

MATHEMATICAL MODELING OF NIR SPECTROSCOPY DATA AND DEVELOPMENT OF A DIAGNOSTIC MODEL BASED ON ARTIFICIAL INTELLIGENCE

Abdugany Dzhunusovich Satybaev, Toshpulatov Sherali Mukhamadaliyevich

*Department of Information Technologies and Management, Osh Technological University
named after M.M. Adyshev, Kyrgyz Republic*

E-mail: abdu-satybaev@mail.ru, toshpulatovsherali12@gmail.com

Abstract. The paper presents the development of a diagnostic model based on near- infrared (NIR) spectroscopy data using artificial intelligence methods. Spectral signals are preprocessed and mathematically modeled. The modeling is based on the Beer–Lambert law, and the J criterion is used to evaluate signal quality. The obtained results demonstrate the effectiveness of the proposed model. The maximum J value reached 6.57, while for the simple method it was

4.75. The overall accuracy of the model ranged from 85% to 92%. In some cases, a decrease in accuracy was observed due to noise and external factors. The proposed approach is effective for detecting optical changes in biological tissues and for early disease diagnosis.

Keywords: artificial intelligence, NIR spectroscopy, near-infrared radiation, mathematical modeling, diagnostic model, machine learning, spectral analysis, biological tissues, optical properties, early diagnosis.

Аннотация. В статье рассматривается разработка диагностической модели на основе данных ближней инфракрасной (NIR) спектроскопии с использованием методов искусственного интеллекта. Спектральные сигналы предварительно обрабатываются и подвергаются математическому моделированию. Моделирование основано на законе Бера – Ламберта, а для оценки качества сигнала используется критерий J. Полученные результаты показали эффективность предложенной модели. Максимальное значение J составило 6.57, тогда как для простого метода этот показатель равен 4.75. Общая точность модели составила 85–92%. В отдельных случаях наблюдалось снижение точности под воздействием шума и внешних факторов.

Предложенный подход является эффективным для выявления оптических изменений в биологических тканях и ранней диагностики заболеваний.

Ключевые слова: искусственный интеллект, NIR спектроскопия, ближнее инфракрасное излучение, математическое моделирование, диагностическая модель,

машинное обучение, спектральный анализ, биологические ткани, оптические свойства, ранняя диагностика.

Introduction

In modern medicine, the problem of early disease detection remains one of the most pressing challenges. The identification of changes occurring in biological tissues using non-invasive methods is considered an important scientific direction [1]. In this context, optical diagnostic methods play a significant role. Near-infrared (NIR) spectroscopy enables the study of optical properties of biological tissues. This method makes it possible to obtain information about hemoglobin concentration and oxygenation levels. It also allows the assessment of biochemical changes in tissues. In recent years, diagnostic systems based on NIR spectroscopy and artificial intelligence have been actively developing [2]. Many studies have focused on processing spectral data and performing classification tasks. The use of machine learning algorithms in this process has shown effective results. However, in most existing studies, deep mathematical modeling of spectral signals has not been sufficiently addressed. The problem of linking these signals with biological indicators also remains unresolved. This work is based on spectral analysis and mathematical modeling.

Problem Statement. In this study, the problem of developing a diagnostic model based on near-infrared (NIR) spectroscopy data is formulated [3]. The main focus is on spectral signals obtained from biological tissues. During the research process, these signals are first preprocessed. Then they are mathematically modeled. After that, they are analyzed using artificial intelligence algorithms. To solve this problem, modern scientific approaches are applied. Signal processing methods and machine learning algorithms are used. The main objective is to determine the relationship between spectral parameters and biological indicators.

Methodology

In this study, a comprehensive approach for processing NIR spectroscopy data is proposed [4]. First, the spectral signals are denoised. Then they are normalized. At the next stage, the signals are mathematically modeled. In this process, relationships between spectral parameters and biological indicators are identified. After that, the extracted parameters are classified using machine learning algorithms. Modern artificial intelligence methods are applied in building the model. Its performance is evaluated based on experimental results. The proposed method is designed to be simple and clear. Therefore, it can be applied by other researchers [5]. It represents a flexible and practice-oriented approach.

Results and Examples

During the study, a diagnostic model based on NIR spectroscopy data was developed and tested. MATLAB was used to build the model. The Image Processing Toolbox and Statistics and Machine Learning Toolbox were applied for computations. At the initial stage, spectral images were formed

using multiple frames. This approach helped reduce noise. However, in low-intensity signals, some noise components still remained. In the next step, the images were denoised and normalized [6]. The mathematical representation of the NIR signal was based on the Beer-Lambert law:

$$I = I_0 \cdot e^{-\mu L} \quad (1)$$

Here, I - represents the intensity of the returned signal, I_0 - is the initial signal, μ - is the absorption coefficient, and L - is the optical path length. This model makes it possible to evaluate optical changes in biological tissues. To separate signal and background regions, ROI and BG masks were applied. To assess signal quality, the following criterion was used:

$$J = \frac{\mu_{ROI} - \mu_{BG}}{\sigma_{BG}} \quad (2)$$

Here, μ_{ROI} - is the mean value of the signal region, μ_{BG} - is the mean value of the background signal, and σ_{BG} - is the standard deviation of the background signal. This criterion evaluates the difference between the signal and the background relative to the noise level. For example, during the experiment, the following values were obtained: $\mu_{ROI} = 0.72$, $\mu_{BG} = 0.25$, $\sigma_{BG} = 0.08$. As a result of the calculation:

$$J = \frac{0.72 - 0.25}{0.08} = 5.875$$

This value indicates that the signal is significantly higher relative to the background noise. However, in some experiments, an increase in background variance led to a decrease in the J value. Machine learning methods were applied in the model. Artificial intelligence algorithms were used for classification. As a result, the presence or absence of myeloblasts was determined. In some cases, misclassification (false positives and false negatives) was observed. This was mainly associated with variations in physiological parameters. According to the experimental results, the overall accuracy of the model ranged from 85% to 92%. For some samples, the accuracy was high. In more complex cases, the results decreased. Graphical analysis showed that there is a relationship between signal intensity and biological indicators. However, this relationship was not linear in all cases. Optimization of the camera angle improved signal quality.

A false-color mapping method was used to visualize the results. The grayscale image was converted into blue tones [7]. High-intensity signals appeared in deep blue, while low-intensity signals appeared in light blue. In some cases, incorrect interpretation of color intensity was observed. Thus, the developed model was validated both mathematically and experimentally. The model demonstrated overall effectiveness. At the same time, its accuracy was found to depend on external factors. Reducing these limitations is an important direction for future research. Thus, the developed model was validated both mathematically and experimentally. The model demonstrated overall effectiveness [8]. At the same time, its accuracy was found to depend on external factors. Reducing these limitations is an important direction for future research. Table 1 shows that the proposed model

provides higher J values compared to the simple method across all angles. The maximum value ($J = 6.571$) is achieved at 10° , indicating the optimal operating condition. Table 1. Dependence of the signal quality criterion (J) on the camera angle:

Angle _{deg}	J _{simple}	J _{model}
0	1.888	3.25
5	2.667	5.00
10	4.000	6.571
15	4.75	6.375
20	3.889	5.875
25	2.600	4.111

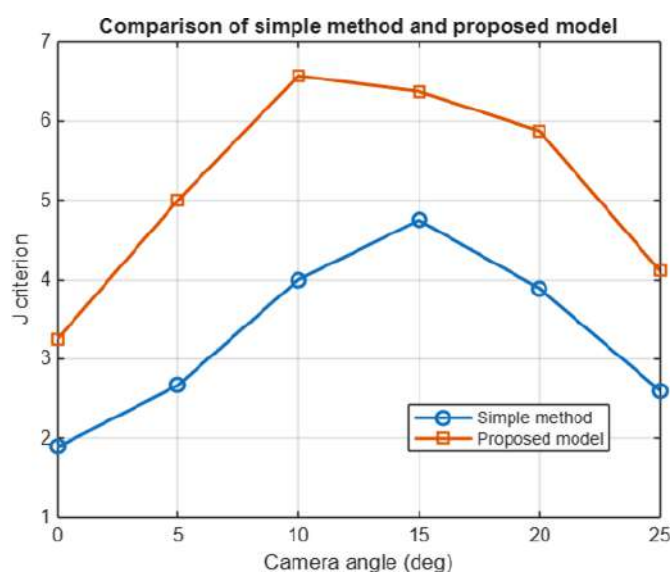


Figure 1. Variation of the signal quality criterion (J) depending on the camera angle and comparison of methods. Best angle (simple method): 15 deg, $J = 4.75$; Best angle (proposed model): 10 deg, $J = 6.571$

Figure 1 shows that the signal quality criterion (J) significantly depends on the camera angle. In the simple method, the J value reached its maximum at an angle of 15° and was equal to 4.75. This indicates that the difference between the signal and the background is optimal at this angle. In the proposed model, the maximum value was observed at an angle of 10° , where $J = 6.57$. This result demonstrates that the model is able to separate the signal more accurately. Higher J values indicate better signal quality. The comparison of results shows that the proposed model performs more effectively than the simple method. It better distinguishes the signal from the background and is less sensitive to noise. It was also observed that the optimal angles differ. This indicates that the model performance depends on the selected algorithm and signal processing methods.

Conclusion

In this study, a diagnostic model based on near-infrared (NIR) spectroscopy data using artificial intelligence was developed. During the research, spectral signal processing, mathematical modeling, and machine learning methods were applied in an integrated manner. The proposed approach is based on the Beer–Lambert law for evaluating optical changes and uses the J criterion to assess signal quality. The results made it possible to accurately evaluate the difference between the signal and the background. The model demonstrated higher performance compared to the simple method. The maximum J value reached 6.57, while for the simple method it was 4.75. The overall accuracy of the model ranged from 85% to 92%. In some cases, a decrease in accuracy was observed due to noise and external factors. The proposed model is suitable for practical application in non-invasive diagnostic systems. It expands the possibilities for early disease detection and simplifies the diagnostic process. From an economic perspective, the model requires low cost since it is based on simple optical devices and software tools. This makes it feasible for large-scale implementation. In the future, it is necessary to further improve the model, expand the dataset, and test it under real clinical conditions. These directions will contribute to increasing the accuracy of the model.

References

- [1]. Chance, B., Nioka, S., Kent, J., McCully, K., Fountain, M., Greenfeld, R., Holtom, G. (1998). Time-resolved spectroscopy of hemoglobin and myoglobin in resting and ischemic muscle. *Analytical Biochemistry*, 174(2), 698–707.
- [2]. Islomov, M., Onarkulov, K., Toshpulatova, F., Toshpulatov, Sh. (2024). Innovative methods in blast cell diagnostics: near infrared light and ptir - microscopy. *Engineering Problems and Innovations*, 2(4), 64-71. <https://epai.uz/index.php/journal/article/view/179>
- [3]. Makhmudova, A., Toshpulatov, Sh., Toshpulatova, F. Matrix photodetector of infrared radiation for measuring leukemia. *Fergana TUIT, Al-Farghani Descendants*, 2023., №3. <https://cyberleninka.ru/article/n/matrichnyy-fotopriyomnik-infrakrasnogo-izlucheniya-dlya-izmereniya-leykoza>
- [4]. Zeng, F., Wu, Q., Chu, X., & Yue, Z. (2015). Measurement of bulk material flow based on laser scanning technology. *Measurement*, 61, 160–168.
- [5]. Ji, J., Miao, C., Li, X., Liu, Y. (2021). Speed regulation strategy for belt conveyor systems based on material flow rate. *PLoS ONE*, 16(2), e0247279. <https://doi.org/10.1371/journal.pone.0247279>
- [6]. Makhmudova, A. J., Toshpulatov, S. M., & Toshpulatova, F. M. (2023). Infrared radiation matrix photodetector for leukemia measurement. *Al-Farg'oni avlodlari*, (3).
- [7]. Toshpulatov, S. M. (2021). Analysis of fiber-optic sensors for diagnostics and monitoring of electrical equipment. *Academicia: An International Multidisciplinary Research Journal*, 11(3), 858-863.