STATISTICAL CLASSIFICATION OF THE STATES OF BIOLOGICAL CELLS TREATED WITH CARBON NANOTUBES BASED ON AFM-IMAGES OF CELL SURFACE

I.E. Starodubtsev[1,2], Yu.S. Kharin[1,2], M.N. Starodubtseva[3,4]
[1] *Research Institute for Applied Problems of Mathematics and Informatics*
[2] *Belarusian State University*
[3] *Gomel Medical State University*
[4] *Institute of Radiobiology National Academy of Sciences of Belarus*
Minsk, Gomel, BELARUS
e-mail: `istarodubtsev.science@gmail.com`

Abstract

The method of statistical classification of biological cells, treated with carbon nanotubes, based on images of cell surface obtained with an atomic force microscope (AFM) is proposed. Each scan line of the original AFM image is considered as a random sequence realization, and the discrete Fourier transform is applied to compute its spectral features. After smoothing, the map of spectral estimates is formed. The informative features are computed as the medians for the set of the spectrogram values. Classification of four classes of cells (control and treated with carbon nanotubes, after 1 hour and 24 hours of incubation) was carried out by the obtained informative features using the decision trees method. The proposed method provides a sufficiently high accuracy classification of cell states after the treatment with carbon nanotubes.

Keywords: data science, AFM-image, carbon nanotube, cell surface, classification

1 Introduction

Atomic force microscopy (AFM) is a modern method of biomedical research which allows studying the relief and the physicomechanical properties of biological cell surfaces at nanoscale level, it makes possible the determination of their type and condition based on complex statistical data [1,2,3].

The aim of this paper it to solve the problem of the statistical classification of AFM-images (the microscale maps of cell surface mechanical properties) of biological cells (glial cells) treated with carbon nanotubes.

The samples of rat glioma cell (C6 cell line) treated with the DNA-single-walled carbon nanotube (NT) complex (incubation time was 1 and 24 hours) were kindly provided by the Biophysics Department of Physics Faculty of Belarusian State University. The structure and distribution of mechanical properties over the cell surface change in time dependent on treatment of cells with NT. Images of cell surfaces were recorded using AFM NT-206 in Research Laboratory of the Gomel State Medical University. For the analysis, the maps of lateral forces of the cell surface of size of 2.5 $\mu m \times 2.5$ μm (256×256 points) were used. AFM-images were processed by the software developed by us using the fftw library [4].

2 Mathematical model

The AFM-image of cell surface is a two-dimensional array *z=z(x,y)*, where x is the vertical coordinate, y is the horizontal coordinate, $x,y \in \{1,2,...,N\}$; z is the value of the sliding friction force at the point *(x, y)*. An AFM-image of size of N $\times$ N points can be considered as a set of N one-dimensional arrays $z = z^{(y)}(x)$ of N points each, located at a distance of a scanning step along the y axis (N is an even number, in experiments N=256).

Original AFM-images were normalized by dividing the values *z(x,y)* by 10^3. Instead of the initial values of z at each y were considered z' is the differences between adjacent values along the x axis, divided by the value of the standard deviation for each line:

$$z'(x,y) = \frac{z(x+1,y) - z(x,y)}{\sqrt{D^{(y)}}},\ x \in \{1, 2, ..., N-1\}.$$

The data obtained in this way corresponds to four classes: Ω_1 - NT-1h (with NT, after 1 hour), Ω_2 - control-1h (without NT, after 1 hour), Ω_3 - NT-24h (with NT, after 24 hours), Ω_4 - control-24h (without NT, after 24 hours).

3 Informative features

Each one-dimensional array $z = z^{(y)}(x)$ with fixed y can be considered as an realization of a random sequence $z = z_x^{(y)}$, $x \in \{1,2,...,N\}$, for which a discrete Fourier transform can be applied:

$$X^{(y)}(\omega_k) = \sum_{n=1}^{N} (z_n^{(y)} - \bar{z}^{(y)}) e^{-j\frac{2\pi kn}{N}},\ k = 0, 1, ..., N-1,$$

where $\bar{z}_n^{(y)} = \frac{1}{N}\sum_{n=1}^{N} z_n^{(y)}$ is the sample mean on the x axis with fixed y, $\omega_k = 2\pi\frac{k}{L}$ is frequency, L is the length of analyzed interval along the x axis.

Based on the sample spectrum $X^{(y)}(\omega_k)$, we calculated the periodogram

$$r^{(y)}(\omega_k) = |X^{(y)}(\omega_k)|^2,$$

and, smoothing it using the Daniel window with a width of m (m=5), we obtained spectral density estimates $R^{(y)}(\omega_k)$ [5]. Estimates $R^{(y)}(\omega_k)$ at $k=\frac{N}{2}+1,...,N-1$ were excluded from our consideration as they repeated the values at $k=0,...,\frac{N}{2}-2$.

For each frequency ω_k we calculated the medians of the spectral density using N values along the y axis:

$$\tilde{R}(\omega_k) = Med\{R^{(1)}(\omega_k), ..., R^{(N)}(\omega_k)\},\ k = 0, ..., \frac{N}{2} - 1,$$

that were used as informative features of the original AFM image. We will call $\tilde{R}(\omega_k)$: $k = 0, ..., \frac{N}{2}-1$ as the spectrogram of AFM image of cell surface under this study. Values $\tilde{R}(\omega_k) : k = 71, ..., 128$ were excluded from our consideration because the corresponding periods $T_k = \frac{2\pi}{\omega_k}$ were smaller than the scanning step.

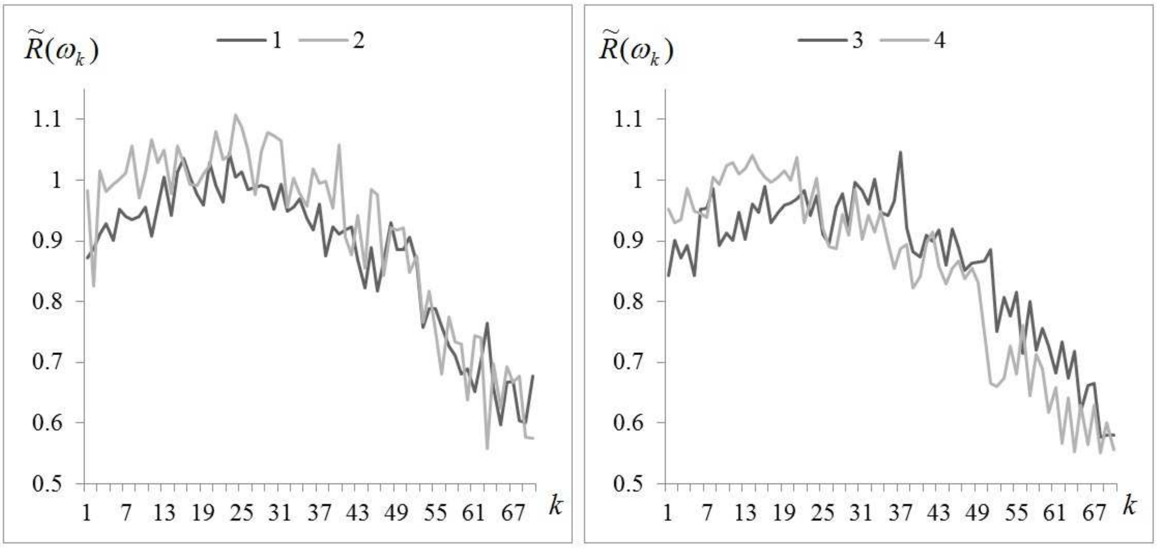

Figure 1 –Averaged over entire sample spectrograms of AFM images of surfaces for classes Ω_1 (1), Ω_2 (2), Ω_3 (3) and Ω_4 (4).

4 Statistical classification

We used the decision trees algorithm for classification, to build decision trees we used the C&RT algorithm [6]. The choice of features for constructing a decision tree was carried out using the Gini criterion [6]. Since the training sample is not large enough, the choice of the optimal size of the classification tree was determined using cross-validation [7]. The examination sample size was 30% of the entire sample size.

5 Numerical results

The training sample size was consist of 51 observations (the AFM images of 256 × 256 dots). Examples of averaged spectrograms for pairs of classes $\{\Omega_1, \Omega_2\}$ and $\{\Omega_3, \Omega_4\}$ are shown in Figure 1.

Table 1 presents the values of the accuracy of correct classification.

6 Conclusion

The proposed statistical classification method based on spectral features of the AFM images presenting the maps of mechanical properties of rat glioma cells provides a sufficiently high classification accuracy of cell states after the treatment with carbon nanotubes. The results can be used in the study of the effects of carbon nanotubes on biological cells.

Table 1. Accuracy of correct classification of cell states using the decision trees algorithm, %

Classification for pair of classes $\{\Omega_1, \Omega_2\}$	Ω_1	Ω_2
	93.33%	**75.00%**
Classification for pair of classes $\{\Omega_3, \Omega_4\}$	Ω_3	Ω_4
	72.73%	**92.31%**
Classification for pair of classes $\{\Omega_1, \Omega_3\}$	Ω_1	Ω_3
	80.00%	**100.00%**
Classification for pair of classes $\{\Omega_2, \Omega_4\}$	Ω_2	Ω_4
	66.67%	**92.31%**

References

[1] Starodubtseva M.N., Starodubtsev I.E. , Starodubtsev E.G. (2017). Novel fractal characteristic of atomic force microscopy images. *Micron*. Vol. **96**, pp. 96-102.

[2] Starodubtseva M.N., Mitsura E.F., Starodubtsev I.E., Chelnokova I.A., Yegorenkov N.I., Volkova L.I., Kharin Yu.S. (2018). Nano- and microscale mechanical properties of erythrocytes in hereditary spherocytosis. *J. Biomech.* Vol. **83**, pp. 1-8.

[3] Kharin Yu.S. (1996). *Robustness in Statistical Pattern Recognition*. Kluwer, Dordrecht.

[4] FFTW package: http://www.fftw.org/ (accessed 10 April 2019).

[5] Anderson T. W. (1971). *The Statistical Analysis of Time Series*. J. Wiley, N. Y.

[6] Harrington P. (2012). *Machine Learning in Action*. Manning, N. Y.

[7] Hastie T., Tibshirani R., Friedman J. H. (2009). *The Elements of Statistical Learning* . Springer Publ., N. Y.