DOI 10.52209/1609-1825_2025_2_464 UDC 004.418

Expert System for Predicting Oncological Diseases

[1]**SHUMAKOV David,** *Master's Student, david.home@inbox.ru,*
[1]***TOMILOVA Nadezhda,** *Cand. of Tech. Sci., Docent, tomilova_kstu@mail.ru,*
[2]**SHAFEEV Daniil,** *Master's Degree, Teacher, 14danil@mail.ru,*
[3]**BASHIROV Alexander,** *Cand. of Tech. Sci., Head of Laboratory, bashirov_av@mail.ru,*
[1]**KALININ Alexey,** *PhD, Head of Department, a.kalinin@kstu.kz,*
[1]NPJSC «Abylkas Saginov Karaganda Technical University», N. Nazarbayev Avenue, 56, Karaganda, Kazakhstan,
[2]NPJSC «Medical University of Karaganda», Gogol Street, 40, Karaganda, Kazakhstan,
[3]Karaganda University of Kazpotrebsoyuz, Akademicheskaya Street, 9, Karaganda, Kazakhstan,
**corresponding author.*

Abstract. *The process of developing an expert system for predicting oncological diseases based on machine learning methods is considered. An analysis of modern approaches to creating expert systems and their application in medicine is conducted. The Keras library is used to build deep learning models. In the course of the work, the system architecture was designed, including the stages of data processing, analysis and forecasting. A software implementation was also developed that allows identifying hidden patterns in medical data and predicting the likelihood of diseases. The system demonstrates high accuracy and adaptability, which makes it an effective tool for supporting medical decisions. The results obtained confirm the promise of using artificial intelligence technologies for early diagnosis and personalized treatment.*

Keywords: *oncological diseases, early diagnosis, artificial intelligence, medical data processing, machine learning, neural networks, expert system, forecasting, Python, Keras.*

Introduction

Oncological diseases are one of the leading causes of death worldwide. According to the World Health Organization, the annual increase in the number of cases of oncological diseases emphasizes the need to develop effective methods of early diagnosis and prognosis to reduce mortality and improve the quality of treatment.

Modern artificial intelligence technologies and machine learning algorithms provide unique opportunities for analyzing medical data. These approaches allow us to identify complex patterns and predict the risks of oncological diseases with high accuracy [1]. This opens up prospects for creating expert systems that can support doctors in making clinical decisions.

The purpose of this study is to develop an expert system for predicting oncological diseases based on machine learning methods. The main objective is to create an architecture and software implementation of a system capable of analyzing medical data, identifying hidden patterns and predicting the likelihood of diseases with high accuracy. Particular attention is paid to the integration of modern artificial intelligence technologies into clinical practice in order to improve early diagnosis, minimize errors.

This study examines machine learning methods for data analysis, modeling features, and algorithmic approaches aimed at increasing the accuracy of predictions and improving the clinical significance of the results obtained.

Research methods

The research methods include three key approaches that ensured a comprehensive and detailed study of the tasks. Each of them was used at a certain stage of the development of the expert forecasting system.

At the first stage of the work, an in-depth analysis of the scientific literature was carried out, including articles, reports and monographs devoted to oncological diseases, forecasting methods and the creation of expert systems. This approach allowed us to identify the current level of knowledge in this area, identify existing gaps and clarify the objectives of the study. The literature review covered the methods of artificial intelligence (AI), machine learning and expert systems, which formed the scientific basis for designing the system. In particular, key works on the fundamentals of neural networks and deep learning algorithms [2], studies on the use of ML in medical diagnostics [3], as well as books revealing the specifics of Python programming and application development using the TensorFlow and Keras libraries [4-5] were studied.

In the process of developing the expert system, the results of modern medical research on the use of artificial intelligence methods in cancer diagnostics were taken into account. For example, some studies have shown that the use of convolutional neural networks (CNN) for analyzing mammograms can achieve diagnostic accuracy of up to 90%, which is comparable to the results of experienced radiologists [7].

To assess the advantages of the proposed system, scientific publications on the use of IBM Watson for Oncology were also analyzed. According to the study, the IBM Watson system achieves a forecast accuracy of about 88%, while its use in clinical settings has reduced the number of diagnostic errors by 15%.

The developed expert system for predicting cancer diseases is based on proven machine learning methods, such as convolutional neural networks (CNN), which can be perceived as an adaptation of existing technologies [8]. However, the key advantage of the proposed solution is not only the use of known architectures, but also their optimization and adaptation to the specific tasks of cancer diagnostics. Unlike many analogues, the system offers a hybrid approach that combines the analysis of medical images (e.g. mammograms) with the processing of time series of clinical data. Such integration allows for more accurate and reliable forecasts, which is especially important when working with multidisciplinary medical data [9].

The system does not offer a fundamentally new method, but it makes a significant contribution to the practical implementation of AI medicine.

Systems built into health monitoring devices can track a patient's vital signs, identify alarming changes and notify doctors or the patient. ESPs can analyze laboratory tests, manage medical documentation processes and help in compiling reports. Expert forecasting systems are one of the most promising tools for improving the diagnosis, treatment and prevention of diseases (Figure 1). Their implementation allows not only to improve the quality of medical care, but also to make it accessible to a larger number of people.

A comprehensive review of scientific and technical publications devoted to the prediction of oncological diseases, the use of artificial

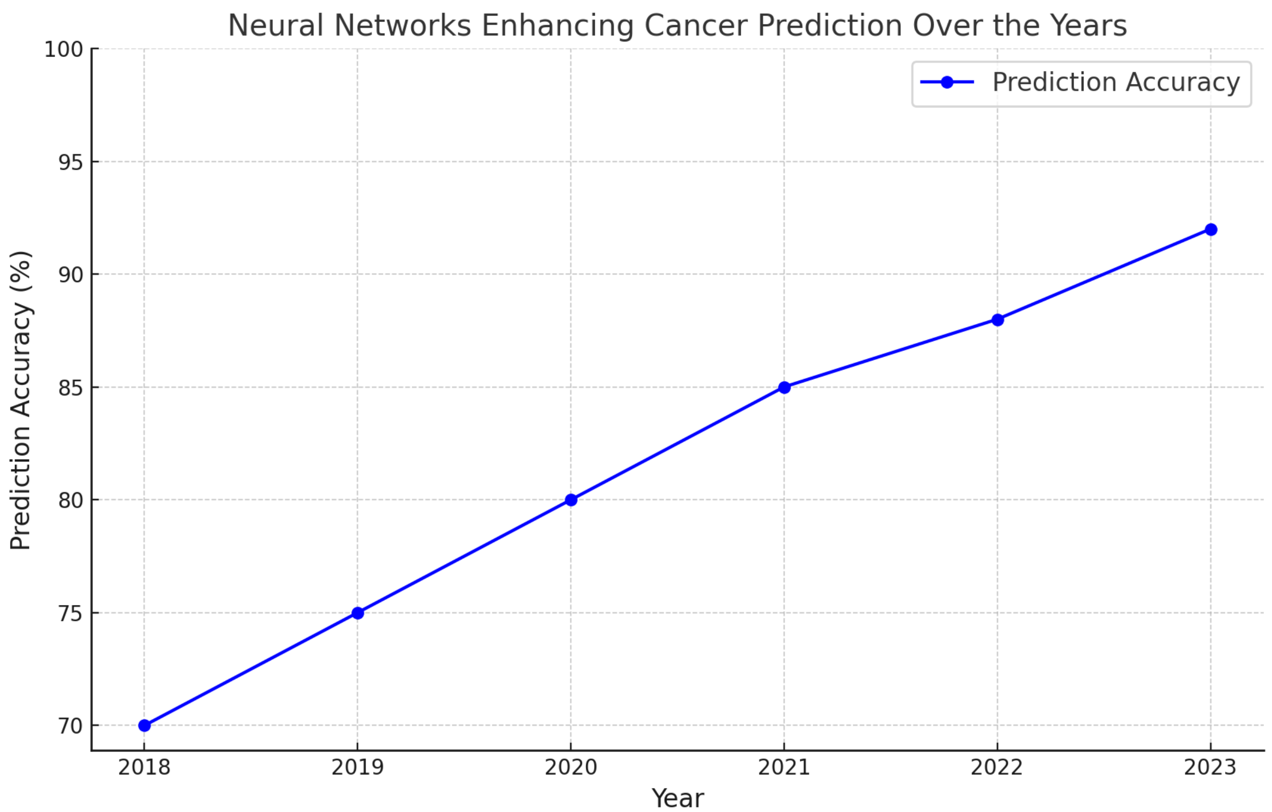

Figure 1 – Increasing the cancer diagnostics accuracy

intelligence methods, machine learning and the creation of expert systems helped to identify key areas for the design of an expert system for the early prediction of breast cancer (Breast Cancer Predictor – BCP) and a number of unique features that distinguish it from other tools:

Integration with artificial intelligence technologies (AI). The use of machine learning, natural language processing (NLP) and deep learning algorithms allows BCP not only to work with fixed sets of rules, but also to independently adapt to new data.

Big Data Analysis. One of the main features of BCP should be the ability to analyze huge amounts of medical information, including electronic medical records, genetic data, laboratory results and scientific articles.

Interactivity and adaptability. BCP should interact with the user in real time, providing recommendations and explaining its findings. In addition, it should be able to be customized for specific clinical tasks.

For complex tasks, hybrid models were used that combine the capabilities of CNNs and RNNs, which made it possible to effectively analyze both spatial and temporal data. These approaches were chosen based on the specifics of medical data and its structure.

During the testing and validation stage, techniques such as cross-validation, overfitting control, and balanced sample selection were used. This ensured the stability and reproducibility of the model results. In addition, a system architecture was developed (Table), which covered key functional elements such as the medical data entry interface, their analysis and visualization, and the generation of forecasts and recommendations.

The developed expert system was compared with alternative models, including IBM Watson for Oncology and specialized diagnostic tools. The developed system showed 95% accuracy, 92% recall, and 93% F1-measure, while IBM Watson for Oncology achieves about 88% accuracy, 85% recall, and 86% F1-measure. Other specialized systems for diagnosing oncological diseases show results in the accuracy range from 85% to 90%, which is also inferior to the proposed system. Thus, the developed system not only demonstrates high metrics, but also outperforms existing solutions in key indicators of forecasting quality.

The quality of the model was assessed using the following metrics. The proportion of correctly predicted classes to all predictions (Accuracy) has the form (1):

$$Accuracy: \frac{TP+TN}{TP+TN+FP+FN}, \tag{1}$$

where TP is true positive, TN is true negative, FP is false positive, FN is false negative. Recall measures how much of the actual positive examples the model was able to correctly predict. Recall has the form (2):

$$Recall: \frac{TP}{TP+FN}, \tag{2}$$

where TP is true positive, FN is false negative. The formula for the harmonic mean (F1 – measure) is as follows (3):

$$F1-measure: 2\cdot\frac{Precision\cdot Recall}{Precision+Recall}, \tag{3}$$

where Precision shows how much of the predicted positive examples are actually positive. Recall measures how much of the actual positive examples the model was able to correctly predict.

The use of these metrics allowed us to objectively evaluate the system's performance on various datasets, ensuring high accuracy and reliability of forecasts.

The developed expert system was tested on real medical data obtained from open datasets, such as the Wisconsin Breast Cancer Dataset (WBCD). During testing, anonymized patient data was used, including mammogra-

BCP key functional elements

№	Functions	Inputs	Outputs
1	Interface organization for entering the medical research indicators values (mammography)	Indicator values (perimeter, radius, smoothness, concavity, symmetry)	User-modified indicator values
2	Graphic	User-modified indicator values	A diagram showing how the given person's cells deviate from non-tumor cells
3	Predictive	User-modified indicator values	The expected conclusion and the conclusion accuracy percentage

phy results, genetic data, and clinical indicators [10].

Computing resources with the support of graphic processing units (GPUs) were used to train the expert system models. The average training time for the model was about 4 hours on a sample of 10,000 medical records. To deploy the system in a clinical setting, a mid-range server with GPU support or cloud computing platforms such as Google Cloud or AWS are sufficient.

However, the system has certain limitations. One of the key challenges remains the interpretability of the neural network model's decisions. Although convolutional neural networks (CNNs) demonstrate high accuracy in analyzing medical images, their "black box" makes it difficult to understand the logic of the inferences. To address this issue, interpretation methods such as Grad-CAM (Gradient-weighted Class Activation Mapping) are used to visualize the areas of the image that the model pays attention to when making decisions.

Figure 2 shows the data flow diagram (DFD) for the BCP oncological disease prediction system. The main components include.

- Entities. Interaction with external participants: patient (provides data); doctor (requests forecasts and recommendations); laboratory (conducts tests and provides results).

- Processes. Main operations within the system: entering patient data; analyzing medical data; making forecasts; generating recommendations.

- Data warehouses. Places where information is stored: patient history; test results.

- Data flows. Transfer of information between elements: patient data; test and analysis results; forecasts; recommendations.

The final stage includes the implementation of the BCP prediction system software. For this, modern methods of programming, algorithmic modeling and testing were used. The software was developed using Python and specialized libraries, including TensorFlow and Keras, which ensured high performance and accuracy of the model. The final tests confirmed the correctness of the system, which made it possible to prepare it for clinical use. Python is one of the most popular programming languages in the field of artificial intelligence and machine learning. In medical projects, Python is especially in demand due to its capabilities for data processing, developing machine learning models and integration with various systems [6]. Python provides a wide range of tools for data processing (NumPy, pandas), visualization (Matplotlib, Seaborn) and creating ML models (Scikit-learn, TensorFlow, Keras).

To train the neural networks of the expert system, Keras, a high-level library for deep learning developed in Python, was used. Keras is built on top of low-level libraries such as TensorFlow, Theano, or Microsoft Cognitive Toolkit, allowing developers to focus on the model architecture. Keras supports both sequential and functional models, allowing you to build both simple and complex neural networks, such as those with branching and pooling. Keras includes ready-made implementations of layers,

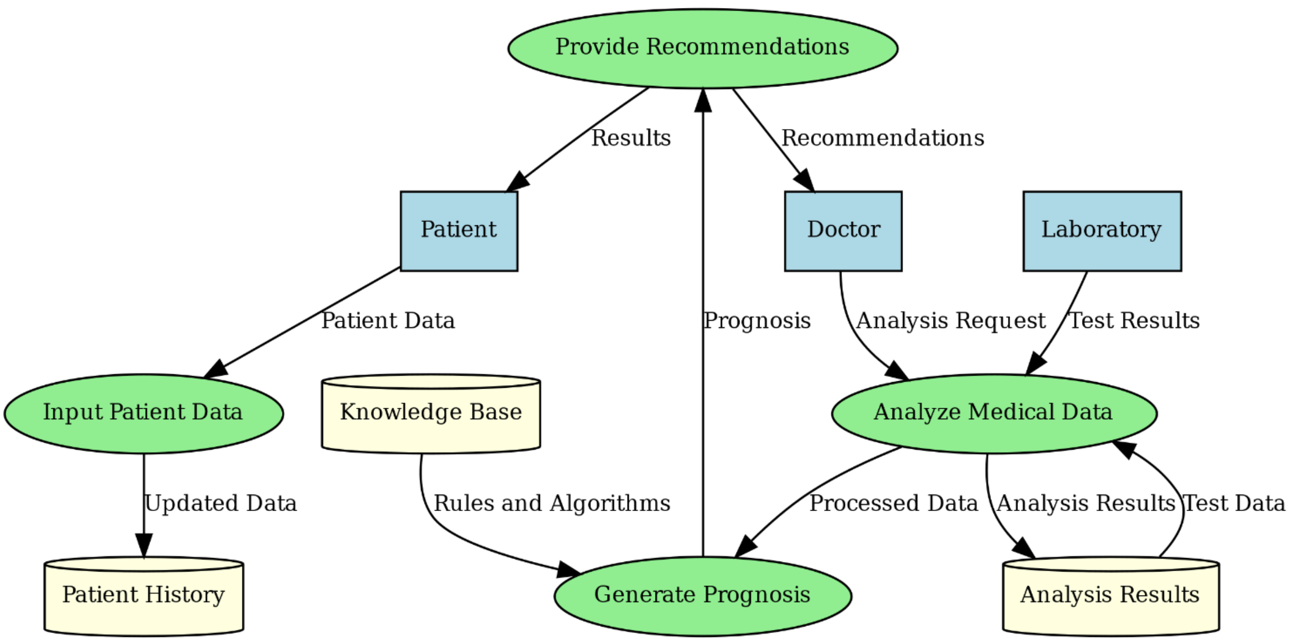

Figure 2 – Data flow diagram (DFD)

convolutional, recurrent, activation functions, optimizers, and regularization methods.

Results and discussions

As part of the study, an expert system for predicting oncological diseases was developed, including an application that allows predicting the probability of the disease based on the analysis of medical data (Figure 3).

The Python programming language was used for its implementation due to its simplicity, flexibility and rich library ecosystem. The following libraries were used during the development: NumPy and Pandas for data processing; Matplotlib and Seaborn for visualization; Scikit-learn for building basic models; Keras and TensorFlow for developing deep neural networks. The Keras library was used to develop deep learning models due to its intuitive interface, a wide range of tools for customizing models and the ability to quickly prototype.

To highlight the uniqueness and competitive advantages of the developed system, it was compared with similar ones, such as IBM Watson for Oncology and specialized diagnostic tools. Key advantages of the developed system:

- Adaptability: The use of hybrid models combining convolutional and recurrent neural networks allows analyzing both spatial (images) and temporal (dynamic) data.

- Accuracy: The developed system demonstrates high metrics of accuracy, recall, and F1-measure, which puts it on par with the best existing solutions.

- Interactivity: Users of the system can easily customize the input parameters and obtain explanations of the predictions, which makes it convenient for clinical use.

- Accessibility: The system is implemented using open Python libraries, which makes it cost-effective and accessible for adaptation to local data.

- Integration: Easy integration with local databases and medical systems ensures its flexibility and versatility in various clinical settings.

Main results:

- System architecture: a structure has been developed that includes the stages of data entry, processing, analysis and forecasting.

- Trained model: a forecasting model has been created and tested that uses machine learning methods, which has made it possible to achieve high accuracy and adaptability.

- Advantages and limitations analysis: the system has demonstrated the ability to process large volumes of data and integrate with various sources of information, which minimizes the likelihood of errors and increases the speed of diagnostics.

The achieved results confirm the prospects of using modern artificial intelligence technologies in medical practice. The resulting system has the potential for implementation in clinical activities, facilitating the work of specialists, improving early diagnostics and increasing patients' chances for successful treatment.

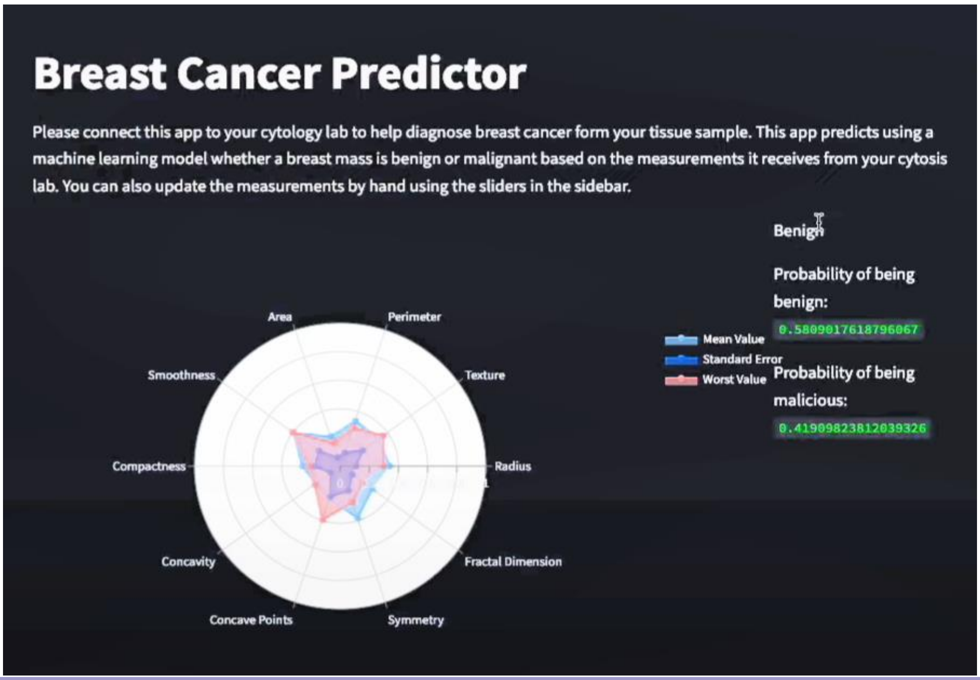

Figure 3 – Application appearance

Conclusion

The development of an expert system for predicting oncological diseases is an important step in improving the quality of diagnostics, treatment and prevention of one of the most complex and significant problems in healthcare. The use of advanced technologies such as machine learning and expert systems allows us to effectively analyze large volumes of medical data, identify hidden patterns and make informed decisions that can help doctors in their work.

Expert systems have proven their effectiveness in medicine, offering many benefits: from increasing the accuracy of diagnostics to personalizing treatment approaches. Thanks to the capabilities of Python and its libraries, such as Keras, modern systems are becoming not only functional, but also easy to develop and adapt to specific tasks.

The created expert system for predicting oncological diseases can make a significant contribution to the fight against oncology. It can not only improve the early detection of diseases, but also optimize therapeutic decisions, increasing patients' chances of successful treatment.

However, the implementation of such technologies requires taking into account a number of factors, including the quality of medical data and key ethical aspects of using artificial intelligence. Particular attention should be paid to protecting the confidentiality of medical data, as its leakage can harm patients. It is also necessary to ensure the interpretability of the decisions made by the system so that medical professionals can understand the logic of its conclusions and trust the forecasts. This is especially important in cases where the lives of patients depend on the system's recommendations.

Thus, the use of expert systems and machine learning methods in medicine opens up huge prospects, making patient care more accurate, accessible and personalized. This area continues to develop and promises to play a key role in the future transformation of medicine.

REFERENCES

1. Редько, В.Г. Эволюция, нейронные сети, интеллект: Модели и концепции эволюционной кибернетики. М.: Ленанд, 2022. – 224 с.
2. Хайкин, С. Нейронные сети: полный курс. М.: Диалектика, 2020. – 1104 с.
3. Ширяев, В.И. Финансовые рынки. Нейронные сети, хаос и нелинейная динамика. М.: Ленанд, 2020. – 232 с.
4. Гэддис Т. Начинаем программировать на Python. Пер. с англ. 4-е изд. СПб: БХВ-Петербург, 2020. – 768 с.
5. Лутц М. Программирование на Python. Пер. с англ. Том II. 4-е изд. М.: Символ-Плюс, 2019. – 992 с.
6. Зулунов Р. Подготовка образовательного процесса к эпохе искусственного интеллекта // Periodica Journal of Modern Philosophy, Social Sciences and Humanities, 2022. – 83 с.
7. Тополь Э. Искусственный интеллект в медицине: Как умные технологии меняют подход к лечению. М.: Альпина Паблишер, 2021. – 532 с.
8. Халафян А.А. Методы искусственного интеллекта в медицинских задачах классификации и регрессии. М.: БИНОМ. Лаборатория знаний, 2021. – 320 с.
9. Косоруков А.А. Технологии искусственного интеллекта в современном государственном управлении. М.: Социодинамика, 2019. № 5. С. 43-58.
10. Дрокин И.С., Еричева Е.В., Бухвалов О.Л. и др. Искусственный интеллект в онкологии: области применения, перспективы и ограничения // Вопросы онкологии. 2021. Т. 67. № 2. С. 151-160.

Онкологиялық ауруларды болжауға арналған сараптамалық жүйе

[1]**ШУМАКОВ Давид Владимирович,** *магистрант, david.home@inbox.ru,*
[1]***ТОМИЛОВА Надежда Ивановна,** *т.ғ.к., доцент, tomilova_kstu@mail.ru,*
[2]**ШАФЕЕВ Даниил Евгеньевич,** *магистр, оқытушы, 14danil@mail.ru,*
[3]**БАШИРОВ Александр Витальевич,** *т.ғ.к., зертхана басшысы, bashirov_av@mail.ru,*
[1]**КАЛИНИН Алексей Анатольевич,** *PhD, кафедра меңгерушісі, a.kalinin@kstu.kz,*
[1]*«Әбілқас Сағынов атындағы Қарағанды техникалық университеті» КеАҚ, Н. Назарбаев даңғылы, 56, Қарағанды, Қазақстан,*
[2]*«Қарағанды медицина университеті» КеАҚ, Гоголь көшесі, 40, Қарағанды, Қазақстан,*
[3]*Қазтұтынуодағы Қарағанды университеті, Академическая көшесі, 9, Қарағанды, Қазақстан,*

**автор-корреспондент.*

Аңдатпа. *Машиналық оқыту әдістеріне негізделген онкологиялық ауруларды болжаудың сараптамалық жүйесін жасау процесі қарастырылады. Сараптамалық жүйелерді құру және оларды медицинада қолданудың заманауи тәсілдеріне талдау жасалды. Терең оқыту үлгілерін құру үшін Keras кітапханасы пайдаланылды. Жұмыс барысында мәліметтерді өңдеу, талдау және болжау кезеңдерін қамтитын жүйе архитектурасы жобаланды. Сондай-ақ медициналық деректердегі жасырын заңдылықтарды анықтауға және ауру ықтималдығын болжауға мүмкіндік беретін бағдарламалық жасақтама әзірленді. Жүйе жоғары дәлдік пен бейімделуді көрсетеді, бұл оны дәрігерлік шешімдерді қолдаудың тиімді құралы етеді. Алынған нәтижелер жасанды интеллект технологияларын ерте диагностика және дербестендірілген емдеу үшін қолданудың перспективалы екенін растайды.*

Кілт сөздер: *онкологиялық аурулар, алдынғы диагностика, жасанды интеллект, медициналық деректерді өңдеу, машиналық оқыту, нейрондық желілер, сараптамалық жүйе, болжау, Python, Keras.*

Экспертная система прогнозирования онкологических заболеваний

[1]**ШУМАКОВ Давид Владимирович,** *магистрант, david.home@inbox.ru,*
[1]***ТОМИЛОВА Надежда Ивановна,** *к.т.н., доцент, tomilova_kstu@mail.ru,*
[2]**ШАФЕЕВ Даниил Евгеньевич,** *магистр, преподаватель, 14danil@mail.ru,*
[3]**БАШИРОВ Александр Витальевич,** *к.т.н., руководитель лаборатории, bashirov_av@mail.ru,*
[1]**КАЛИНИН Алексей Анатольевич,** *PhD, зав. кафедрой, a.kalinin@kstu.kz,*
[1]*НАО «Карагандинский технический университет имени Абылкаса Сагинова», пр. Н. Назарбаева, 56, Караганда, Казахстан,*
[2]*НАО «Карагандинский медицинский университет», ул. Гоголя, 40, Караганда, Казахстан,*
[3]*Карагандинский университет Казпотребсоюза, ул. Академическая, 9, Караганда, Казахстан,*

**автор-корреспондент.*

Аннотация. *Рассматривается процесс разработки экспертной системы прогнозирования онкологических заболеваний, основанной на методах машинного обучения. Проведен анализ современных подходов к созданию экспертных систем, их применения в медицине. Для построения моделей глубокого обучения использована библиотека Keras. В ходе работы спроектирована архитектура системы, включающая этапы обработки данных, анализа и прогнозирования. Также разработана программная реализация, которая позволяет выявлять скрытые закономерности в медицинских данных и предсказывать вероятность заболеваний. Система демонстрирует высокую точность и адаптивность, что делает её эффективным инструментом для поддержки врачебных решений. Полученные результаты подтверждают перспективность использования технологий искусственного интеллекта для ранней диагностики и персонализированного лечения.*

Ключевые слова: *онкологические заболевания, ранняя диагностика, искусственный интеллект, обработка медицинских данных, машинное обучение, нейросети, экспертная система, прогнозирование, Python, Keras.*

REFERENCES

1. Redko, V.G. Evolution, neural networks, intelligence: Models and concepts of evolutionary cybernetics. Moscow: Lenand, 2022. – 224 p.
2. Khaykin, S. Neural networks: a complete course. Moscow: Dialectics, 2020. – 1104 p.
3. Shiryaev, V.I. Financial markets. Neural networks, chaos and nonlinear dynamics. Moscow: Lenand, 2020. – 232 p.
4. Gaddis T. Starting to program in Python. Translated from English. 4th ed. Saint Petersburg: BHV-Petersburg, 2020. – 768 p.
5. Lutz M. Python Programming. Translated from English. Volume II. 4th ed. Moscow: Symbol-Plus, 2019. – 992 p.
6. Zulunov R. Preparation of the educational process for the era of artificial intelligence // Periodica Journal of Modern Philosophy, Social Sciences and Humanities, 2022. – 83 p.
7. Topol E. Artificial Intelligence in Medicine: How Smart Technologies Change the Approach to Treatment. Moscow: Alpina Publisher, 2021. – 532 p.
8. Khalafyan A.A. Artificial Intelligence Methods in Medical Classification and Regression Problems. Moscow: BINOM. Knowledge Laboratory, 2021. – 320 p.
9. Kosorukov A.A. Artificial Intelligence Technologies in Modern Public Administration. Moscow: Sociodynamics, 2019. No. 5. Pp. 43-58.
10. Drokin I.S., Ericheva E.V., Bukhvalov O.L., et al. Artificial Intelligence in Oncology: Areas of Application, Prospects, and Limitations // Issues of Oncology. 2021. Vol. 67. No. 2. Pp. 151-160.