Luciferase-based bioassay for rapid pollutants detection and classification by means of multilayer artificial neural networks

Ivan A. Denisov

Laboratory of bioluminescent biotechnologies, Siberian Federal University, 660041, Krasnoyarsk, Russia

Abstract

Biosensors for rapid environmental pollution detection can be designed with biomodule based on the bacterial bioluminescent system. Usually this method returns total value of toxicity and does not allow to distinguish pollutants types. Herein we demonstrate the classification of pollutants by the kinetic analysis utilizing artificial neural networks with multilayer perceptron architecture. The kinetics of light emission of NAD(P)H:FMN-oxidoreductase-luciferase bioluminescent reaction was measured for clean water and in the presence of three environment pollutants (1,4-benzoquinone, copper sulfate and 1,3-dihydroxybenzene) separately with various concentrations. The efficiency of using multilayer perceptron with sigmoid activation function for processing of kinetics of light emission was estimated. It was shown that multilayer perceptrons allowing to distinguish pollutant class and concentration after sufficient training. The architecture consisted of 61 inputs neurons, 3 hidden layers and 3 output neurons was found optimum in sense of learning time for classification of three pollutants. Usage of simplest activation function sigmoid and backpropagation method for multilayer perceptron teaching providing the results been useful for smart signal processing in computational modules of biosensors.
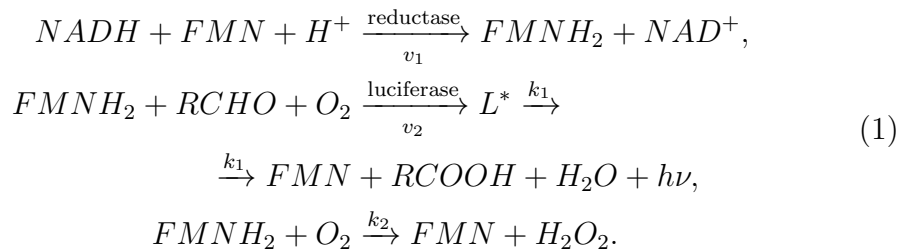
Keywords: Bioluminescence, Luciferase, Bioassay, Artificial Neural Networks, Perceptron, Machine Learning

1. Introduction

Bioluminescence inhibition assays based on luminous bacteria [1, 2, 3] and components of their bioluminescence system [4, 5] are used for detection of organic and inorganic pollutants in water. The method of pollutants detection is based on their interaction with bacterial luciferase and other components of bacterial luminescence system, that is leading to quenching of light emission [6]. These classical methods possess high sensitivity to various pollutants, however they do not allow to identify pollutant type [7].

Bioluminescence *in vitro* in extracts of luminous bacteria requires two enzymes: NAD(P)H:FMN-oxidoreductase for proton transfer and luciferase for light emission catalysis [8, 9]. The reaction of light emission is described by the equations:

$$NADH + FMN + H^+ \xrightarrow[v_1]{\text{reductase}} FMNH_2 + NAD^+,$$

$$FMNH_2 + RCHO + O_2 \xrightarrow[v_2]{\text{luciferase}} L^* \xrightarrow{k_1}$$

$$\xrightarrow{k_1} FMN + RCOOH + H_2O + h\nu,$$

$$FMNH_2 + O_2 \xrightarrow{k_2} FMN + H_2O_2.$$

(1)

Different pollutants often have various mechanisms of interaction with components of this bacterial luminescent system [10], that can lead to the different light emission kinetics. Classification could be made by means of shape analysis. However, it is extremely complicated, owing to the complexity of theoretical models describing interaction of various chemicals with components of luminescent system. Even the first stage of the reaction has several speed coefficients which wary from one bacteria strain to another [11, 12]. Some speed coefficients are still unknown or floating, because of natural reagents instability of NAD(P)H:FMN-oxidoreductase-luciferase bioluminescent system. That is delimiting the development of full mathematical model suitable for fitting methods and also leading to the idea of using machine learning methods for signal processing.

There are two mainstream machine learning methods suitable for classification and regression analysis: artificial neural networks (ANN) and support vector machines (SVM). Detailed comparison of these competitive methods was out of focus of our study. We wanted to keep computational algorithms "as simple as possible" for future usage in computational signal processing

modules of biosensors and point-of-care devices [13]. That is why we decided to evaluate applicability of the welled-studied multilayer perceptron (MLP) architecture with computationally cheap sigmoid activation functions and the backpropagation method for its teaching [14].

Basic concepts of artificial neural networks (ANN) modeling and its application in pharmaceutical research is described in the work [15]. Trained ANN should represent phenomenological model of bioluminescence decay process using available experimental data about kinetics. The efficiency of ANN for multicomponent kinetic determinations was proved in a series of works [16, 17]. ANN is suitable for the detection and classification of neurotoxins [18], the analysis of spectrum generated from an optical phenolic biosensor [19] and for processing the signals from bioelectronic tongue for urea detection [20]. The applicability of ANN for analysis of bioluminescent inhibition kinetics was not described in the literature.

Several architectures were required to be compared to choose the minimal number of hidden layers in MLP for obtaining good predictability and less teaching time with sigmoid activation function. Despite the fact that the universal approximation theorem postulates single hidden layer as sufficient for any mapping from input to output layer, the theorem does not say that a single hidden layer is optimum in the sense of learning time, ease of implementation, or (more importantly) generalization [21].

The purpose of our work was to check the suitability of multilayer perceptrons for estimation of the type and the concentration of pollutant based on light emission kinetics from the two-enzyme bioluminescent reaction.

2. Materials and methods

2.1. Measurement of the light emission kinetics

Lyophilized bacterial luciferase (EC 1.14.14.3) from *Photobacterium leiognathi* (recombinant strain of *E. coli*) and NAD(P)H:FMN-oxidoreductase (EC 1.6.99.3) from *Vibrio fischeri* were produced at the Laboratory of Bacterial Bioluminescence (Institute of Biophysics, Siberian Branch of the Russian Academy of Sciences, Krasnoyarsk). Flavin mononucleotide (FMN, CHEBI:17621), reduced nicotinamide adenine dinucleotide (NADH, CHEBI:16908), tetradecanal ($C_{14}H_{28}O$, CHEBI:84067) were purchased from Merck.

The light emission kinetics was measured using microplate reader TriStar LB941 (Berthold, Germany) with two injectors. Solution for the first

injector consisted of 1.19×10^{-4} M of FMN, 9.46×10^{-6} M of tetrade-canal, 1.66×10^{-2} M of ethanol, 6.45×10^{-8} M of bacterial luciferase and 4.3 units of activity NAD(P)H:FMN-oxidoreductase per liter. One unit of NAD(P)H:FMN-oxidoreductase activity was defined as 1 μmol of NADH degraded per minute. Solution for the second injector initiated reactions, consisted of the 1.9×10^{-5} M of NADH. All solutions were prepared with 0.05 M potassium phosphate buffer (pH 6.8) purchased from Fluka.

Three classes of model pollutants with AR grade were used: quinone (1,4-benzoquinone, $C_6H_4O_2$, CHEBI: 16509), copper (II) sulfate ($CuSO_4$, CHEBI: 23414) and resorcinol (1,3-dihydroxybenzene, $C_6H_4(OH)_2$, CHEBI: 27810). Model pollutants were prepared at different concentrations (Table 1), which were chosen preliminary to have one oder of influence on the NAD(P)H:FMN-oxidoreductase-luciferase system activity.

Table 1: The initial concentrations of the model pollutants in wells

Quinone, $\times 10^{-5}M$	$CuSO_4$, $\times 10^{-4}M$	Resorcinol, $\times 10^{-2}M$
10.0	10.0	10.0
5.00	7.50	7.50
2.50	5.63	5.63
1.25	4.22	4.22
0.94	3.16	3.16
0.70	2.37	2.37
0.53	1.78	1.78
0.40	1.33	1.33
0.30	1.00	1.00
0.22	0.75	0.70
0.67	0.56	0.50
	0.42	
	0.32	
	0.24	
	0.18	

The pollutants were added in wells of white opaque 96-well plate by pipette in volume of 10 μL. Then just before measurement the reagents of bioluminescence reaction were added in well with two injectors of microplate reader with volumes of 50 μL each. So the total volume of liquid in each well before measurement was 110 μL. After reagents injection light emission kinetics was measured 61 times with 0.1 second exposure time and a half of a

second delays between measurements. Such kinetic curve was obtained twice for each concentration.

Temperature fluctuations can affect tetradecanal solubility and change amplitude of kinetic curves. To control this effect reference measurements were carried out in every third well on a plate (Fig. 1). To obtain reverence curve 10 μL of distilled water instead of pollutants had been added. Measured reference curves then were used for the normalization of kinetic curves inhibited by the pollutants. The averaging of many reference curves gave a "global" reference curve. The correction was made by the multiplication of each curve point to relation of value of a global reference value to an average of two values of neighboring reference curves according the equation:

$$\langle a \rangle = \frac{3}{2}\langle r \rangle \left(\frac{a_1}{2r_1 + r_2} + \frac{a_2}{r_1 + 2r_2} \right). \tag{2}$$
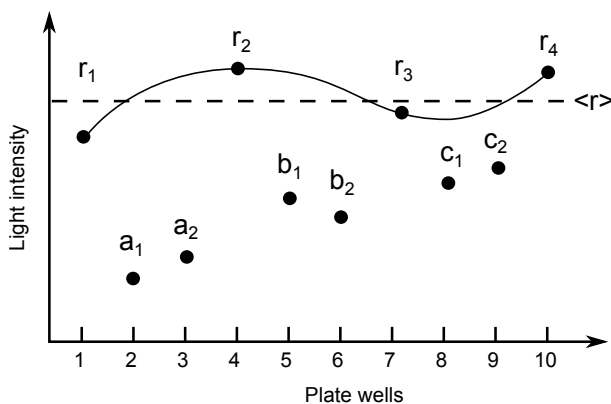

Figure 1: Schematic representation of reference value deviation. r_1, r_2, r_3, r_4 — light intensity for distilled water after some fixed time in different plate wells, a_1, a_2 — light intensity for sample a, b_1, b_2 — light intensity for sample b and c_1, c_2 — light intensity for sample c.

2.2. ANN training and evaluation

ANN had multilayer perceptron architecture with 61 inputs (one input for each point of kinetic curve) and 3 outputs for pollutant type determination. The arrays of inputs and outputs for ANN teaching were placed in two text documents. Each line of inputs contained results of luminescence measurements. Each line of outputs contained information in a form of "toxicity

vector". Components of such vector corresponded to presence of one of the pollutants: (1, 0, 0) — quinone, (0, 1, 0) — copper sulfate and (0, 0, 1) — resorcinol.

The sigmoid function was used as the activation function and the back-propagation method was used for ANN teaching.

The software for numerical experiments (perceptron training, prediction and evaluation) was developed on Component Pascal programming language in the framework BlackBox Component Builder (Oberon microsystems Inc., Switzerland) from the basic principles [22]. The developed application has been published under the open-source license (*http://ann.molpit.org*).
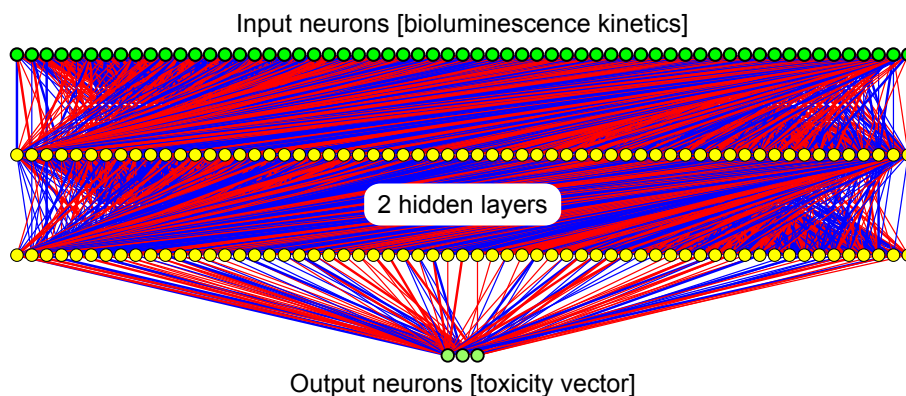

Figure 2: The example of the obtained perceptron for pollutants classification (blue lines are negative and red — positive weights).

The standard leave-one-out cross-validation (LOOCV) strategy was used to evaluate achieved ANN models. All curves were used in training of ANN except one (testing) which was used later for evaluation of prediction quality. Learning time, learning cycles and number of attempts to train one net were estimated separately with all the data for the MLP with from 1 to 6 hidden layers.

In the case of regression analysis there were concentrations of a pollutant in the output of perceptron. The orders of concentration values are very small (Table 1). The required target errors would be even less. During the regression analysis such small output for ANN can lead to number rounding errors while training and prediction. So the concentrations of copper sulfate were multiplied to 100 and concentrations of quinone were multiplied to 10000. All the kinetic curves were normalized by the division each point by

6

maximal intensity (213680 RLU).

3. Results and discussions

The obtained dependencies of the bioluminescence kinetics from pollutants type and concentration in the wells are shown in Fig. 3. The results showed that the dynamics of bioluminescence intensity can be specific to the pollutant type. The quinone kinetics demonstrates a delay in bioluminescence response before attaining peak intensity and this delay was found to increase with concomitant rise in its concentration. The rate of quenching of luminescence response increased with rise of copper sulfate concentration. Faster quenching changes inclination of curves while decreasing total amount of light emitted. Resorcinol decreases the maximal intensity, however light emission keeps approximately same rate over a period after rising until the end of test.

The difference in the light emission kinetics can be explained by the difference in the mechanochemical approach followed by the individual pollutants and various targets in bioluminescence reaction. Ideas about influence of phenols and quinones on triple enzyme systems had been discussed earlier [25], however there is still no mathematical model which can holistically explain and predict the type of pollutants based on unique bioluminescent kinetics. Despite the absence of mathematical model, in practice, existing correlation anyway can be used to determine pollutants classes and concentrations without fitting methods.

The obtained kinetic data was used to teach multi-layer perceptron employing backpropagation algorithm. The results of leave-one-out cross-validation for the perceptron configuration with 2 hidden layers is shown in Fig. 4. The deviations of "toxicity vectors" take place on boundary concentrations only. That happens because there are less example points around boundary concentrations for perceptron to learn from. That means that prediction is smooth in data range area, however is not effective in extrapolated areas.

Another perceptrons with 3 hidden layers were used to predict pollutants concentrations. The concentrations obtained by the prediction correlate with expected known values with 0.95 Pearson coefficient. The results of leave-one-out cross-validation is demonstrated in the Fig. 5.

The efficiency of using perceptron with from 1 to 6 hidden layers were compared (Fig. 6). The number of neurons in a hidden layers were equal to the number of inputs. Perceptron with 3 hidden layer was found to be
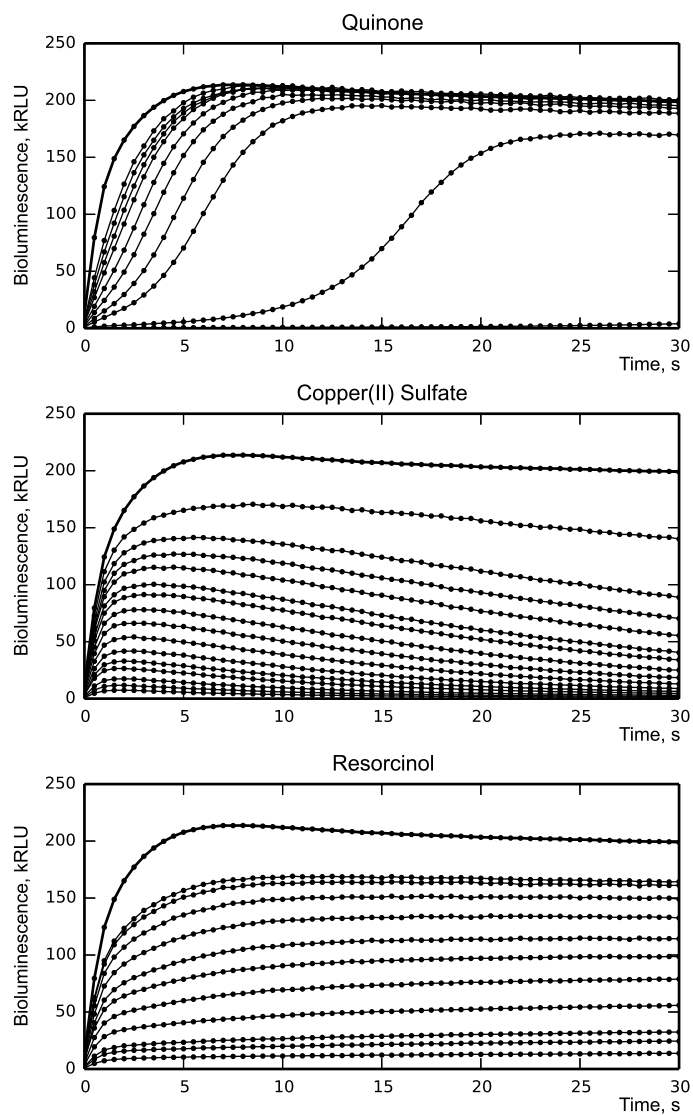
Figure 3: Bioluminescence kinetics in a presence of three model pollutants. The bold lines on the top are references. The concentrations in plate wells from top to bottom: Quinone (1,4-benzoquinone) $[0.15, 0.20, 0.27, 0.36, 0.48, 0.64, 0.85, 1.1, 2.3, 4.5] \times 10^{-6}$ M. Copper sulfate $[0.16, 0.22, 0.29, 0.38, 0.51, 0.68, 0.91, 1.2, 1.6, 2.1, 2.9, 3.8, 5.1, 6.8, 9.1] \times 10^{-5}$ M. Resorcinol (1,3-dihydroxybenzene) $[0.45, 0.64, 0.91, 1.2, 1.6, 2.1, 2.8, 3.8, 5.1, 6.8, 9.1] \times 10^{-3}$ M.

optimal in sense of learning time for pollutants classification. It was shown that this result are independent from the required prediction accuracy (target
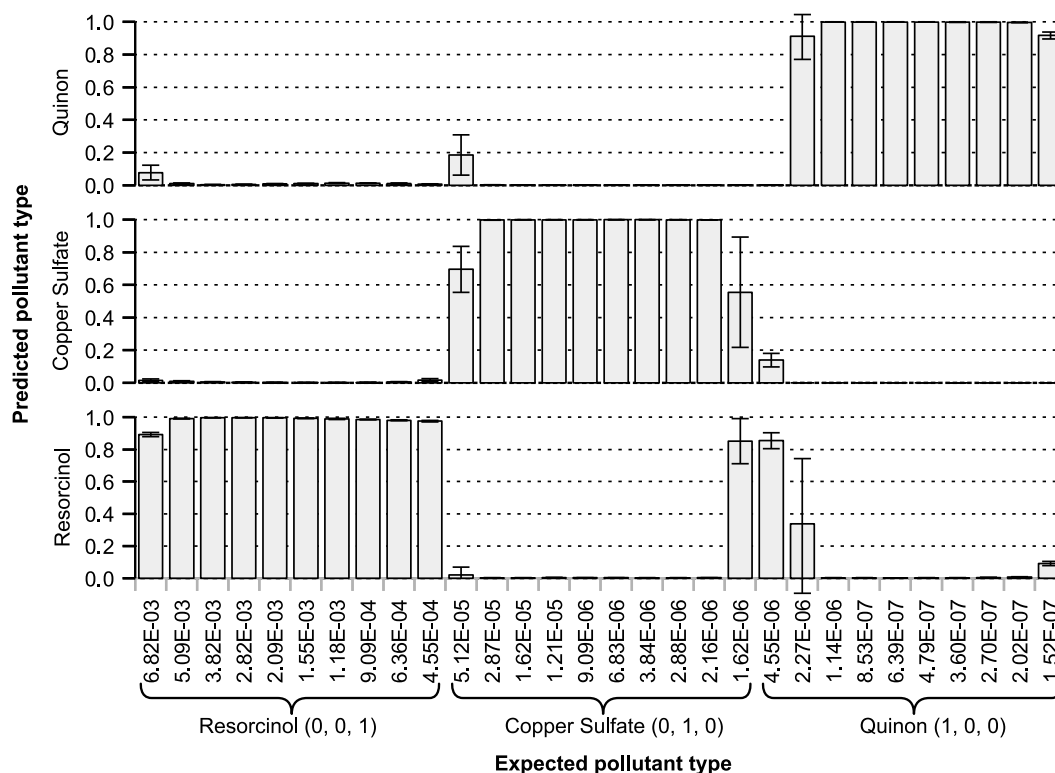
8

Figure 4: The results of leave-one-out cross-validation test for prediction of a pollutant type by MLP with 2 hidden layers. Columns: 1–10 — resorcinol expected, 11–20 — copper sulfate expected, 21–30 — quinone expected. The height of the bar representing the component of predicted "toxicity vector".

error).

We suggest that in case of more complex situations where pollutant types, their number and proportion in a sample increases, the amount of hidden layers and number of neurons in each layer should be increased. We propose to repeat developed evaluation algorithm to define optimum architecture for more complex cases.

4. Conclusion

The efficiency of using artificial neural networks in the analysis of bioluminescent signals for identification of different groups of environment pollutants (phenols, quinones, heavy metal ions) was estimated. The optimum multi-
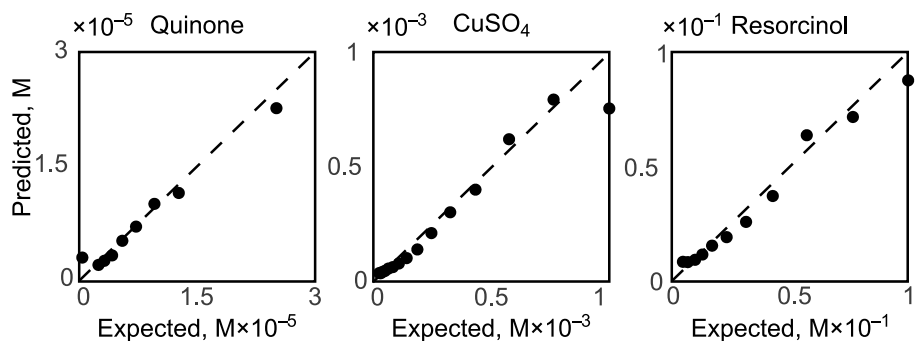
9

Figure 5: The results of leave-one-out cross-validation test for prediction of a pollutant concentration. In the figure shown the dependency of predicted concentration from expected. The correlation coefficients (Pearson) are: 0.99 — for quinone, 0.97 — for copper sulfate and 0.98 — for resorcinol.
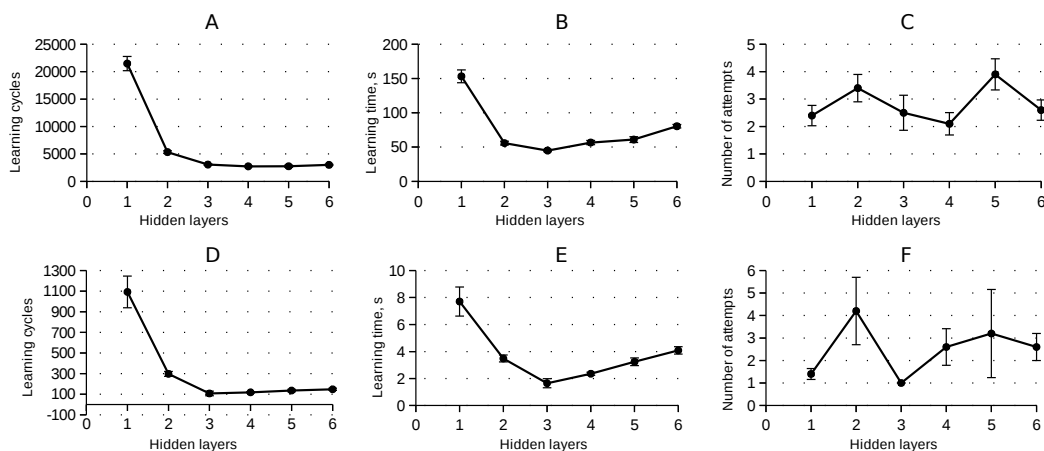

Figure 6: The characteristics of multilayer perceptron learning for pollutants classification depending on the amount of hiddel layers. A, B, C — the learning cycles, learning time and requared attempts to get one net in the case of 0.1 target error, D, E, F — the same characteristics in the case of 0.5 target error.

layer perceptron architecture has three hidden layers in sense of learning time that does not contradict to universal approximation theorem.

Results showed that artificial neural networks with multilayer perceptron architecture is suitable in identifying the type of pollutant and its concentration, if it has specific influence on luminescent kinetics. The results can be useful for rapid determination of the pollution source. For example, if there are several industries upstream with specific emissions, such method

will allow to distinguish in a minute which object emits pollutants.

The demonstrated method is data driven, however it solves the problem of kinetics analysis bypassing creation of mathematical model for chemical processes if there are enough experimental data for perceptron training. The machine learning methods can essentially expand application of inhibition bioluminescent analysis on the basis of bacterial bioluminescence for detection and identification of environmental pollutants.

Further research on using multilayer perceptron for bioluminescent kinetic analysis can by made for mixtures of pollutants of different types.

Acknowledgments

References

[1] S. Girotti, E. N. Ferri, M. G. Fumo, E. Maiolini, Monitoring of environmental pollutants by bioluminescent bacteria., Analytica chimica acta 608 (1) (2008) 2–29. doi:10.1016/j.aca.2007.12.008.

[2] C. J. Kelly, C. Hsiung, C. a. Lajoie, Kinetic analysis of bacterial bioluminescence, Biotechnology and bioengineering 81 (3) (2003) 370–378. doi:10.1002/bit.10475.

[3] S. E. Medvedeva, N. A. Tyulkova, A. M. Kuznetsov, E. K. Rodicheva, Bioluminescent bioassays based on luminous bacteria, Journal of Siberian Federal University. Biology 4 (2) (2009) 418–452.

[4] V. A. Kratasyuk, E. N. Esimbekova, Applications of luminous bacteria enzymes in toxicology, Combinatorial chemistry & high throughput screening 18 (10) (2015) 952–959.

[5] E. V. Vetrova, N. S. Kudryasheva, V. A. Kratasyuk, Redox compounds influence on the NAD(P)H:FMN-oxidoreductase-luciferase bioluminescent system., Photochemical & photobiological sciences : Official journal

of the European Photochemistry Association and the European Society for Photobiology 6 (1) (2007) 35–40. doi:10.1039/b608152e.

[6] J. Hastings, K. Nealson, Bacterial bioluminescence, Annual Reviews in Microbiology 31 (1977) 549–595.

[7] S. Parvez, C. Venkataraman, S. Mukherji, A review on advantages of implementing luminescence inhibition test (Vibrio fischeri) for acute toxicity prediction of chemicals., Environment international 32 (2) (2006) 265–268. doi:10.1016/j.envint.2005.08.022.

[8] V. Petushkov, G. Kratasyuk, N. Rodionova, A. Fish, P. Belobrov, Two-enzyme NADH:FMN-oxidoreductase-luciferase system from luminescent bacteria, Biochemistry Acad. Sci. USSR 49 (4) (1984) 593–604.

[9] A. Roda, M. Guardigli, E. Michelini, M. Mirasoli, Bioluminescence in analytical chemistry and in vivo imaging, TrAC Trends in Analytical Chemistry 28 (3) (2009) 307–322. doi:10.1016/j.trac.2008.11.015.

[10] N. Kudryasheva, V. Kratasyuk, P. Belobrov, Bioluminescent analysis. the action of toxicants: Physical-chemical regularities of the toxicants effects, Analytical letters 27 (15) (1994) 37–41.

[11] H. Abu-Soud, L. S. Mullins, T. O. Baldwin, F. M. Raushel, Stopped-flow kinetic analysis of the bacterial luciferase reaction, Biochemistry 31 (15) (1992) 3807–3813.

[12] H. M. Abu-Soud, A. C. Clark, W. A. Francisco, T. O. Baldwin, F. M. Raushel, Kinetic destabilization of the hydroperoxy flavin intermediate by site-directed modification of the reactive thiol in bacterial luciferase, The Journal of Biological Chemistry 268 (11) (1993) 7699–7706.

[13] V. Gubala, L. F. Harris, A. J. Ricco, M. X. Tan, D. E. Williams, Point of care diagnostics: Status and future, Analytical Chemistry 84 (2) (2012) 487–515. doi:10.1021/ac2030199.

[14] E. Kussul, T. Baidyk, D. C. Wunsch, Neural networks and micromechanics, Springer, Heidelberg ; New York, 2010.

[15] S. Agatonovic-Kustrin, R. Beresford, Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research,

Journal of pharmaceutical and biomedical analysis 22 (5) (2000) 717–727.

[16] M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, M. Redn, Artificial neural networks for multicomponent kinetic determinations., Analytical chemistry 67 (24) (1995) 4477–83.

[17] S. Ventura, M. Silva, D. Prez-Bendito, C. Hervs, Artificial neural networks for estimation of kinetic analytical parameters, Analytical Chemistry 67 (24) (1995) 1521–1525.

[18] G. Gholmieh, S. Courellis, S. Fakheri, E. Cheung, V. Marmarelis, M. Baudry, T. Berger, Detection and classification of neurotoxins using a novel short-term plasticity quantification method, Biosensors and Bioelectronics 18 (12) (2003) 1467–1478. doi:10.1016/S0956-5663(03)00120-9.

[19] J. Abdullah, M. Ahmad, L. Y. Heng, N. Karuppiah, H. Sidek, Evaluation of an optical phenolic biosensor signal employing artificial neural networks, Sensors and Actuators B: Chemical 134 (2) (2008) 959–965. doi:10.1016/j.snb.2008.07.009.

[20] M. Gutirrez, S. Alegret, M. del Valle, Bioelectronic tongue for the simultaneous determination of urea, creatinine and alkaline ions in clinical samples., Biosensors & bioelectronics 23 (6) (2008) 795–802. doi:10.1016/j.bios.2007.08.019.

[21] S. S. Haykin, S. S. Haykin, Neural networks and learning machines, 3rd Edition, Prentice Hall, New York, 2009.

[22] P. D. Wasserman, Neural computing: Theory and practice, Van Nostrand Reinhold, New York, 1989.

[23] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten, The WEKA data mining software: an update, ACM SIGKDD Explorations Newsletter 11 (1) (2009) 10. doi:10.1145/1656274.1656278.

[24] S. K. Shevade, S. S. Keerthi, C. Bhattacharyya, K. R. K. Murthy, Improvements to the SMO algorithm for SVM regression, IEEE transactions on neural networks 11 (5) (2000) 11881193.

[25] N. S. Kudryasheva, I. Y. Kudinova, E. N. Esimbekova, V. A. Kratasyuk, D. I. Stom, The influence of quinones and phenols on the triple NAD(H)-dependent enzyme systems, Chemosphere 38 (4) (1999) 751–758.