscopic state of the system. In information theory, Entropy is seen as the rate of generation of new information in discrete signals [2]. It quantifies the probability density function of the distribution of the measured values, portraying the static properties of the system. The amount of uncertainty for the experiment which have the possible result's probabilities $P = \{p_1, p_2, \dots, p_n\}$, is called the Shannon Entropy H of the distribution P and is given by the negative average of the logarithm of the distribution

$$H = - \langle \log P \rangle = - \sum_{i=1}^n p_i \log p_i \quad (1)$$

Nowadays, technology allows measuring biological systems with high precision. The ever-decreasing prices of mass storage byte allow for the recording and deposition of terabytes of measurement data. Also, quickly increasing computer power makes the situation tempting to study densely sampled recorded signals. However, maintaining high-precision recordings is not only redundant but often leads the conclusions astray. This work raises the problem of selecting the appropriate sampling frequency needed to obtain as much valuable information about the signal as possible. In the context of information entropy, this aspect seems to be of great importance. In some situations, less is more, as we will illustrate this issue with the recordings of the ion channels' activity.

Dermal fibroblasts are predominant cells within the dermis layer of skin [3]. They play a crucial role in the regulation of skin physiology and pathology, including production and organization of connective tissue, which allows for recovery of the skin from injury [4]. Fibroblasts also have secretory functions, i.e., they release multiple growth factors and cytokines, which are relevant in the stimulation of cell proliferation and

Abbreviations: KNN, K-Nearest Neighbors; H, Shannon Entropy; MSE, Multiscale Sample Entropy; SEN, Spectral Entropy; SE, Sample Entropy; SGD, Stochastic Gradient Descent; PSD, Power Spectrum Density.

* Corresponding author at: Chelkowski Institute of Physics, University of Silesia in Katowice, 41-500 Chorzow, Poland.

E-mail address: lukasz.machura@us.edu.pl (L. Machura).

<https://doi.org/10.1016/j.bspc.2022.103680>

Received 10 August 2021; Received in revised form 4 March 2022; Accepted 9 April 2022

Available online 19 April 2022

1746-8094/© 2022 Elsevier Ltd. All rights reserved.

apoptosis, immune responses, and generation of inflammatory processes as well as production and organization of the extracellular matrix [5]. The proper realization of the mentioned functions frequently depends on mitochondrial energy metabolism. In turn, ion channels' activity is a key factor in mitochondrial metabolism, the efficiency of oxidative phosphorylation, and cytoprotection. It has been proposed that potassium transport through the mitochondrial inner membrane via mitochondrial potassium channels is an essential player in cell life and death [6]. From this perspective, fibroblasts [7], and particularly the mitochondrial potassium channels, can be considered as drug targets [8].

In this work, we analyze the activity of the large-conductance voltage- and Ca^{2+} -activated potassium channels (mitoBK channels) present in mitochondria of the primary human dermal fibroblast cell line. Due to their relatively sizeable single-channel conductance (up to 300 pS), they can very rapidly and efficiently regulate the mitochondrial membrane potential, respiration, and production of reactive oxygen species [6]. The mitoBK channels are encoded by the *Kcnma1* (*Slo1*) gene, which is the same as their cell membrane counter-parts BK which had been broader characterized hitherto [9]. Despite the genetic match, the exonic composition of the mitoBK channels can significantly vary from the ones expressed in the cell membrane due to alternative splicing or post-translational modifications. Most importantly, the mitoBK channels are expressed when the *Kcnma1* undergoes splicing to the DEC isoform [10], which implies some functional differences between the channel variants from the plasma membrane and the mitochondrial membranes. Despite the existence of the cryo-EM structures of the *Slo1* channel from the plasma membrane [11], the molecular structure of the mitochondrial channel variant is still not available. Thus, regardless of the existence of some initial studies on the stable BK channel conformations which were performed by means of molecular dynamics [11], their real utility in the context of mitochondrial channel variants can be disputable since there is no one-to-one structural correspondence between the BK and mitoBK channels. From this perspective, the best solution to make some basic inferences about the mitoBK channel system dynamics is to carry out a detailed analysis of the channel activity in a natural timescale of the process.

The electrophysiological techniques, such as patch clamp, allow us to capture ions' movement through single-channel proteins in real-time [12]. The recorded signal, being a series of single-channel currents, resembles a channel protein's conformational diffusion. This kind of diffusion determines the complexity and, generally, the observable structural features of the patch clamp time series. One can assume that in most cases, the channel's conformations are sufficiently distant from each other structurally and energetically, to notably differ from the other ones by current amplitude, lifespan, or possible system of conformational switching enabling to enter a given state and exit it. Thus, the presence and abundance of distinct channel conformations are suspected to affect the complexity of the experimental data.

The ion current through mitoBK channels is characterized by a high degree of complexity, largely dependent on the measurement conditions. Thus, despite long-term studies on gating machinery, not all aspects of the channel's activity mechanism are already resolved. As mentioned before, to obtain comprehensive knowledge about the system, we take into consideration the dynamical diversity of mitoBK ion current time series. In this work, we analyze the entropy of single-channel data registered at different levels of hyper- and depolarization of the membrane.

The information-based measures such as Spectral Entropy (SEN) or Sample Entropy (SE) and its multiscale variant - Multiscale Sample Entropy (MSE) [13,14] give valuable knowledge about the system's ability to adapt in the ever-changing environment and has been successfully applied in an extensive range of biological data analyses, also in the different types of electrophysiological signals [15–17]. SEN quantifies the energy distribution in the frequency domain, illustrating the static complexity in a like manner to the Shannon Entropy. On the other hand, SE and its multiscale version MSE evaluate the system's

dynamical complexity. However, a relatively low number of works addresses the information entropy in context of the ion channels activity, particularly BK channels or their mitochondrial analogues (mitoBK) [18–20]. The investigation of signal complexity employing Entropy measures can ensure a deeper insight into the actual character of the mitoBK biosystem. The activity of the channel protein is suspected to be regulated by interacting mechanisms that.

- can operate across multiple spatial and temporal scales, e.g., movement of the functional domains of the channel vs. fluctuations of the membrane or protein–protein interactions,
- may contain deterministic and stochastic components, e.g., Coulombic or Lennard–Jones interactions, which influence gating, vs. thermal fluctuations within the pore–gate domain.

Considering the MSE method's potential utility in studies of ion channel dynamics, its application in the patch-clamp signal analysis allows evaluating the relative complexity of a switching mechanism between the channel microsystem substates (stable or metastable conformations) at given external conditions. The higher values of the entropy, the more intricate mechanism should govern the ion channel protein's conformational diffusion. Moreover, the changes of entropy with the timescale enable us to infer whether, with raising observation time, new complex structures are unraveled within the signal or not. This may be interpreted in terms of the existence of large-scale factors affecting the single-molecule dynamics, as the large-scale membrane fluctuations. For more accurate classification of the analyzed time series concerning the Entropy values over the range of different scales, the classification algorithms were implemented in the form of the K-Nearest Neighbors technique (KNN) and Stochastic Gradient Descent (SGD) method.

This work is organized as follows. In the first subsection of the next part we briefly explain the experimental setup of obtaining the recordings of single-channel currents in mitoBK channels. In the second subsection one we introduce all addressed quantifiers of complexity and the methods of the numerical analysis of data. In the Results section the introduced characteristics are calculated for the data sampled at the optimal frequency of 10 kHz. The selection of the aforementioned frequency is elaborated in detail as well. In Section 4 the MSE and its statistical analysis is provided. Section 5 shows the possible separation of recordings at different membrane potentials. The manuscript ends with the discussion and acknowledgments.

2. Materials and methods

2.1. Experiment

2.1.1. Cell culture and mitoplast preparation

The mitoBK channels investigated in this work were isolated from the primary human dermal fibroblasts cell line (HDFa) [6], which was available commercially (ATCC-PCS-201–012 line). In this aim, fibroblast cells were cultured in DMEM medium supplemented with 10% fetal bovine serum, 2 mM L-glutamine, 100 U/ml penicillin, and 100 $\mu\text{g}/\text{ml}$ streptomycin at 37°C in a humidified atmosphere with 5% CO_2 . Every fourth day, the cells were reseeded.

Mitochondria were prepared from the fibroblast cells as previously described in [6]. In short, mitochondria were obtained after a series of appropriately adjusted centrifugations, homogenizations, and pellet resuspensions at 4°C. First, fibroblasts were placed in phosphate-buffered saline (PBS) and centrifuged at 400 \times g for 5 min. The obtained cell pellet was then resuspended in a calcium-free preparation solution (250 mM sucrose, 5 mM HEPES and 1 mM EGTA, pH 7.2) and homogenized (Wheaton homogenizer, U.S.A.). The homogenate was centrifuged at 9200 \times g for 10 min, and the obtained pellet was once more suspended and centrifuged at 700–750 \times g for 10 min. The supernatant was transferred to a new tube and centrifuged at 9200 \times g for

10 min, which allowed to get the pelleted mitochondria. The mitochondria pellet was resuspended in about 0.3 ml of the preparation solution. The entire procedure was performed at 4°C.

To prepare mitoplasts, the fibroblast mitochondria were incubated in a hypotonic solution (5 mM HEPES and 200 μ M CaCl₂, pH 7.2) for ca. 1 min for inducing swelling and breakage of the outer membrane. Finally, to restore the medium's isotonicity, a hypertonic solution (750 mM KCl, 30 mM HEPES, and 200 μ M CaCl₂, pH 7.2) was added.

2.1.2. Electrophysiology

The recordings of single-channel currents were obtained through the patch-clamp technique in the mitoplast-attached inside-out mode. The pipettes used in experiments were made from borosilicate glass (Harvard, UK) pulled using Flaming/Brown puller. They reached a resistance of 10–20 M Ω . The isotonic solution filling of the patch-clamp pipette contained: 150 mM KCl, 200 μ M CaCl₂ and 10 mM HEPES at pH 7.2, which allowed for reaching the full Ca²⁺-activation of the investigated mitoBK channels.

All patch-clamp experiments were carried out in a voltage-clamp mode at pipette potentials fixed at -60, -40, -20, 20, 40, and 60 mV using the patch-clamp amplifier Axopatch 200B (Molecular Devices Corporation, U.S.A.). The obtained signal was low-pass filtered at a frequency of 1 kHz and sampled at 100 kHz. The measurement error of single-channel currents was $\Delta I = 1 \times 10^{-6}$ pA, which was determined by the possible measurement resolution of the equipment. The single-channel recordings were recorded at room temperature. Each experimental time series comprised $N = 1.98\text{--}2.00 \times 10^5$ current values.

2.2. Numerical analysis

2.2.1. Spectral entropy

Frequency-based entropy quantifiers express the uniformity of signal energy distribution in the frequency domain. It takes maximum value for flat and minimum for single peaked power spectrum density (PSD). Spectral Entropy (SEN), as an analogue to Shannon information entropy [2], is defined as a logarithmic average of the normalized components p_i of the i -th density components $PSD(f_i)$ [21]

$$p_i = \frac{PSD(f_i)}{\sum_i PSD(f_i)} \quad SEN = - \sum_i p_i \log p_i \quad (2)$$

Higher values of SEN suggest uniformity in the distribution of the signal energy. The PSD is calculated using two classical methods, the Discrete Fourier Transform (DFT) and its averaged version, the Welch method [22].

2.2.2. Sample entropy

Sample Entropy is a measure of uncertainty closely related to the Metric Entropy, a dynamical complexity measure introduced by Kolmogorov in the late 50' of the last century [23,24]. For classical systems, it is a measure of the degree of the irregularity inherent in the system's dynamics. This technique to determine changing system complexity from data was first introduced by Pincus as Approximate Entropy (*ApEn*) [25]. It was later updated to *SampEn* [26] where self matches are not taken into account as in the original *ApEn*.

We consider time series $X = \{x_i\}_{i=1}^N$ which consists of N data points. A set of vectors $U_m(i) = \{x_i, \dots, x_{i+m-1}\}_{i=1}^{N-m+1}$ represent m consecutive values of series starting with the i -th data point. The length m of the vectors is often called the embedding dimension. The difference between two sets $U_m(i)$ and $U_m(j)$ is taken as the Chebyshev distance defined as the maximum absolute difference between their scalar components

$$d[U_m(i), U_m(j)] = \max_{k=0, \dots, m-1} (|x(i+k) - x(j+k)|) \quad (3)$$

To measure this distance, one needs to heuristically determine the

tolerance threshold r referred to as a similarity criterion or the distance threshold for two template vectors. It is usually taken within the range between 10% and 20% of the standard deviation σ of X [27]. If the absolute difference of any pair of the components is larger than r , i.e. $d[U_m(i), U_m(j)] > r$, the vectors are not similar. With this definition we can calculate the probability $C_i^m(r)$ that any $U_m(i)$ vector is close to any other vector $U_m(j)$. The $n_i^m(r)$ is a number of $U_m(j)$ vectors ($1 \leq j \leq N-m$, $j \neq i$) that are close enough to the pattern vector $U_m(i)$

$$C_i^m(r) = \frac{n_i^m(r)}{N-m} \quad (4)$$

The probability $C^m(r)$ that any two vectors are within r of each other is taken as an average over the possible pattern vectors $U_m(i)$

$$C^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} C_i^m(r) \quad (5)$$

The negative logarithm of the conditional probability that two sequences which are similar for m points remain comparable to the $m+1$ points, defines the Sample Entropy

$$SampEn(m, r, N) = - \ln \left[\frac{C^{m+1}(r)}{C^m(r)} \right] \quad (6)$$

For the above calculations $j \neq i$. In the following the values of $m = 1$ and $r = 0.2\sigma$ have been used.

2.2.3. Multiscale entropy

Multiscale Entropy (MSE) is an extension of *SampEn* estimated for different time scales. We address different time scales by implementing the coarse-graining procedure of resampling the series [21]. Typically a copy y of coarse-grained time series is taken by averaging the data points in each of the non-overlapping windows j with the length corresponding to the desired scale τ

$$y_j^\tau = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i, \quad 1 \leq j \leq \frac{N}{\tau} \quad (7)$$

For each y_j^τ series the value of *SampEn* (6) is calculated. The plot of *SampEn* versus τ provides the requested MSE curve.

Whenever we talk about entropy in an article, we take the original formulation of Shannon [2] into account, where the logarithm base equals two, so the unit of entropy is called Shannon (Sh) or bit.

2.3. K-nearest neighbors

Despite its simplicity, the KNN algorithm is an effective tool widely used in the class of problems with labeled data classification, including electrophysiological time series data [28,29]. Each data point, represented by n -features, is placed in n -dimensional space. The number of k -neighbors for the new point is selected before assigning this data to an existing class. Next, the distance between a new data point and all other training data points is calculated. As a distance metric, the Euclidean formula is most often used [8]. In the final part, the class of new points is selected based on the majority vote (the class with the largest amount of nearest points to the new data is selected).

$$d(x_i, x_j) = \sqrt{\sum_{i=1}^N (x_i - x_j)^2}, \quad (8)$$

2.3.1. Support vector machine and stochastic gradient descent

Support Vector Machine (SVM) is a supervised learning model. This classifier aims to determine the hyperplanes separating cases belonging to different classes with a maximum margin. In our implementation, the Stochastic Gradient Descent (SGD) minimizes a loss function in the Linear Support Vector Machine technique. The SGD is a similar technique to standard Gradient Descent, which is used, e.g., in a logistic

regression model. It is faster because it estimates only the derivative of a single random instance, not all the points. SGD is a popular algorithm for optimizing a wide range of models, including (linear) support vector machines, logistic regression, or graphical models.

3. Results

3.1. Sampling frequency selection

Preliminary stage of this analysis include the selection of the sampling rate, which closely corresponds to the time-scale of the system's dynamics. The measuring equipment allowed us to gain a probing frequency of 100 kHz to record the mitoBK channel's activity. Here we postulate, that the high sampling rate would have a relatively significant influence on the information entropy measures. This, in turn, instead of meaningful results, we would effectively observe the impact of the oversampling. In the following we standardized the length of the data to 198000 points (almost 2 s long recordings) for all samples.

In the mid-sixties of the last century, Derksen and Verwee presented the power spectral density (PSD) of myelinated axons as being inversely proportional to the frequency f [30]. The flicker or pink noise reflects the ion channels dynamics' complexity, given by random switches between the channel's open and close states. What is defined as a channel noise has its origin in several distinct properties of the complex membrane structure [31] and the complex equilibrium protein dynamics that influence channel conductance. Therefore, such noise is not simply a manifestation of nonequilibrium transport phenomena but rather echoes several agents' interplay influencing the channel dynamics [32]. In addition to real ion channels, similar characteristics were also found in nanofabricated synthetic pores [33].

Although we are able to study the complex dynamics of the processes taking place in membranes almost independently for each of the components, it is still unclear what specific mechanism leads to pink noise, which is an averaged effect of ion channel dynamics. Any process with doubly harmonic diminution produces pink noise over a frequency bandwidth proportional to the number N of evenly spaced on real line eigenvalues of the diagonalizable time-evolution generator [34]. In such systems PSD exhibits three distinct segments: constant flat for low frequencies ($f < 3a/2\pi^2$), decreasing as $1/f$ for moderate frequencies ($3a/2\pi^2 < f \lesssim aN/4\pi^2$) and as $1/f^2$ for fast dynamics ($f \gtrsim aN/4\pi^2$) for some nonnegative constant a .

The presence of such complex scaling of PSD can indicate an actual complex $f^{-\alpha}$, $\alpha = 1, 2$ characteristics known for ion channels. We will test the PSD against the probing frequency and search for the best representation of the intricate channel noise dynamics for further entropy investigations. In Fig. 1 we present the PSD calculated for several

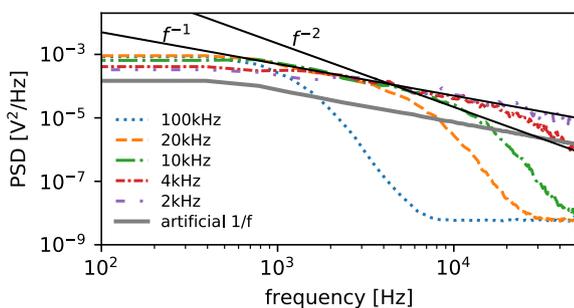


Fig. 1. Power spectral density as a function of the frequency calculated for artificial pink noise (solid gray line) as well as for the actual channel measurement for several sampling rates: 100 kHz (blue dotted line), 20 kHz (orange dashed line), 10 kHz (green dashed-dotted line), 4 kHz (red dense dashed-dotted line) and 2 kHz (purple dashed-double-dotted line). Two solid lines mark the $f^{-\alpha}$, $\alpha = 1, 2$ and serve as a guide. PSD was calculated with the Welch method.

sampling frequencies. The minimal rational sampling rate is 2 kHz due to the restrictions imposed by the low-pass filtering, as the sampling rate is supposed to be greater or equal to doubled threshold frequency in low-pass filtering. In such a situation, 100 kHz (blue dotted line), 20 kHz (orange dashed line), 10 kHz (green dashed-dotted line), 4 kHz (red dense dashed-dotted line) and finally, the limiting sampling rate 2 kHz (purple dashed-double-dotted line) are displayed. On top of that, the characteristic of the artificial $1/f$ noise is shown with a gray solid line for comparison.

The PSD of the original sampling rate (blue dotted line) does not show any of the features essential for the pink noise. Reducing the sampling to 20 kHz, either by keeping every 5th data point or by averaging windows of consecutive 5 points, brings the periodogram closer to the desired shape, as the distinct window of $PSD \propto f^{-1}$ decrease becomes apparent (see the middle part of the orange dashed line). Further reduction of the sampling rate reveals the demanded peculiarities. For the sampling rate of 10 kHz (green dashed-dotted line), we can identify all three regions: flat, f^{-1} and f^{-2} . This condition stays present for even lower rates, however, with the cost of reduction of f^{-2} part. The latter region disappears as we decrease the sampling even more and become absent for the limiting sampling rate of 2 kHz (purple dashed-double-dotted line). Interestingly in such a case, the PSD shows considerable similarities to the artificial $1/f$ signal (solid gray line).

The closer inspection of the PSD obtained for the 10 kHz allow for the exact calculation of the specific value of $a \simeq 5140$ and the corresponding number of eigenvalues $N = 42$. All three regimes with in-between borders marked by the red arrowheads are visible in Fig. 2.

The phase diagram presented in the two lower panels of Fig. 3 summarizes the sampling influence on the information contained in the PSD. It is divided into four regions depending on the rate of spectral density decay. The light green color shows the frequency range for which the PSD is approximately constant (f^0). Green and dark-green sections correspond to the $1/f$ and $1/f^2$ characteristics, respectively. The area marked in red corresponds to PSD decay faster than $1/f^2$. One can see from the diagram that there are more sampling frequencies which show all three consecutive scaling functions describing sections of PSD f^n , $n = 0, -1, -2$, which one need to relate the channel dynamics to the doubly harmonic diminution process [34], although the highest chance of finding such properties lies around 10 kHz. This finding holds for other values of the membrane potential, independent of hyper- and depolarization states. Unfortunately, for the original recording frequency of 100 kHz, we did not observe the necessary relation for the $1/f$ decay. c.f. the solid blue line in upper panels of Fig. 3.

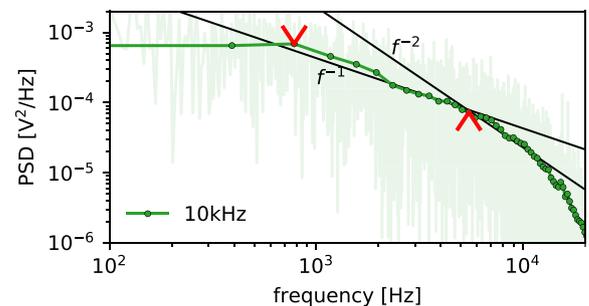


Fig. 2. Power spectral density as a function of the frequency calculated for the sampling rate of 10 kHz. PSD calculated using the Welch method (solid green line) is shadowed by the results of the discrete Fourier transform (DFT, light green in the background) manifest all three distinct scaling regimes: constant flat if $f \lesssim 780$ Hz (red down arrowhead), $1/f$ for $780 \text{ Hz} \lesssim f \lesssim 5.5 \text{ kHz}$ (red up arrowhead) and $1/f^2$ for the fastest dynamics. It is because of the logarithmic scale that data points get denser toward the right end of the graph.

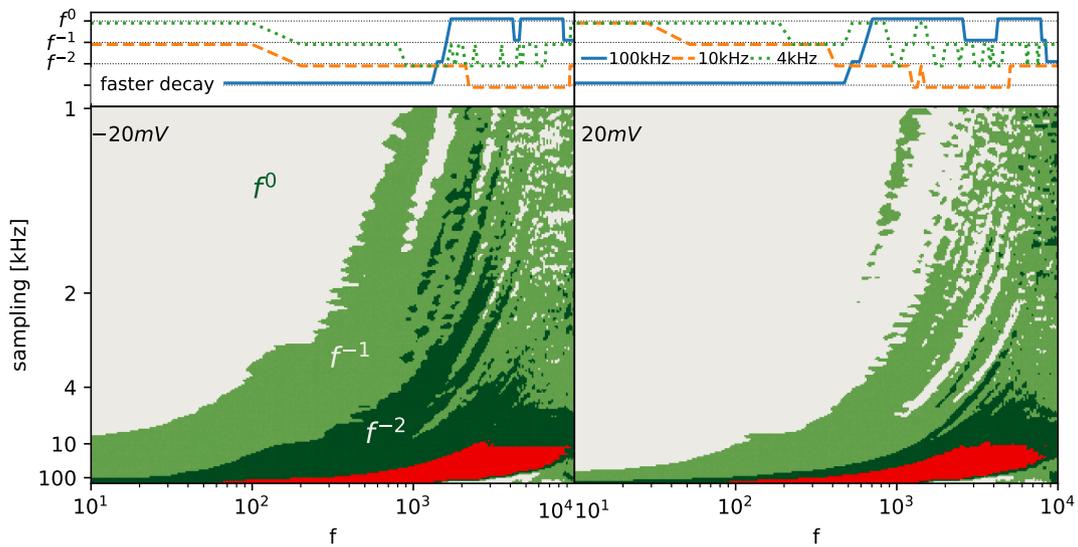


Fig. 3. Phase diagram of the influence of sampling frequency on the PSD characteristics for the signal recorded at the representative magnitude of hyperpolarization (left, $-20mV$) and depolarization (right, $20mV$). Four colors in two lower panels (light-green, green, dark-green, and red) correspond to four decay rates (flat constant, f^{-1} , f^{-2} and faster than the latter), respectively. The upper panels present the exponent variation for three selected values of the sampling frequencies (solid blue for 100 kHz, dashed orange for 10 kHz and dotted green for 4 kHz).

3.2. Spectral entropy

In Fig. 4 the SEN (2) averaged over all recordings for hyperpolarization (a), and depolarization (b) is presented for the sampling frequencies in question. We want to point out no frequency-domain information loss (or gain) if one decides to drop the sampling from 100 kHz to 10 kHz. In other words - all the correlations hidden behind densely sampled data will still be present in a ten times smaller sample. The determination of PSD is insensitive to the nonlinearities that might be present in the signal. Next, we address the sample entropy to get rid of this gap.

3.3. Sample entropy

Despite the drastic difference between PSD calculated for data with 100 kHz and ten times lower sampling frequency, it turns out there is no actual visual discrepancy in the time dependence of ionic current, c.f. corresponding black solid (100 kHz) and green dashed lines (10 kHz) in Fig. 5 for comparison. There is, however, a visible difference in variability of the latter data sets and the same signal but sampled with 4 kHz (solid red line).

In the context of signal analysis, information entropy measures the

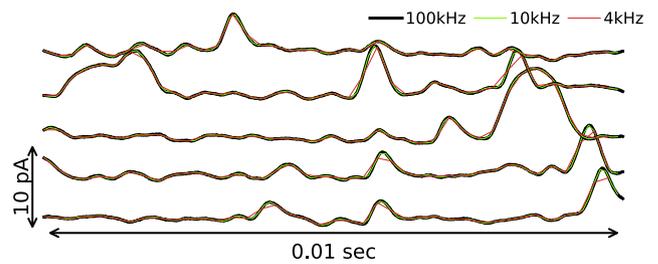


Fig. 5. Similar plots of ionic current with sampling rates of 100 kHz (black solid), 10 kHz (lawn-green dashed) and 4 kHz (red solid) shown against the same signal source but with 4 kHz rate (solid red) for single recording at depolarized membrane potential of $20mV$.

meaningful structural richness of the data points sequences. The smaller the variation in the series, the lower the value of information entropy. This has consequences in decreasing entropy values along with increasing sampling rate of mitoBK channel's data as is presented in Fig. 6, where the MSE curves are shown for different sampling rates, similarly to the section above. We can observe that the lowest sample

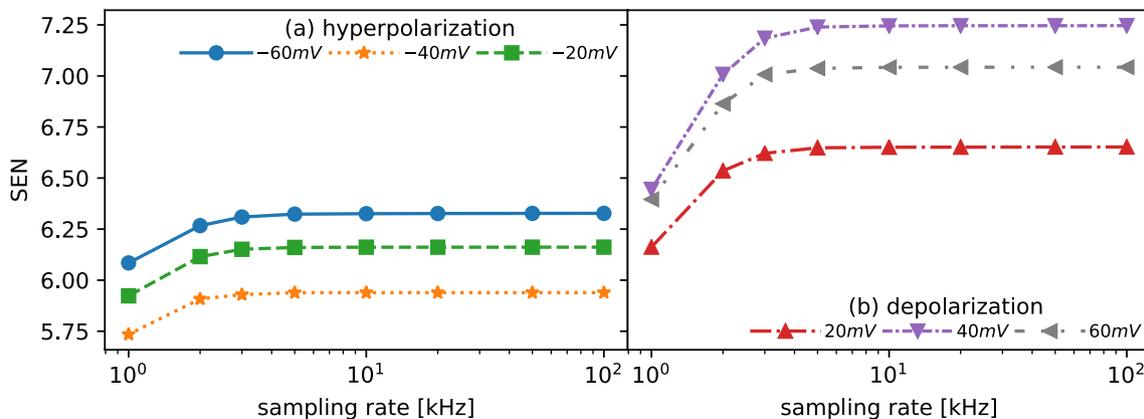


Fig. 4. The values of SEN calculated over range of sampling frequencies for the time series registered at different potential $U_C [-60, 60]$ mV with the step of 20 mV for hyperpolarization (a) and depolarization (b).

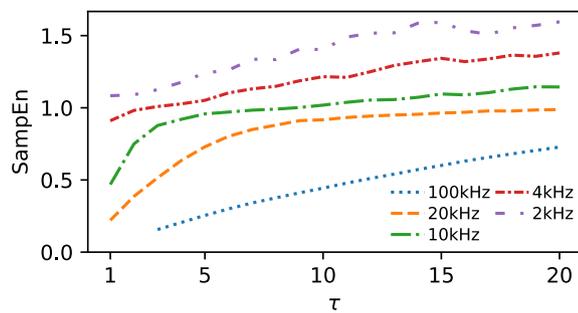


Fig. 6. Sample Entropy as a function of the scale calculated for the different sampling rates: 100 kHz (blue dotted line), 20 kHz (orange dashed line), 10 kHz (green dashed-dotted line), 4 kHz (red dense dashed-dotted line) and 2 kHz (purple dashed-double-dotted line).

entropy values occur for the highest sampling rate across the whole scaling range. Besides, the shape of the MSE curve in this case mostly differ from the others. The dependency looks nearly linear, c.f. the blue dotted line in Fig. 6.

In contrast, when the sampling rate is too rough, the observed changes of entropy may not correspond to the system's real complexity, as seen on PSD above. This effect may stem from omitting the crucial events in system dynamics. In such a case, one cannot observe the whole path of entering a given state and exiting it during the conformational switching. Thus, one can introduce an artificial gain in complexity (higher entropy values than expected) due to ignoring intermediate stages linking the remote conformational states. The second consequence undersampling causes is the apparent timescale of conformational diffusion of channel protein being slower than the actual one. So the discrepancies between the rapid components (movement of channel pore-forming helices) and the slow components (interprotein interactions, membrane fluctuations) of the observed single-channel current are lower than in reality. This, in turn, may lead to the almost linear dependency of entropy on a scale, as in the case of the lower valued sampling cases of 4 kHz (red dense dashed-dotted line) and 2 kHz (purple dashed-double-dotted line) in Fig. 6.

3.4. 10 kHz as an optimal sampling frequency

The above arguments indicate 10 kHz sampling frequency as possibly the most adequate representation of the peculiarities of the dynamics of ion current changes within the mitoBK channels activity. It is also believed that experimental resolution of 10 kHz can mostly correspond to the actual time scale of the mitoBK channel's dynamics. Fortunately, for the chosen sampling frequency, one can mark the most significant changes in entropy for small scales, giving valuable information about the nature of the tested signal's complexity. Increasing the scale from 1 to 4 unravels the signal's new structures, which probably are affected mainly by the relatively rapid components of the analyzed process, i.e., channel helices' movement. The weakly increasing trend observable for larger scales stems mainly from the relatively slow components of the systems' dynamics like the membrane fluctuations. The choice of 10 kHz let us keep more details about the signal. It allows us to monitor the global system dynamics in real-time without oversampling on the one hand and losing too much information about the mitoBK conformational diffusion observed in the form of ion current fluctuations on the other.

4. Multiscale sample entropy

The central result of this work is presented in Fig. 7, where the average SampEn values are calculated from 11 independent time series of single-channel currents at fixed membrane potential and displayed separately at membrane hyperpolarization (a) and depolarization (b).

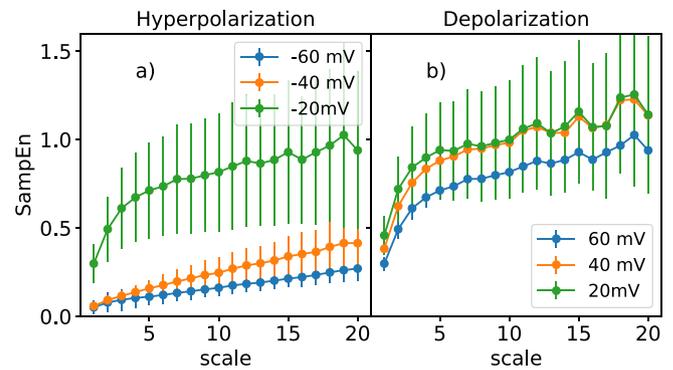


Fig. 7. The mean values of SampEn entropy calculated over a range of scales sc [1, 20] for the time series registered at different potentials U_C [−60, 60] mV with the step of 20 mV for hyperpolarization (a) and depolarization (b) sampled at 10 kHz. The average values of SampEn at each individual scale are presented with standard deviations.

The MSE curves are assigned to the three different potential values: $U = [20\text{mV}, 40\text{mV}, 60\text{mV}]$ for the depolarization of the mitochondrial membrane, and $U = [-20\text{mV}, -40\text{mV}, -60\text{mV}]$ in the case of hyperpolarization. For the later (see Fig. 7a), the state recorded at -20mV potential stands out the most, with the highest values of SampEn across all time scales. In turn, at membrane depolarization, the 60mV potential has the significantly lowest entropy values over the whole range of scales. The system's complexity is highest in terms of intermediate values of open state probability at each observation scale. In those terms, the energetic barriers separating the functionally open and closed conformations are relatively low, so the channel can quite freely fluctuate between the manifolds of open and closed substates. Therefore, at the intermediate potential (-20mV), the complexity of channel switching is relatively high.

The shape of curves indicates an increase in entropy along with scales. The most rapid boost is visible for the small scales. The average values of SampEn are also found to be greater at membrane depolarization. This suggests that the conformational space within the open and closed states may not be symmetric, as well as the changes in probabilities of switching between the recognized substates are not analogical when one deeply lowers or increases membrane potential. In general, the conformational dynamics seems to be more complex when the channel is activated by voltage. In other words, the complexity of switching between different channel substates increases when the membrane potential is high, which leads to a decrease of the system predictability and, in turn, an increase of the sample entropy.

4.0.1. Results of statistical analysis of MSE

The normality test of the entropy values was calculated via Shapiro–Wilk formula. The values of entropy are investigated at each scale. In depolarization, except for the scales 9 and 10, all entropy values are characterized by a normal distribution ($p > 0.05$ calculated via Shapiro–Wilk formula at selected significance level $\alpha = 0.05$). For depolarization state, the t-test and the non-parametric ANOVA Kruskal–Wallis statistic calculated for the scales 9 and 10 showed no statistically significant differences between the entropy values ($p > 0.05$) assigned to the potential values (20, 40, and 60 mV) (see Fig. 7b). The other situation occurs for hyperpolarization, where the statistics rejected the hypothesis about the normal distribution of entropy values for most cases. The Kruskal–Wallis test showed statistically significant differences between entropy values for all scales ($p < 0.05$) (see Fig. 7a). In addition, a comparison of depolarization, and hyperpolarization states was performed. As the non-parametric analog of the t-test for comparison between two independent groups, the Mann–Whitney U test was used. The hyper-, and depolarization at each stage of potential, yielded the

following results:

- $|U_{20mV}|$: normal distribution of SampEn values detected for all scales via Shapiro–Wilk test ($p > 0.05$). The t-test confirmed the statistically significant difference between the stage of hyper- and depolarization only for the small range of scales 1–6 ($p < 0.05$).
- $|U_{40mV}|$: the hypothesis about the normal distribution of SampEn rejected for all scales. The Mann–Whitney U statistics confirmed statistically significant differences between hyper- and depolarization ($p < 0.05$ for all the scales).
- $|U_{60mV}|$: the hypothesis about the normal distribution of SampEn rejected for all the scales. Mann–Whitney U statistics confirmed statistically significant differences between hyper- and depolarization ($p < 0.05$ for all the scales).

5. Separation of data at different membrane potentials

To study the differences between the MSE entropy curves in a more detailed way, we considered all the scales. Next, we implemented the machine learning techniques to find the numerical point separability index between the classes representing the experimental data obtained at different membrane potentials. We implemented the Support Vector Machine method and K-Nearest Neighbors technique for evaluating the classification accuracy between the different potential due to Multiscale Entropy values. The prediction of accuracy is described by Eq. 9. True Positives (TP) represent the samples that were classified correctly as right. True Negatives (TN) are classified correctly as wrong. Likewise,

we have the False Negative (FN), and False Positive (FP).

$$Acc = \frac{TP + TN}{TP + TN + FN + FP} \cdot 100\% \quad (9)$$

For the visualization of the separability of points, we used the decision surface of multi-class plots. Fig. 8 characterizes the separation of data at different membrane polarization states according to MSE entropy values. The individual points characterizes the MSE values, respectively for $MSE_{\tau \in [10,20]}$ for ordinate axis, and $MSE_{\tau \in [0,10]}$ for the abscissa. What is necessary to emphasize, the points in Fig. 8 are shifted and do not characterize the precise entropy values due to the earlier data standardization, which is required for the machine learning algorithm. The subtraction of the mean and division by the standard deviation artificially shifted the points towards the negative values. The better separability is visible for the data registered at membrane hyperpolarization (see the statistical analysis, Section 4.0.1). The different color points represent each class of potential at hyper-, and depolarization state separately: -20 mV, -40 mV, -60 mV for hyperpolarization (see Fig. 8 and 20 mV, 40 mV, 60 mV for depolarization (see Fig. 8b). The background colors show the decision surfaces. For each of the N classes, a binary classifier between one class and the others is constructed. We can observe better separability and a greater value of classifier accuracy for the channel currents recorded at different states of membrane hyperpolarization ($SVM_{score} = 0.753$) than for the cases at membrane depolarization ($SVM_{score} = 0.404$), which can also be visible directly from MSE curves dependency (see Fig. 7). At each stage of membrane depolarization, we observe a high voltage activation of the channel.

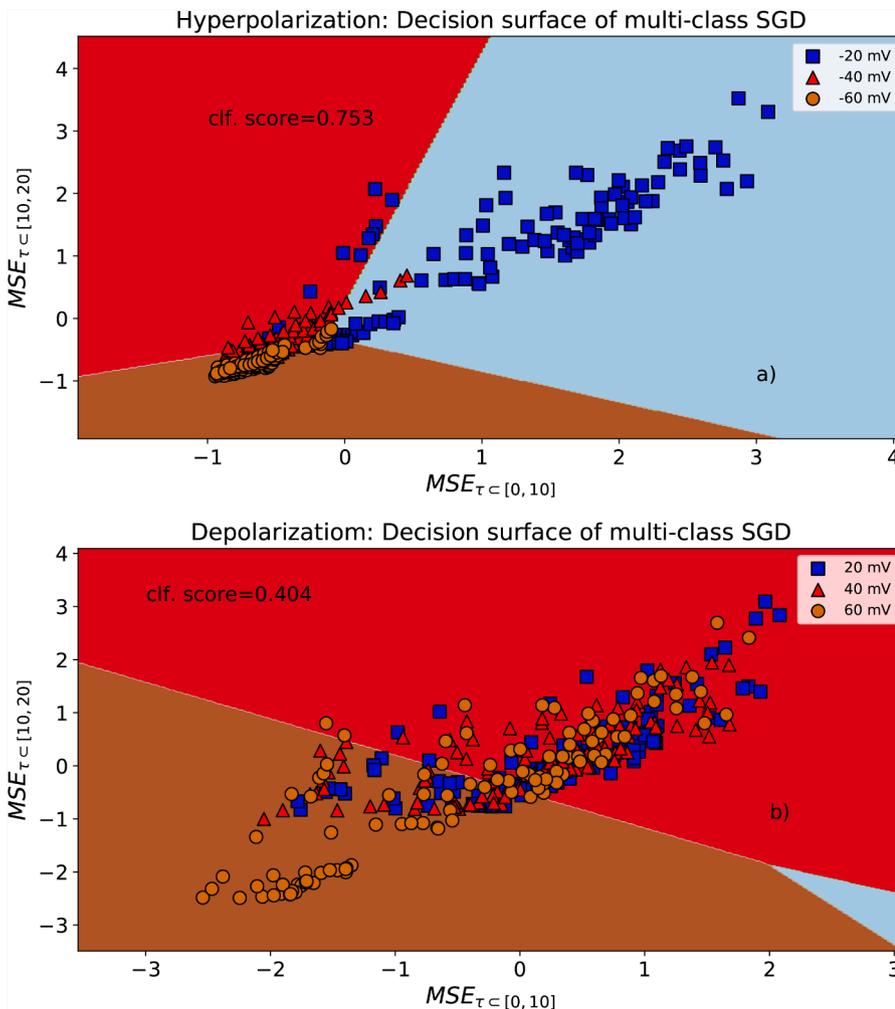


Fig. 8. The Decision Surfaces of Support Vector Machine assisted by Support Vector Machine algorithm optimized by the Stochastic Gradient Descent Classifier: panel a characterizes the points assigned to the membrane hyperpolarization; subfigure b presents the potential values recognized as membrane depolarization. The points represent the class of -60 mV (brown circles), -40 mV (red triangles) and -20 mV (blue squares) for depolarization; 60 mV (brown circles), 40 mV (red triangles) and 20 mV (blue squares) for hyperpolarization. Each point in the plane is characterized by 2 entropy values: the ordinate axis describe the MSE_{τ} for $\tau \in [0, 10]$; the abscissa axis represents MSE_{τ} for $\tau \in [10, 20]$. The process of standardization shifts the data points toward negative values.

Especially in the case of 40 mV and 60 mV, the open state probability approach its maximal value. In those terms, the channel's voltage sensor is located possibly close to the outer side mitochondrial membrane, and there is a minor variation within the available conformational states of the channel. In consequence, at the high voltage, the single-channel current traces have similar characteristics. Table 1 summarizes these results, supplementing them with the results of the KNN classifier. Both methods (KNN and SVM) give consistent results after implementation of k-fold cross-validation method which is a procedure that let us obtain the average results of different configuration of test and training data. Table 2.

Fig. 9 compares the hyper- and depolarization states at different values of membrane potential. We can observe an excellent separation between the states for $|U_{40mV}|$ and $|U_{60mV}|$ with the accuracy of classification 0.988 and 0.967. The most difficult is to distinguish between states at $|U_{20mV}|$ (the rate of about 0.58). The results calculated by both classifiers are presented in Table 1.

6. Discussion

PSD and MSE analyses both confirmed the complex, pink noise characteristics of the system's dynamics. The apparent nonlinear features of the ion channel system strongly depend on the sampling rate at patch-clamp recording. Thus, a crucial preliminary task before starting the numerical analysis of the electrophysiological experiment is to appropriately choose the sampling rate to capture the system's dynamics in real-time. On the one hand, it is important not to reduce the information about the conformational dynamics of the channel protein by imposing too low a probing frequency. On the other hand, the oversampling causes that each data point to carry less useful information than it could bring when the sampling rate would approach the suitable timescale of the system's dynamics.

Our results showed that such a compromise is obtained for the sampling of 10 kHz. In terms of probing frequency of 10 kHz (or higher), the three distinct segments describing pink noise dynamics typical for ion channels [32] become visible within the PSD plot. The MSE plots suggest a significant loss in information for signal probing of over 10 kHz (oversampling).

This result can be counterintuitive since the limited temporal resolution of measurement may imply failure to detect brief channel openings and shutoffs [35]. According to the studies presented in [36] increasing sampling resolution from 22 to ca. 8 μ s doubled the number of detected rapid shutoffs (shorter than 100 μ s). Nevertheless, the results [36] refer to an acetylcholine receptor that is structurally and functionally not as complex as the mitoBK channel. The possible mutual interactions between the ligand-sensing domain, voltage-sensing domain, and the channel gate in the mitoBK channel (that can be analogous to the BK channel [37]) may result in shifting the time scale of gating towards larger values.

The signal in the form of single-channel currents is a composition of several large- and small-scale components influencing pore-gate dynamics, including not only a sole channel protein fluctuations stemming from the equilibrium protein dynamics itself, but the proteins conformational diffusion is deeply affected by other processes within the membrane patch [31,32]. The analysis of the patch-clamp recordings employing PSD and MSE allowed us to highlight the rapid and slow components of the analyzed processes (as different regimens within the PSD and MSE plots, independently). It turns out that in small scales 1–4,

Table 1

Accuracy score calculated by the KNN and SVM compared for MSE at different membrane potential states.

Membrane potential	Acc_{KNN}	Acc_{SVM}
Depolarization (20 mV, 40 mV, 60 mV)	0.420	0.404
Hyperpolarization (–20 mV, –40 mV, –60 mV)	0.758	0.753

Table 2

Classifier score calculated by the KNN and SVM compared for MSE at different membrane potential states.

Depolarization vs Hyperpolarization	Acc_{KNN}	Acc_{SVM}
20/–20	0.602	0.580
40/–40	0.988	0.988
60/–60	0.960	0.967

one can observe a significant gain of entropy (and complexity of system dynamics), which is probably most affected by the rapid components as a movement of the channel helices. This gain may be explained by the fact that the random fluctuations don't exhibit too high complexity [38]. Thus, at the smallest timescale, when the signal can be mostly biased by thermal fluctuations of pore-forming helices of the channel, the observed entropy is not as high as in the larger time scales for which the complex channel dynamics seems to be relatively more complex. In turn, the low, almost linear increase in entropy at larger scales suggests that some new structures are still unraveled when one increases the time scale of process observation, which resembles the presence and impact of large-scale processes, like the ones which refer to the state of the membrane.

Referring to the characteristics of conformational dynamics of mitoBK channels in fibroblasts during voltage activation, one can observe that the most complex structure of the signal corresponds to the intermediate voltage values. In those terms, the channel may quite freely change the conformations in a relatively intricate way, which may be caused by relatively small energetic barriers separating the substates within the open and closed manifolds of channel states. In contrast, when the voltage sensor is at deep depolarization or hyperpolarization, the channel tends to switch in a more systematic way. The energetic landscape of the system predefines a set of highly preferred channel conformations and the most probable method of switching between them. Thus the complexity of single-channel currents is lower in those terms. Interestingly, the system of available channel conformations is more complicated in its voltage-activated state compared to the non-activated one at membrane hyperpolarization.

In this work, we showed the utility of PSD, SEN, and MSE in preliminary studies of conformational dynamics of ion channels. This is particularly important in channels where the molecular structure is not yet available, as in mitoBK channels in complex with β 3 subunits in fibroblast cells. The presented methodology can be considered a powerful tool for analyzing single-channel recordings, which gives a piece of complementary information to the standard kinetic analysis. The kinetic analysis enables the average description of the system dynamics, e.g., calculation of open state probability, mean dwell-times of the recognized states, and the switching rates. The nonlinear methods calculated across many timescales allow for a global look at the system and extract the information about the component processes that effectively affect the observed dynamics at different scales. What is also worth emphasizing, a good separation of analyzed states received via the Multiscale Entropy indicates a tremendous future potential for implementing information entropy techniques and studying the ion channels activity, especially when combined with machine learning methodology.

CRedit authorship contribution statement

Lukasz Machura: Conceptualization, Methodology, Software, Formal analysis, Writing - original draft, Writing - review & editing, Visualization, Supervision. **Agata Wawrzekiewicz-Jalowiecka:** Data curation, Writing - original draft. **Piotr Bednarczyk:** Investigation, Resources. **Paulina Trybek:** Conceptualization, Methodology, Software, Formal analysis, Writing - original draft, Writing - review & editing, Visualization.

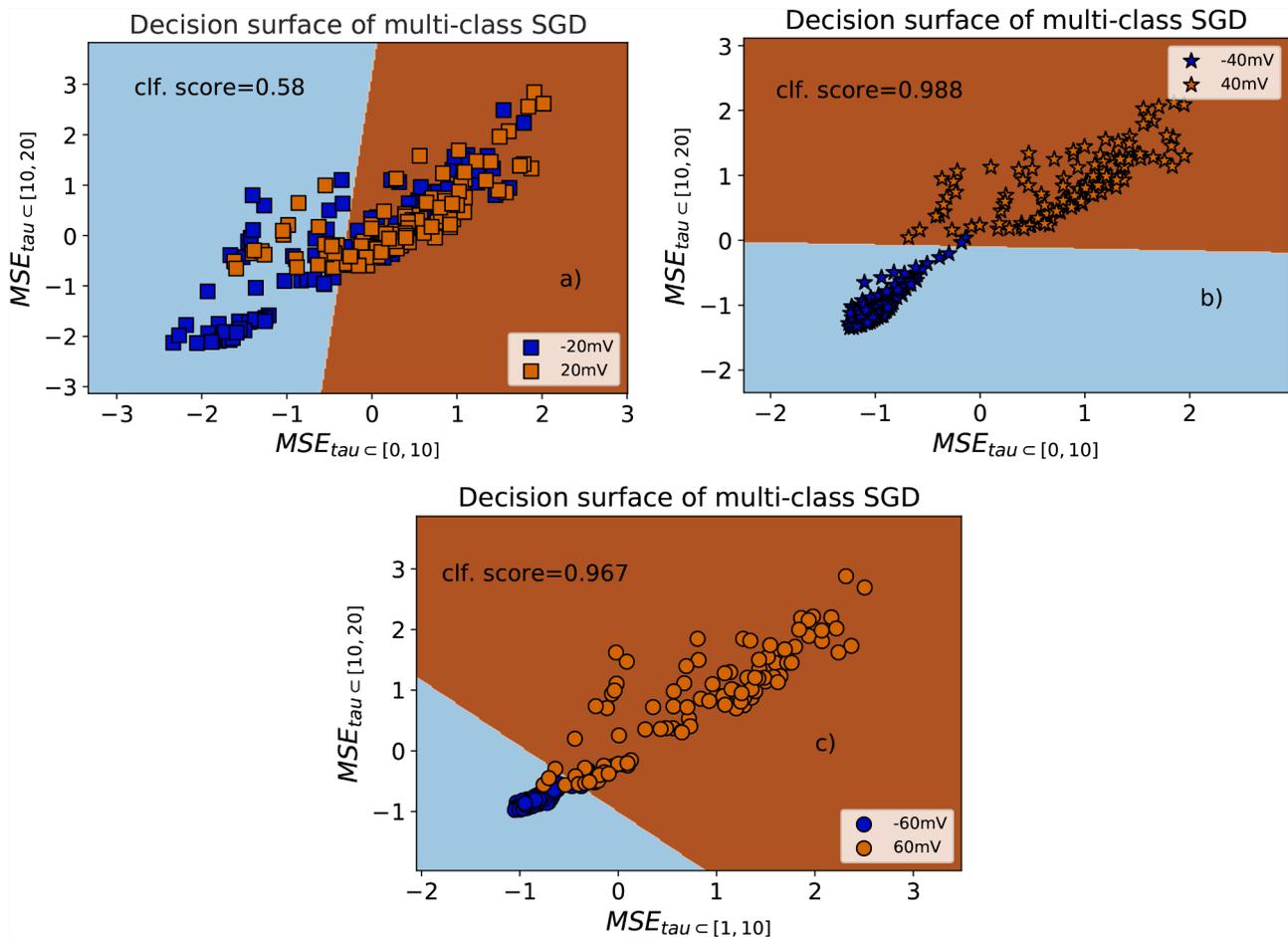


Fig. 9. The Decision Surfaces of Support Vector Machine Classifier assisted by the Stochastic Gradient Descent. Comparison of hyper- and depolarization state: figure a characterizes the $U = -20$ mV; panel b compares state for $U = -40$ mV; panel c is assigned to the potential state of $U = -60$ mV. Each point in the plane is characterized by 2 entropy values: the ordinate axis describe the MSE_{τ} for $\tau \in [0, 10]$; the abscissa axis represents MSE_{τ} for $\tau \in [10, 20]$.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Agata Wawrzkiwicz-Jalowiecka was supported by the Silesian University of Technology Grant for young researchers 04/040/BKM21/0177 (statute project). All authors declare no conflicts of interests.

References

- [1] N. Johnson, *Simply complexity: A clear guide to complexity theory*, Simon and Schuster, 2009.
- [2] C.E. Shannon, A mathematical theory of communication, *Bell Syst. Techn. J.* 27 (3) (1948) 379–423.
- [3] J.M. Sorrell, A.I. Caplan, Fibroblast heterogeneity: more than skin deep, *J. Cell Sci.* 117 (5) (2004) 667–675.
- [4] J. Janda, V. Nfonsam, F. Calienes, J.E. Sligh, J. Jandova, Modulation of ros levels in fibroblasts by altering mitochondria regulates the process of wound healing, *Arch. Dermatol. Res.* 308 (4) (2016) 239–248.
- [5] A. Stunova, L. Vistejnova, Dermal fibroblasts—a heterogeneous population with regulatory function in wound healing, *Cytokine Growth Factor Rev.* 39 (2018) 137–150.
- [6] A. Kicinska, B. Augustynek, B. Kulawiak, W. Jarmuszkiwicz, A. Szewczyk, P. Bednarczyk, A large-conductance calcium-regulated k^+ channel in human dermal fibroblast mitochondria, *Biochem. J.* 473 (23) (2016) 4457–4471.
- [7] M. Zada, U. Pattamatta, A. White, Modulation of fibroblasts in conjunctival wound healing, *Ophthalmology* 125 (2) (2018) 179–192.
- [8] R.P. Kampa, A. Kicinska, W. Jarmuszkiwicz, M. Pasikowska-Piwko, B. Dolegowska, R. Debowska, A. Szewczyk, P. Bednarczyk, Naringenin as an opener of mitochondrial potassium channels in dermal fibroblasts, *Exp. Dermatol.* 28 (5) (2019) 543–550.
- [9] E. Balderas, J. Zhang, E. Stefani, L. Toro, Mitochondrial $bkca$ channel, *Front. Physiol.* 6 (2015) 104.
- [10] H. Singh, R. Lu, J.C. Bopassa, A.L. Meredith, E. Stefani, L. Toro, $mitobkca$ is encoded by the $kcna1$ gene, and a splicing sequence defines its mitochondrial location, *Proc. Nat. Acad. Sci.* 110 (26) (2013) 10836–10841.
- [11] X. Tao, R. MacKinnon, Molecular structures of the human $slc1$ k^+ channel in complex with $\beta 4$, *Elife* 8 (2019), e51409.
- [12] E. Neher, B. Sakmann, Single-channel currents recorded from membrane of denervated frog muscle fibres, *Nature* 260 (5554) (1976) 799–802.
- [13] M. Costa, A.L. Goldberger, C.-K. Peng, Multiscale entropy analysis of complex physiologic time series, *Phys. Rev. Lett.* 89 (6) (2002), 068102.
- [14] M. Costa, A.L. Goldberger, C.-K. Peng, Multiscale entropy analysis of biological signals, *Phys. Rev. E* 71 (2) (2005), 021906.
- [15] T. Angsuwatanakul, J. O'Reilly, K. Ounjai, B. Kaewkamnerdpong, K. Iramina, Multiscale entropy as a new feature for eeg and fmrs analysis, *Entropy* 22 (2) (2020) 189.
- [16] D. Dharmapranj, L. Dykes, A.D. McGavigan, P. Kuklik, K. Pope, A.N. Ganesan, Information theory and atrial fibrillation (af): A review, *Front. Physiol.* 9 (2018) 957.
- [17] P. Trybek, M. Nowakowski, J. Salowka, J. Spiechowicz, L. Machura, Sample entropy of semg signals at different stages of rectal cancer treatment, *Entropy* 20 (11) (2018) 863.
- [18] A. Wawrzkiwicz-Jalowiecka, P. Trybek, L. Machura, B. Dworakowska, Z. J. Grzywna, Mechanosensitivity of the bk channels in human glioblastoma cells: Kinetics and dynamical complexity, *J. Membr. Biol.* 251 (5–6) (2018) 667–679.
- [19] A. Wawrzkiwicz-Jalowiecka, P. Trybek, B. Dworakowska, L. Machura, Multifractal properties of bk channels' currents in human glioblastoma cells, *J. Phys. Chem. B.*
- [20] A. Wawrzkiwicz-Jalowiecka, P. Trybek, L. Machura, P. Bednarczyk, Dynamical diversity of mitochondrial bk channels located in different cell types, *Biosystems* 199 (2021), 104310.

- [21] J.L. Semmlow, B. Griffel, *Biosignal and medical image processing*, CRC Press, 2014.
- [22] P. Welch, The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms, *IEEE Trans. Audio Electroacoustics* 15 (2) (1967) 70–73.
- [23] A.N. Kolmogorov, New metric invariant of transitive dynamical systems and endomorphisms of lebesgue spaces, in: *Dokl. Akad. Nauk SSSR*, Vol. 119, 1958, pp. 861–864.
- [24] A.N. Kolmogorov, Entropy per unit time as a metric invariant of automorphisms, in: *Dokl. Akad. Nauk SSSR*, Vol. 124, 1959, pp. 754–755.
- [25] S.M. Pincus, Approximate entropy as a measure of system complexity, *Proc. Nat. Acad. Sci.* 88 (6) (1991) 2297–2301.
- [26] J.S. Richman, J.R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, *Am. J. Physiol.-Heart Circulatory Physiol.* 278 (6) (2000) H2039–H2049.
- [27] M. Costa, C.-K. Peng, A.L. Goldberger, J.M. Hausdorff, Multiscale entropy analysis of human gait dynamics, *Physica A: Stat. Mech. Its Appl.* 330 (1) (2003) 53–60.
- [28] W.A. Chaovalitwongse, Y.-J. Fan, R.C. Sachdeo, On the time series k-nearest neighbor classification of abnormal brain activity, *IEEE Trans. Syst., Man, Cybern.-Part A: Syst. Humans* 37 (6) (2007) 1005–1016.
- [29] N.E.M. Isa, A. Amir, M.Z. Ilyas, M.S. Razalli, The performance analysis of k-nearest neighbors (k-nn) algorithm for motor imagery classification based on eeg signal, in: *MATEC web of conferences*, Vol. 140, EDP Sciences, 2017, p. 01024.
- [30] H. Derksen, A. Verveen, Fluctuations of resting neural membrane potential, *Science* 151 (3716) (1966) 1388–1389.
- [31] K. Diba, H.A. Lester, C. Koch, Intrinsic noise in cultured hippocampal neurons: experiment and modeling, *J. Neurosci.* 24 (43) (2004) 9723–9733.
- [32] S.M. Bezrukov, M. Winterhalter, Examining noise sources at the single-molecule level: 1/f noise of an open maltoporin channel, *Phys. Rev. Lett.* 85 (1) (2000) 202.
- [33] Z. Siwy, A. Fuliński, Origin of 1/f α noise in membrane channel currents, *Phys. Rev. Lett.* 89 (15) (2002), 158101.
- [34] P.M. Riechers, J.P. Crutchfield, Fraudulent white noise: Flat power spectra belie arbitrarily complex processes, *Phys. Rev. Res.* 3 (1) (2021), 013170.
- [35] L. Sivilotti, D. Colquhoun, In praise of single channel kinetics, *J. Gen. Physiol.* 148 (2) (2016) 79–88.
- [36] N. Mukhtasimova, C.J. daCosta, S.M. Sine, Improved resolution of single channel dwell times reveals mechanisms of binding, priming, and gating in muscle achr, *J. Gen. Physiol.* 148 (1) (2016) 43–63.
- [37] Y. Geng, Z. Deng, G. Zhang, G. Budelli, A. Butler, P. Yuan, J. Cui, L. Salkoff, K. L. Magleby, Coupling of ca²⁺ and voltage activation in bk channels through the α helix/voltage sensor interface, *Proc. Nat. Acad. Sci.* 117 (25) (2020) 14512–14521.
- [38] M. Costa, A. Goldberger, C.-K. Peng, Multiscale entropy to distinguish physiologic and synthetic rr time series, in: *Computers in Cardiology, 2002, IEEE, 2002*, pp. 137–140.