Technical Physics Letters, 2025, Vol. 51, No. 9

Approaches to detecting the presence of blood pathology: synergy of hardware implementation of spectrophotometry methods and data mining

© M.S. Mazing, A.Yu. Zaitceva

Institute of Analytical Instrument Making, Russian Academy of Sciences, St. Petersburg, Russia
E-mail: mazmari@mail.ru

Received May 5, 2025
Revised June 20, 2025
Accepted June 24, 2025

The approaches to detecting blood system pathologies based on non-invasive methods of spectrophotometry in the visible and near infrared ranges and intelligent data analysis are proposed. Blood system pathologies are diseases that affect blood components, such as red and white blood cells, platelets and plasma. A group of patients with various pathologies, such as erythrocytosis, anemia and leukemia, was studied. The effectiveness of a combined approach combining multichannel optical spectroscopic analysis with modern machine learning algorithms for processing broadband spectral characteristics of biological tissues in order to differentiate pathological and physiological conditions is shown..

Keywords: spectrophotometry, blood system pathology, data mining, linear discriminant analysis.

DOI: 10.61011/TPL.2025.09.61864.8098

The cardiovascular system is one of the most important systems of the body. Together with blood, it connects tissue cells and organs into a single whole, supplying nutrients and oxygen to almost all cells and transporting carbon dioxide and other metabolic products away from them [1]. Microscopy of blood films has remained the primary method for analyzing the composition of formed elements of blood for one hundred and fifty years. However, automatic analyzers have become widespread in the last decade [2,3]. The biochemical composition of blood plasma is assessed using a variety of methods, which are chosen depending on the specific substance that needs to be analyzed. Atomic emission spectroscopy (also known as flame photometry) has earlier been used widely to assess the ionic composition of blood plasma. This method is relatively simple and inexpensive, but has recently been superseded by such techniques as turbidimetry and colorimetry, which were implemented successfully in automatic analyzers [4].

The key disadvantage of classical methods for analysis of certain blood components is their invasiveness. In view of this, constant attempts are being made to develop minimally invasive or non-invasive methods for assessing the blood system (dry chemistry, assessment of biomarkers in tissue fluid, sweat, saliva, etc.) [5].

One possible way toward completely non-invasive assessment of the blood system is the detection of its pathologies. Spectroscopy is used for this purpose. Specifically, while oxyhemoglobin absorbs light with a wavelength of 660 and 940 nm, carboxyhemoglobin absorbs at a wavelength of 539 and 570 nm. Therefore, one may pass radiation through tissues or analyze radiation returned from tissues to determine the number of pathological agents, detecting deviations from the conditionally normal state of biological tissues and estimating the degree of this deviation [6].

The hardware and software complex (HSC) proposed in the present study is specific, first, in using a set of weakly selective sensors sensitive to components of the biological environment of the body and, second, in that the measurement results are analyzed using statistical methods for processing of large data arrays [7]. The basis of the L234-DIOD complex developed at the Institute for Analytical Instrumentation of the Russian Academy of Sciences (St. Petersburg) is a set of modern analytical sensors manufactured by AMS (model as7265x), each of which measures a specific parameter (characteristic) of the sample under study (tissue, liquid, etc.). The AS72651/652/653 chipset is calibrated at the factory. Each analytical sensor includes six optical channels with silicon interference filters with a Gaussian transmission profile. The set of 18 spectral channels covers a range of 410–940 nm with an optimized sampling step (25–50 nm). The module is fitted with an optical emission system (a broadband white and two near-infrared SMD LEDs), which is sufficient for various spectroscopic analysis techniques. The hardware complex implements the principle of reflective spectrophotometry. The system supports two operating modes: pulsed with software-configurable parameters and continuous monitoring. Power is supplied by a lithium-ion battery with a capacity of 500 mA · h ensuring more than 5 h of independent operation. Two scenarios of operation are implemented: online mode with real-time data transmission and offline mode with recording of readings into built-in memory (logging) and subsequent uploading to a personal computer.

Systems of intelligent processing of multidimensional information, which allow, with a high degree of automation, to extract data on the qualitative parameters of multicomponent biological environments being analyzed from arrays of analytical signals, were used in the HSC. The operating principle of the L234-DIOD HSC is illustrated in Fig. 1.

The outer casing of the electronic sensor unit was designed for convenience and safety of operation of the optical system. Having reviewed the existing techniques for manufacture of electronic device casings, we decided to

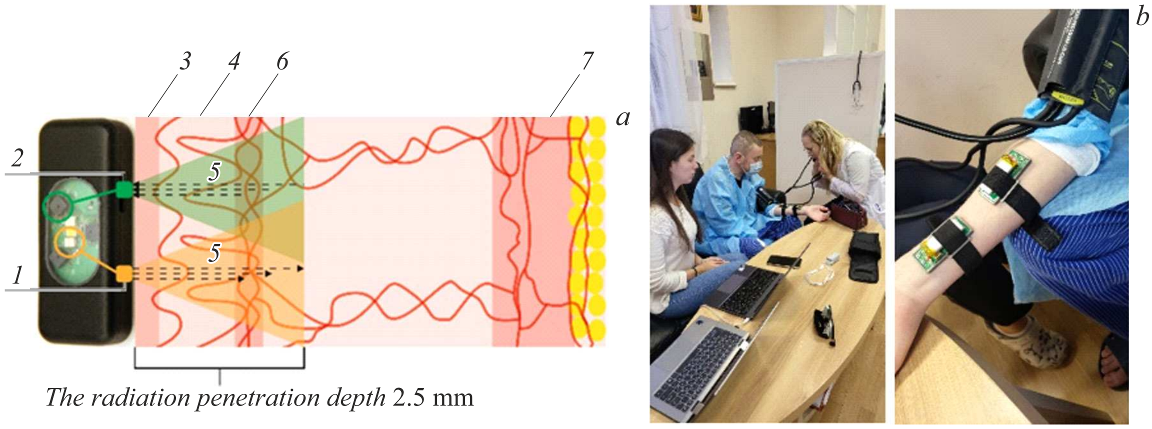

Figure 1. *a* — Operating principle of the L234-DIOD HSC. *1* – Radiation source, *2* — radiation detector, *3* — epidermis, *4* — dermis, *5* — radiation, *6* — superficial vascular plexus, and *7* — deep vascular plexus. *b* — Experimental identification of blood system pathology (Institute for Analytical Instrumentation, St. Petersburg).

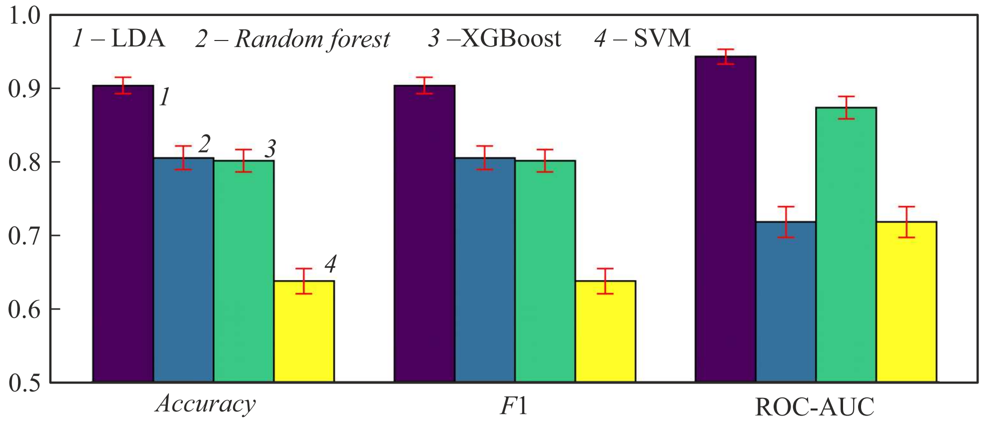

Figure 2. Graphical representation of the comparison of metrics of different models obtained by cross-validation. The confidence interval is indicated.

fabricate the outer casing of the electronic sensor unit using modern additive processes.

Two groups of test subjects participated in the experimental study: 31 patients of the hematology department with various blood pathologies and 17 nominally healthy subjects without identified pathologies in the control group. The sample included patients with different hematological and systemic diseases (anemia, thrombocytopenia, leukocytosis, erythrocytosis, polycythemia, and oncological pathologies). Non-invasive spectrophotometric monitoring with the L234-DIOD HSC was carried out under standard clinical conditions. The portable optical complex was secured on the skin surface in the right cheek area, which is characterized by minimum epidermis thickness and maximum transparency, for assessing the state of the microvascular and tissue system. A series of five–seven consecutive measurements were performed for each participant at 5 ± 2 s intervals over a standardized time period (60 s). The obtained data were transferred to a computer for subsequent processing and analysis.

The data processing module is a mathematical model for optimum classification of conditions into pathological and normal ones. The model with the highest accuracy in cross-validation was considered to be the optimum one. This approach was chosen due to the sample size, which is too small to divide it into training and validation sets.

The experimental data were a set of spectral measurements at 18 wavelengths for each study participant. Since the present study did not involve an analysis of temporal dynamics of the functional state, the median value of each series of measurements was used for each spectral channel.

Thus, the dataset was a matrix of spectrophotometric data of the object–feature type with a size of 48×15. Its rows corresponded to the test subjects, and the median values at the corresponding wavelength were listed in columns. This matrix was also accompanied by a vector of binary labels of pathological and normal classes.

The chosen classification methods were linear discriminant analysis (LDA), random forest, gradient boosting (in the XGBoost implementation), and support vector

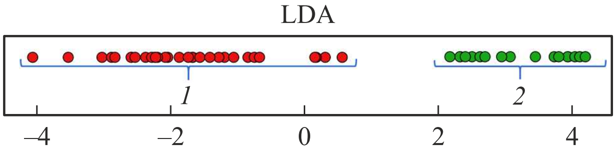

Figure 3. Graphical representation of the change in response in the presence of blood system pathology for the data set obtained using the L234-DIOD HSC in a new space defined by one linear discriminant component (LDA). *1* — Patients with blood system pathology; *2* — nominally healthy patients.

machine (SVM) with a radial basis function kernel. The following metrics averaged over the cross-validation folds were used to compare the quality of models: accuracy ($Accuracy$), F-measure ($F1$), and area under the ROC curve (ROC-AUC). LDA was found to provide the best results in cross-validation: $Accuracy = 0.901 \pm 0.011$, $F1 = 0.91 \pm 0.012$, and ROC-AUC $= 0.94 \pm 0.01$ (Fig. 2).

Figure 3 illustrates the change in response in the presence of blood system pathology for the data set obtained using the L234-DIOD HSC in a new space defined by one linear discriminant component (LDA).

A clear separation between two groups of spectrophotometric data is evident, verifying the effectiveness of the chosen classification method. The results were presented both graphically in Fig. 3 and as a one-component LDA equation. The equation for the discriminant component takes the form

$$\begin{aligned} \mathrm{LDA} = {} & 0.003I_{410} + 0.005I_{435} - 0.046I_{460} \\ & + 0.035I_{485} - 0.001I_{510} + 0.002I_{560} \\ & - 0.008I_{585} - 0.054I_{610} + 0.002I_{645} \\ & + 0.013I_{680} - 0.005I_{705} + 0.028I_{810} \\ & - 0.001I_{860} + 0.009I_{900} - 0.003I_{940}. \end{aligned} \tag{1}$$

Having analyzed the LDA weight coefficients, we found that the greatest contribution to classification is produced by spectral characteristics in the visible region: 610 nm (corresponds to the absorption peak of oxyhemoglobin), 460 nm (characteristic of methemoglobin), and 485 nm (characteristic of reduced hemoglobin (deoxyhemoglobin)). This is potentially indicative of a violation of the oxygen transport function of blood in pathology. The following infrared wavelengths also contribute to classification: 810 nm (close to the isosbestic point of oxyhemoglobin and deoxyhemoglobin) and 900 nm (a significant difference between the absorption coefficients of oxyhemoglobin and deoxyhemoglobin is observed here). This suggests the presence of characteristic patterns of optical response of biological tissues in each of the studied groups. The obtained data verify the applicability of the proposed approach in non-invasive diagnostics, and the identified patterns open up opportunities for developing new algorithms for hematological disease screening.

The results of this study indicate that the designed sensor systems for assessment of the state of biological media and tissues are highly efficient. The non-invasive optical method for detection of blood pathology via visible and near-infrared spectrophotometry and intelligent data analysis is an express, fast, and inexpensive technique that may be used widely both in sports medicine (to identify manipulations with blood doping) and in clinical research and practical healthcare [8,9].

Funding

This study was supported financially by a grant from the Russian Science Foundation (project No. 24-21-00404) .

Compliance with ethical standards

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Informed voluntary consent was obtained from each study participant.

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] L.M. Forbes, Am. J. Respir. Cell Mol. Biol., **72** (4), 456 (2025). DOI: 10.1165/rcmb.2024-0178LE

[2] R.H.G. Schwinger, Cardiovasc. Diagn. Therapy, **11** (1), 263 (2021). DOI: 10.21037/cdt-20-302

[3] M.N. Zenina, E.R. Shilova, N.Yu. Chernysh, Vestn. Gematol., **17** (4), 24 (2021) (in Russian).

[4] A.A. Astakhov, V.V. Kazartsev, K.V. Kuchkin, J. Barg, Mod. Technol. Med., **14** (3), 42 (2022). DOI: 10.17691/stm2022.14.3.05

[5] N.G. Kostsova, I.D. Dzhopua, O.A. Dogotar, A.V. Adilkhanov, I.S. Nikitin, Sovrem. Probl. Zdravookhr. Med. Stat., No. 4, 872 (2023) (in Russian). DOI: 10.24412/2312-2935-2023-4-872-886

[6] V.V. Gnoevykh, Yu.A. Shorokhova, A.Yu. Smirnova, E.V. Efremova, Russ. Arch. Intern. Med., **13** (1), 75 (2023). DOI: 10.20514/2226-6704-2023-13-1-75-80

[7] G. Gabrieli, M. Manica, P. Ruch, Electrochem. Soc. Meeting Abstracts, **244**, 2919 (2023). DOI: 10.1149/MA2023-02622919mtgabs

[8] W. Schmidt, N. Prommer, Eur. J. Appl. Physiol., **95**, 486 (2005). DOI: 10.1007/s00421-005-0050-3

[9] X. Li, Y. Li, H. Wei, C. Wang, B. Liu, Sensors, **24** (11), 3602 (2024). DOI: 10.3390/s24113602

Translated by D.Safin