UDC 519.254:519.257
DOI: 10.21668/health.risk/2025.2.14.eng

Read online

Research article

RANDOM FOREST METHOD FOR INTERPRETING RESULTS OBTAINED BY BIOLUMINESCENCE ANALYSIS OF SALIVA IN PERSONALIZED DIAGNOSTICS

G.V. Zhukova[1], P.A. Martyshuk[1], E.R. Afer[1], A.N. Shuvaev[1], N.A. Rozanova[2], D.V. Sergeev[2], V.A. Kratasyuk[1, 3]

[1]Siberian Federal University, 79 Svobodnyi Av., Krasnoyarsk, 660041, Russian Federation
[2]Research Center for Neurology, 80 Volokolamskoe highway, Moscow, 125367, Russian Federation
[3]Institute of Biophysics, Siberian Department of the RAS, separate division of the Federal Research Center Krasnoyarsk Scientific Center of the Siberian Department of the Russian Academy of Sciences, 50 Akademgorodok, build. 50, Krasnoyarsk, 660036, Russian Federation

Development of personalized medicine and biotechnologies is directly linked to obtaining relevant data, which largely depend on individual characteristics of examined patients. Permissible ranges of analyzed indicators that are commonly used in conventional medicine do not always describe a patient's health adequately. It seems necessary to search for such data analysis techniques, which allow considering variable individual peculiarities of patients' bodies and their lifestyles.

The aim of this study is to determine whether it is possible to use the Random Forest method for biomedical data analysis in order to achieve correct interpretation of results obtained by personalized diagnostic tests. Bioluminescent testing is used as an example since it estimates effects produced by various characteristics of examined patients and their living conditions. The method allows minimizing risks of incorrect diagnosis and adjusting monitoring schemes for specific patients.

This study relies on using the results obtained by diagnosing workloads of railway workers using the bioluminescence method. A patient's health is assessed by examining effects produced by a patient's saliva on intensity of the bi-enzyme system luminescence: NAD(P)H:FMN oxidoreductase + luciferase. This analysis is integral and responses to many factors, each of which can influence the analysis result. Effectiveness of various methods for data analysis is assessed on an example group made of traffic controllers employed by the Krasnoyarsk Branch of Russian Railways JSC. Both statistical methods and the Random Forest machine learning algorithm were used for data analysis.

As a result, our study has revealed that it is advisable to use the Random Forest method for assessing significance of some biochemical saliva indicators to predict health of railway workers. The method makes it possible to identify the most significant factors and create graphs to show partial influence exerted by various factors on the target variable. This study allows optimizing the system for health diagnostics using integral bioluminescence analysis. The Random Forest method can become a component of a personalized bioluminescent biosensor for assessing effects produced by stress and workloads on the body.

Keywords: *personalized diagnostics, machine learning, data analysis, multifactorial analysis, saliva, bioluminescence, biosensor, signal systems.*

Galina V. Zhukova – Junior Researcher of Bioluminescent Technologies Laboratory (e-mail: gvivanova@sfu-kras.ru; tel.: +7 (913) 047-88-37; ORCID: https://orcid.org/0000-0002-8646-1224).

Polina A. Martishuk – student of the Institute of Fundamental Biology and Biotechnology (e-mail: p.martishuk@yandex.ru; tel.: + 7 (904) 896-97-34).

Egor R. Afer – student of the Institute of Fundamental Biology and Biotechnology (e-mail: egorafer@yandex.ru; tel.: +7 (902) 957-64-89).

Andrey N. Shuvaev – Candidate of Physical and Mathematical Sciences, Head of the Basic Department of Biomedical Systems and Complexes of the Institute of Fundamental Biology and Biotechnology (e-mail: ashuvaev@sfu-kras.ru; tel.: +7 (950) 997-77-92; ORCID: https://orcid.org/0000-0002-3887-1413).

Natalia A. Rozanova – graduate student, laboratory research assistant at Brain Institute's Neurobiology and Tissue Engineering Laboratory (e-mail: nataliarozanovaa@gmail.com; tel.: +7 (968) 030-47-28; ORCID: https://orcid.org/0000-0001-9619-4679).

Dmitry V. Sergeev – Candidate of Medical Sciences, Senior Researcher, Neurologist (e-mail: sergeev@neurology.ru; tel.: +7 (926) 860-08-97; ORCID: https://orcid.org/0000-0002-9130-1292).

Valentina A. Kratasyuk – Doctor of Biological Sciences, Professor, Head of Biophysical Department of the Institute of Fundamental Biology and Biotechnology; Leading Researcher (e-mail: valkrat@mail.ru; ORCID: https://orcid.org/0000-0001-6764-5231).

At present, the personalized approach in medicine is becoming more and more popular and this leads to growing volumes of individual data obtained through diagnostics. Such data are usually influenced by multiple factors including age, lifestyle, working conditions, and stress states. Permissible ranges of analyzed health indicators often give no opportunity to describe patients' health or to analyze health risks considering their individual peculiarities. Saliva is a promising biological fluid, which can be used for solving this problem. It has several advantages when used in diagnostics: its collection is fast, easy, inexpensive, and non-invasive; in addition, it can reflect the physiological and pathological state of the body [1–3]. These issues have not been resolved yet for a new bioluminescent non-invasive biotest based on assessing effects produced by saliva on the bi-enzyme system luminescence: NAD(P)H:FMN oxidoreductase + luciferase and the relationship between the saliva composition and pathological and physiological states of the body [4]. Analysis results are influenced by many factors [5] and this makes it necessary to find a method for estimating their contributions into biomedical data considering greater variety of patients; personalized characteristics; to find a way to average these data and to establish criteria that describe both the normal state and deviations from it.

There are three common groups of methods for analyzing biomedical data: statistical methods, clusterization, and machine learning[1] [6–11]. The statistical component can be substantially involved in machine learning algorithms. Statistical analysis usually works with small data volumes and is aimed at getting an insight into interrelations between variables and at testing hypotheses. Machine learning is more actively used in biotechnologies since it allows automating various processes, analyzing larger and more complicated data arrays, and is focused on prediction accuracy and classification when solving tasks without an obvious variable-result. This makes it possible to analyze diverse data types and structures without any strict assumptions [6]. Statistical analysis starts with a hypothesis about a main factor, which describes the body state, and its results are used for testing it. This analysis relies on previously defined models about data distribution (for example, normal distribution)[2].

Correlation analysis is considered the basic statistical method for saliva analysis [7]. In particular, computed Spearman's rank correlation coefficient estimates intensity and direction of a correlation between two ranked variables and provides an insight into how well this correlation is described with a monotonic function. This does not require any assumptions about distribution of attributes due to the test being non-parametric. Next, saliva analysis often involves using the non-parametric Mann – Whitney test to identify authentic statistical differences between two independent groups per the level of an attribute [7, 10] even for small samples[2]. These methods make it possible to find correlations between indicators and conduct primary analysis of a data array. However, experience gained by using basic statistical methods does not always allow establishing key factors able to influence a target indicator of saliva.

Machine learning algorithms, which are often used for detecting diseases of various body systems, can be another solution to the issue of insufficient statistical analysis stock-

---

[1] Makarova N.V. Statisticheskii analiz mediko-biologicheskikh dannykh s ispol'zovaniem paketov statisticheskikh programm Statistica, SPSS, NPSS, SYSTAT [Statistical analysis of biomedical data using Statistica, SPSS, NPSS, SYSTAT statistical software packages]: methodical guide. In: Professor S.S. Aleksanin ed. Saint Petersburg, Polygraphic Center of the Saint Petersburg University of the EMERCOM of Russia Publ., 2012, 179 p. (in Russian); Shorokhova I.S., Kislyak N.V., Mariev O.S. Statisticheskie metody analiza [Statistical analysis methods]: manual, the 2nd ed. Moscow, FLINTA Publ., 2017, 300 p. (in Russian).

[2] Makarova N.V. Statisticheskii analiz mediko-biologicheskikh dannykh s ispol'zovaniem paketov statisticheskikh programm Statistica, SPSS, NPSS, SYSTAT [Statistical analysis of biomedical data using Statistica, SPSS, NPSS, SYSTAT statistical software packages]: methodical guide. In: Professor S.S. Aleksanin ed. Saint Petersburg, Polygraphic Center of the Saint Petersburg University of the EMERCOM of Russia Publ., 2012, 179 p. (in Russian)

piles for interpreting biomedical data about saliva [12–18]. We can choose an algorithm to establish regularities per previously marked data arrays judging from specificity of a task (searching for significant factors affecting saliva indicators). Data marks provide a feedback for a model thereby adjusting the results and making prediction more accurate. To analyze data on saliva, we can use a family of algorithms based on creating decision trees such as CART, Random Forest and boosting [13, 15, 16, 19]. Random Forest is among the most popular methods for solving non-deep machine learning tasks[3]. In addition to identifying a class of an object and finding the precise value of a target variable, the Random Forest algorithm predicts significance of each model parameter for decision-making considering its level against the tree top, how frequent this parameter can be found in a decision node, and the number of objects correctly classified by using it. Parameters found closer to the tree top are the most significant [20]. The Random Forest algorithm creates a partial dependence plot (PDP) for each model parameter. For regression, the plot shows averaged relationships between an attribute and a target value. For classification, an averaged relationship is built between likelihood that a class of an object is identified correctly and an attribute used in creating a model[4].

The aim of this study was to determine whether it was possible to use the Random Forest method for biomedical data analysis in order to achieve correct interpretation of results obtained by personalized diagnostic tests in a bioluminescent biosensor for estimating impacts of stress and workloads on the human body.

Materials and methods. To estimate effectiveness of data analysis, we used the results obtained by diagnostic tests performed in a group of traffic controllers employed by the Krasnoyarsk Branch of Russian Railways JSC (n = 43). The tests were performed every day for a month according to work schedules in 2022 and 2023. Resampling was used to extend the sample; re-learning was controlled by dividing the sample into the training and testing samples. The examined indicators were established by analyzing saliva collected two times a day: prior to a work shift and after it. The following saliva indicators were estimated: pH, lactate concentration, residual luminescence (luciferase index) of the bienzyme system of luminescent bacteria: NAD(P)H:FMN oxidoreductase + luciferase (LI, %) [21], as well as biochemical blood indicators and other results obtained by periodical medical check-ups of personnel accomplished at the Clinical Hospital 'Russian Railways Medicine'. We also analyzed the results obtained by a self-survey, which included subjective assessment of stress level and work capacity prior to a work shift and after it, data about daily routines, food consumption, use of drugs and energy drinks, and tobacco smoking.

All studies were conducted in conformity with the principles of biomedical ethics and approved by the local ethics committee of the Siberian Federal University (Krasnoyarsk; the meeting protocol No. 5 dated November 11, 2019).

Each participant gave a voluntary informed written consent to take part in this research after being provided with explanation of its potential risks and advantages as well as its essence.

Machine learning for classification and regression by the Random Forest algorithm was conducted in Python 3 using the Scikit-learn library. Factor significance is established in this library per the Gini Index improvement after division per each of these factors within classification or per declining dispersion of prediction per this factor within regression. The study relied on using a possibility provided by the Scikit-learn library to visualize partial influence exerted by each factor on the target variable. Statistical analysis was accomplished with the R

[3] Géron A. Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, 2nd ed. O'Reilly Media Publ., 2019, 848 p.

[4] Jain A., Fandango A., Kapoor A. Tensorflow Machine Learning Projects. Packt Publishing, 2018, 322 p.

Table 1

Factors used for testing effectiveness of the program for data analysis

Field name	Measuring units / scores	Range
Lactate concentration in saliva	mmol/l	0.2–80.7
Blood glucose	mmol/l	3.2–9.8
Total cholesterol	mol/l	3.4–6.8
pH	pH	3–8
Sampling time	Prior to work shift (morning) – 1 After work shift (evening) – 0	0.1
Stress level (attitude to work)	Very anxious – 3 Anxious – 2 Calm – 1 No anxiety – 0	0–3
A time gap between having a smoke and sampling	Minutes	0–120
Sex	Female – 1 Male – 0	0.1
Smoking	Yes – 1 No – 0	0.1
Time of a day (morning / evening)	Morning – 1; evening – 0	0.1
LI (luciferase index)	%	1.08–142.50

Table 2

Correlations between some factors and the target variable (LI)

Factors	Correlation coefficient
Attitude to work (females)	-0.618
Time gap between having a smoke and sampling (females)	-0.699
Blood glucose, mmol/l (females)	0.446
Total cholesterol, mol/l (males)	0.671

software using basic libraries and the ggplot2 library. Differences were deemed significant at p-values below or equal to 0.05.

Results and discussion. After preliminary processing, the database containing biomedical data consisted of 310 lines. To extend it, resampling was made allowing a four-fold increase in the number of lines by randomly creating new lines out of the already existing ones. The sample was then divided into the training sample (80 % of the initial sample) and the testing sample (20 % of the initial sample). Variables provided in Table 1 were used for further analysis.

Correlation coefficients between target parameters from Table 1 were estimated using statistical data analysis as preprocessing for identifying the most promising factors to be included in machine learning.

Correlation analysis revealed significant negative correlations between the target variable LI and the participants' attitude to work per the scale between 1 and 5 and a time gap after having a smoke for female participants. Positive significant correlations were established between blood glucose (females) and total cholesterol (males). The correlation coefficients provided in Table 2 are significant at $p < 0.05$.

Random Forest, an ensemble machine learning algorithm based on decision tree building, was selected as a promising method for analysis. Two types of tasks were solved in the study: classification (using Random Forest Classifier of the Scikit-learn library), where the algorithm predicted belonging to groups of high, medium, or low LI values, and regression (using Random Forest Regressor of the Scikit-learn library) involving prediction of the LI exact value per the set factors. The most significant factors were identified and partial dependence plots were created. In both cases, we estimated prediction accuracy and significance of each factor per its influence on LI (Figure 1) and analyzed partial influence exerted on LI by a given factor depending on its level. Factor significance and standard deviation were calculated based on the results

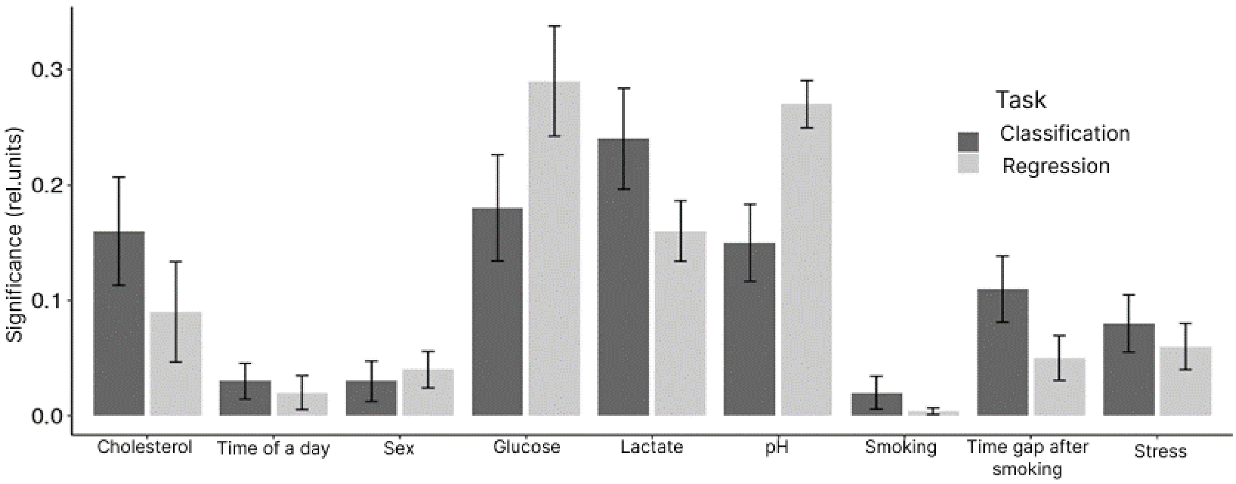

Figure 1. Distribution of factors per significance of their influence on LI in regression and classification tasks

obtained by building 2000 trees in each task and were established per the Gini Index based on a decline in uncertainty when breaking nodes: the higher a factor was in a tree, the more significant it was considered.

The hypothesis that each factor differed significantly from zero was tested using the Student's t-test with subsequent Bonferroni correction per multiple comparisons; as a result we revealed significant difference for each attribute. Three factors turned out to be the most significant per their influence on LI both in classification and regression, namely, lactate concentration, blood glucose, and pH level. In addition, total blood cholesterol can be considered significant in classification.

It is noteworthy that smoking had different significance; if the very fact of a participant being a smoker had practically zero significance, then a time gap between having a smoke and sampling was authentically higher for both tasks (*p*-values were 1.18E-07 for classification and 1.26E-06 for regression). This discrepancy can be explained by a rapid leveling of the effect produced by smoking on saliva.

Mean values of the learning metrics equaled the following: MAE = 28.73; MS = 880.99: R^2 = 0.175 in regression. Prediction accuracy equaled 95 % in classification. This high classification accuracy at a low R^2 value is due to a different nature of regression and classification tasks. In classification, the model recognizes classes per non-linear combinations of attributes whereas the regression model suffers from high response variance.

Partial influence was also estimated for two types of tasks (Figures 2 and 3).

Blood glucose level higher than 6 mmol/l is the threshold, above which this factor ceases to exert any influence on the luciferase index (Figure 2). The highest likelihood of a LI value closer to the upper bound of the assumed normal range is observed when blood glucose is low. For pH, a level below 5 has practically no influence on the target value level. However, as pH grows, its effect on LI is described with ambiguous dynamics and high dispersion. Lactate concentration in saliva has effects on the luciferase index value at low levels (lower than 8 mmol/l).

For classification, blood glucose level showed a relationship opposite to that observed for regression as influence exerted by this factor grew (Figure 3).

The dynamics itself was shaped saturated and significance leveled off at the blood glucose level of approximately 5 mmol/l. The lactate and pH plots had shapes similar to that in regression. When lactate concentration reached 8 mmol/l, a clear transition appeared in the significance level, after which any influence on the luciferase index disappeared. The relationship with pH had a rather chaotic dynamic with the overall trend to grow as the factor level grew.

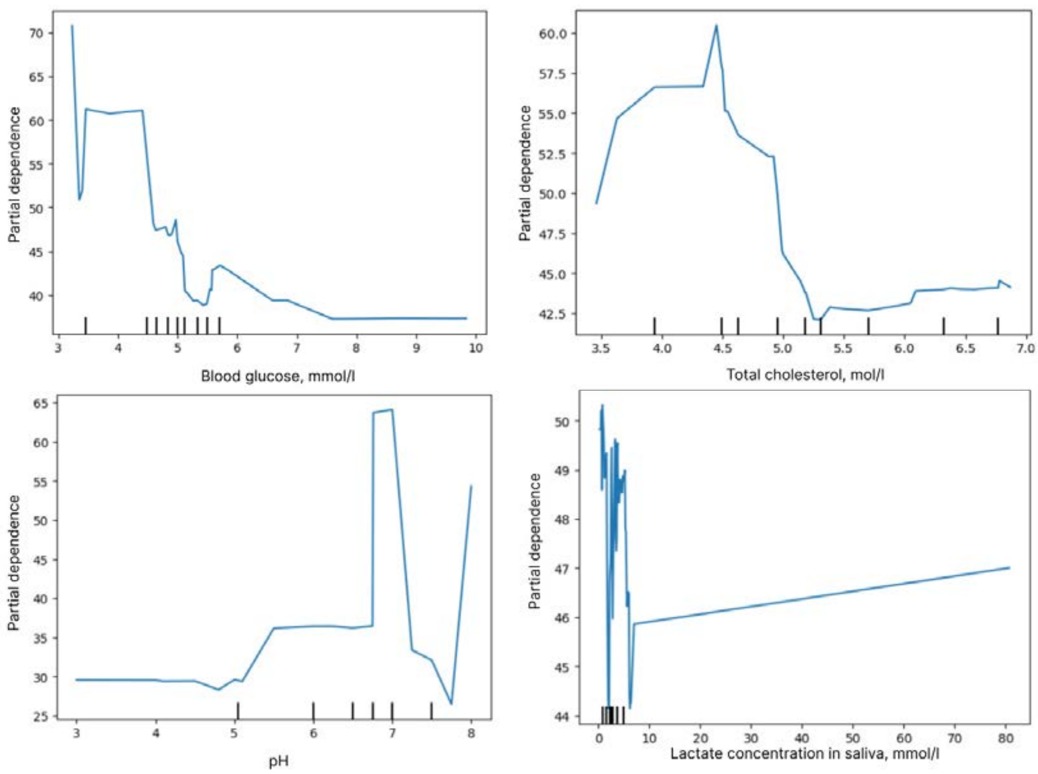

Figure 2. Partial dependence plots to show partial influence of selected factors on the target variable (LI) for regression task

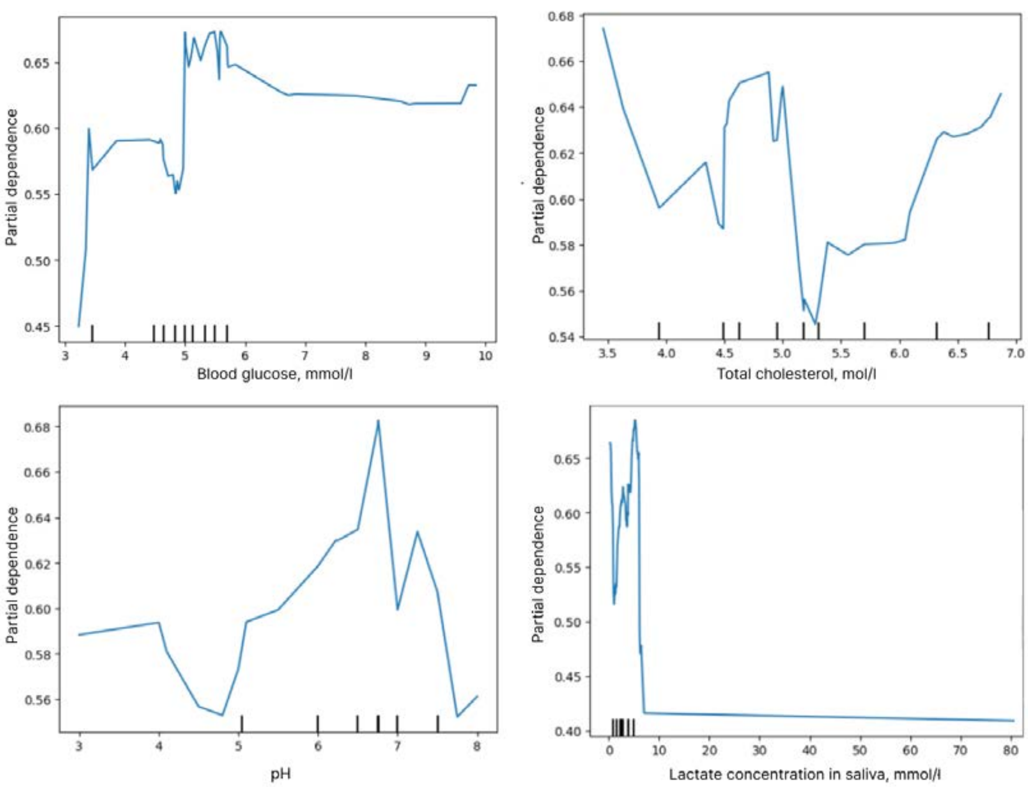

Figure 3. Partial dependence plots to show partial influence of selected factors on the target variable (LI) for classification task

Conclusion. The selected factors were analyzed using the Random Forest algorithm; the analysis showed that not all parameters were equally significant for predicting the target indicator. Blood glucose demonstrated opposite effects in the regression and classification tasks: in regression, its influence decreased when its level grew above 6 mmol/l whereas in classification the factor significance leveled off when the blood glucose level was approximately 5 mmol/l. This may indicate some complex non-linear interrelations between blood glucose and the state of the analyzed material. The pH level also had different influence depending on the task. In regression, pH effects on the luciferase index of the bioluminescence system were considerably variable, especially as it increased. In classification, dynamics of pH influence also remained uncertain and this complicated clear understanding of its role in changes of the body state despite its great significance for predicting LI levels. Lactate concentration had a precise transition value (8 mmol/l), after which its influence on the luciferase index decreased considerably. This indicates possible existence of a threshold value, after which lactate ceased to be a significant factor; this may be quite useful for practical use within workers' health monitoring. Despite the low R^2 value (0.175), the model identifies significant factors quite successfully and demonstrates high prediction accuracy. This indicates strong non-linearity and potential existence of hidden variables, which have not been covered in the current data set. Our study did not include several potentially significant factors in the analysis such as physical activity, stress situation outside workplace, and individual metabolism peculiarities although inclusion of these parameters might have considerably increased the model accuracy and usefulness. In future, we plan to compare effectiveness of the Random Forest method, conventional linear regression, logistic regression, XGBoost, and neural networks.

The Random Forest method is suitable for selecting informative factors, which describe peculiarities of personal characteristics of the body in its normal state and in case of deviations from it. It allows minimizing the number of factors influencing LI in case a test is preliminary adjusted for personalized work. The Random Forest method can become a component of a personalized bioluminescent biosensor for assessing effects produced by stress and workloads on the body. The suggested data analysis can be used for analyzing and interpreting results obtained by various research methods; it makes it possible to analyze big data volumes and reveal hidden regularities as well as to minimize risks of incorrect diagnosis and to adjust a monitoring scheme for patients, the latter being a crucial aspect in biotechnological and medical research.

Financial support. The work was supported by a grant from the Ministry of Science and Higher Education of the Russian Federation for implementation of major scientific projects in priority areas of scientific and technological development (Project No. 075-15-2024-638).

Competing interests. The authors declare the absence of obvious and potential competing interests related to the publication of this article.

References

1. Tabak L.A. Point-of-care diagnostics enter the mouth. *Ann. N Y Acad. Sci.*, 2007, vol. 1098, pp. 7–14. DOI: 10.1196/annals.1384.043

2. Kubala E., Strzelecka P., Grzegocka M., Lietz-Kijak D., Gronwald H., Skomro P., Kijak E. A Review of Selected Studies That Determine the Physical and Chemical Properties of Saliva in the Field of Dental Treatment. *BioMed Res. Int.*, 2018, vol. 2018, pp. 6572381. DOI: 10.1155/2018/657238

3. Kaczor-Urbanowicz K.E., Carreras-Presas C.M., Aro K., Tu M., Garcia-Godoy F., Wong D.T. Saliva diagnostics – Current views and directions. *Exp. Biol. Med. (Maywood)*, 2017, vol. 242, no. 5, pp. 459–472. DOI: 10.1177/1535370216681550

4. Esimbekova E.N., Torgashina I.G., Kalyabina V.P., Kratasyuk V.A. Enzymatic Biotesting: Scientific Basis and Application. *Contemp. Probl. Ecol.*, 2021, vol. 14, pp. 290–304. DOI: 10.1134/S1995425521030069

5. Kratasyuk V.A., Stepanova L.V., Ranjan R., Sutormin O.S., Pande S., Zhukova G.V., Miller O.M., Maznyak N.V., Kolenchukova O.A. A noninvasive and qualitative bioluminescent assay for express diagnostics of athletes' responses to physical exertion. *Luminescence*, 2021, vol. 36, no. 2, pp. 384–390. DOI: 10.1002/bio.3954

6. Jordan M.I., Mitchell T.M. Machine learning: Trends, perspectives, and prospects. *Science*, 2015, vol. 349, no. 6245, pp. 255–260. DOI: 10.1126/science.aaa8415

7. Bel'skaya L., Sarf E. The use of Fourier transform IR spectroscopy of saliva for rapid assessment of the level of lipid peroxidation products. *Biomedical Chemistry: Research and Methods*, 2019, vol. 2, no. 2, pp. e00094. DOI: 10.18097/BMCRM00094 (in Russian).

8. Bel'skaya L.V., Sarf E.A., Kosenok V.K. Correlation interrelations between the composition of saliva and blood plasmain norm. *Klinicheskaya laboratornaya diagnostika*, 2018, vol. 63, no. 8, pp. 477–482. DOI: 10.18821/0869-2084-2018-63-8-477-482 (in Russian).

9. Nikonorova M.L. Building optimal models for analysis of biomedical data. *Vestnik novykh meditsinskikh tekhnologii*, 2021, vol. 28, no. 1, pp. 55–59. DOI: 10.24412/1609-2163-2021-1-55-59 (in Russian).

10. Berestneva O.G., Osadchaya I.A., Nemerov E.V. Methods for studying the medical data structure. *Vestnik nauki Sibiri*, 2012, vol. 1, no. 2, pp. 333–338 (in Russian).

11. Klimenko A.V., Slashchev I.S. Klasternyi analiz dannykh [Cluster data analysis]. *Vestnik nauki*, 2019, no. 1 (10), vol. 1, pp. 159–163 (in Russian).

12. Braz D.C., Popolin Neto M., Shimizu F.M., Sá A.C., Lima R.S., Gobbi A.L., Melendez M.E., Arantes L.M.R.B. [et al.]. Using machine learning and an electronic tongue for discriminating saliva samples from oral cavity cancer patients and healthy individual. *Talanta*, 2022, vol. 243, pp. 123327. DOI: 10.1016/j.talanta.2022.123327

13. Lee E., Park S., Um S., Kim S., Lee J., Jang J., Jeong H.-O., Shin J. [et al.]. Microbiome of Saliva and Plaque in Children According to Age and Dental Caries Experience. *Diagnostics (Basel)*, 2021, vol. 11, no. 8, pp. 1324. DOI: 10.3390/diagnostics11081324

14. Kouznetsova V.L., Li J., Romm E., Tsigelny I.F. Finding distinctions between oral cancer and periodontitis using saliva metabolites and machine learning. *Oral Dis.*, 2021, vol. 27, no. 3, pp. 484–493. DOI: 10.1111/odi.13591

15. Zarrin P.S., Roeckendorf N., Wenger C. In-Vitro Classification of Saliva Samples of COPD Patients and Healthy Controls Using Machine Learning Tools. *IEEE Access*, 2020, vol. 8, pp. 168053–168060. DOI: 10.1109/ACCESS.2020.3023971

16. Murata T., Yanagisawa T., Kurihara T., Kaneko M., Ota S., Enomoto A., Tomita M., Sugimoto M. [et al.]. Salivary metabolomics with alternative decision tree-based machine learning methods for breast cancer discrimination. *Breast Cancer Res. Treat.*, 2019, vol. 177, no. 3, pp. 591–601. DOI: 10.1007/s10549-019-05330-9

17. Rabbani N., Kim G.Y.E., Suarez C.J., Chen J.H. Applications of machine learning in routine laboratory medicine: Current state and future directions. *Clin. Biochem.*, 2022, vol. 103, pp. 1–7. DOI: 10.1016/j.clinbiochem.2022.02.011

18. Li X., Zheng J., Ma X., Zhang B., Zhang J., Wang W., Sun C., Wang Y. [et al.]. The oral microbiome of pregnant women facilitates gestational diabetes discrimination. *J. Genet. Genomics*, 2021, vol. 48, no. 1, pp. 32–39. DOI: 10.1016/j.jgg.2020.11.006

19. Kuwabara H., Katsumata K., Iwabuchi A., Udo R., Tago T., Kasahara K., Mazaki J., Enomoto M. [et al.]. Salivary metabolomics with machine learning for colorectal cancer detection. *Cancer Sci.*, 2022, vol. 113, no. 9, pp. 3234–3243. DOI: 10.1111/cas.15472

20. Alam M.Z., Rahman M.S., Rahman M.S. A Random Forest based predictor for medical data classification using feature ranking. *Informatics in Medicine Unlocked*, 2019, vol. 15, pp. 100180. DOI: 10.1016/j.imu.2019.100180

21. Stepanova L.V., Kolenchukova O.A., Zhukova G.V., Sutormin O.S., Kratasyuk V.A. The Use of the Bioluminescent Enzyme Bioassay for the Analysis of Saliva of Railway Transport Workers to Monitor the Functional State of the Body in the Conditions of Labor Activity. *Biofizika*, 2024, vol. 69, no. 3, pp. 674–683. DOI: 10.31857/S0006302924030224 (in Russian).

Zhukova G.V., Martyshuk P.A., Afer E.R., Shuvaev A.N., Rozanova N.A., Sergeev D.V., Kratasyuk V.A. Random Forest method for interpreting results obtained by bioluminescence analysis of saliva in personalized diagnostics. Health Risk Analysis, 2025, no. 2, pp. 166–174. DOI: 10.21668/health.risk/2025.2.14.eng

Received: 22.12.2024
Approved: 07.05.2025
Accepted for publication: 14.06.2025