DOI: 10.18721/JCSTCS.13402
УДК 004.852

MODERN POSSIBILITIES OF USING AI METHODS IN THE ANALYSIS OF BIOMEDICAL DATA

I.A. Gasanova[1], *D.S. Prelovskii*[1], *V.A. Yurkin*[1],
P.D. Drobintsev[1], *A.O. Drobintseva*[2]

[1] Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation;
[2] Saint Petersburg State Pediatric Medical University, St. Petersburg, Russian Federation

Nowadays, one of the key indicators that have a great impact on the evolution of society is artificial intelligence. AI and Big Data technologies are widely used to analyze biomedical data. This article describes what artificial intelligence and Big Data are and what are the modern possibilities of using their methods and technologies. The statistics showing the growth in the use of Big Data and AI technologies in medical research are presented. The main types of artificial neural networks used in this area are considered, as well as examples of the successful use of Big Data technologies in medicine. The effectiveness of the use of special computer programs in the field of health care, which allows detecting diseases at early stages, are demonstrated. The key technological and ethical problems of introducing artificial intelligence technologies into medicine are considered, the difficulties of implementation, integration and dissemination of technologies are shown. Special attention is paid to the use of AI in the fight against the global pandemic, the COVID-19 coronavirus infection. The methods of using AI in various countries for collecting data, analyzing and then building a model of the spread and mutation of coronavirus under various scenarios of the development of the situation and the introduction of special restrictive measures, as well as predicting their effectiveness are analyzed.

Keywords: Big Data, predicative analysis, healthcare, cancer detection, neural networks.

Citation: Gasanova I.A., Prelovskii D.S., Yurkin V.A., Drobintsev P.D., Drobintseva A.O. Modern possibilities of using AI methods in the analysis of biomedical data. Computing, Telecommunications and Control, 2020, Vol. 13, No. 4, Pp. 21–33. DOI: 10.18721/JCSTCS.13402

СОВРЕМЕННЫЕ ВОЗМОЖНОСТИ ИСПОЛЬЗОВАНИЯ МЕТОДОВ ИИ В АНАЛИЗЕ БИОМЕДИЦИНСКИХ ДАННЫХ

И.А. Гасанова[1], *Д.С. Преловский*[1], *В.А. Юркин*[1],
П.Д. Дробинцев[1], *А.О. Дробинцева*[2]

[1] Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Российская Федерация;
[2] Санкт-Петербургский государственный педиатрический медицинский университет, Санкт-Петербург, Российская Федерация

На сегодняшний день одним из ключевых показателей, оказывающих большое влияние на эволюцию общества, является искусственный интеллект (ИИ). ИИ и технологии Big Data широко применяются для анализа биомедицинских данных. В статье рассказывается о том, чем является искусственный интеллект и Big Data и каковы современные

возможности использования их методов и технологий. Приведена статистика, показывающая рост использования технологий Big Data и ИИ в медицинских исследованиях. Рассмотрены основные типы искусственных нейронных сетей, используемых в этой области, а также приведены примеры успешного применения технологий Big Data в медицине. Продемонстрирована эффективность использования специальных компьютерных программ в сфере здравоохранения, позволяющих выявлять заболевания на ранних стадиях. Рассмотрены ключевые технологические и этические проблемы внедрения технологий искусственного интеллекта в медицину, показаны трудности реализации, интеграции и распространения технологий. Отдельное внимание уделено использованию ИИ в борьбе с глобальной пандемией – коронавирусной инфекцией COVID-19. Изучены способы применения ИИ в различных странах для сбора данных, анализа и последующего построения модели распространения и мутации коронавируса при различных сценариях развития ситуации и введения специальных ограничительных мер, а также прогнозирование их эффективности.

Ключевые слова: Большие данные, предикативный анализ, здравоохранение, обнаружение рака, нейронные сети.

Ссылка при цитировании: Гасанова И.А., Преловский Д.С., Юркин В.А., Дробинцев П.Д., Дробинцева А.О. Современные возможности использования методов ИИ в анализе биомедицинских данных // Computing, Telecommunications and Control. 2020. Vol. 13. No. 4. Pp. 21–33. DOI: 10.18721/JCSTCS.13404

Introduction

Nowadays, the whole world and his wife talk about Big Data, but only a few actually know what it is and how it works. Big data technology is a set of various tools, approaches, methods for analyzing structured and unstructured information and on its basis the subsequent production of new information to solve specific problems and goals defined by the customer (market, consumer, etc.). In 2011, research and consulting company Gartner noted Big Data as the number two trend in the information technology infrastructure (after virtualization) [1]. At present, Big Data technologies can no longer be unambiguously attributed to the IT sphere, since they have become an important part of the structures of management, business, industry, medicine and science, constantly increasing their share of participation.

According to the Russian IT holding IBS, in May 2015, the global amount of data exceeded 6.5 zettabytes. By 2025, it is predicted that humanity will generate 400–440 zettabytes of information, according to a report by The Data Age 2025, which was prepared by analysts at the American research company IDC [2]. The report notes that most of the data will be generated by the enterprises themselves, and not ordinary consumers. By 2027, the global Big Data market is projected to grow to $ 103 billion, more than double its expected market size in 2018. The growth is projected to be up to (about) 8 billion US dollars per year, or 10–15 % of the total market turnover. The largest share of 45 % will be achieved in the field of software production, and will be the largest share on the market [3] (Fig. 1).

Big Data is a part of the information that can be obtained by digitizing the entire spectrum of sources available to us, which can be processed by any means and recorded using sensors. The rationality and prospects of the application of Big Data technologies in medicine, as well as in the whole health concept in recent years has been widely discussed by the professional community. In the field of healthcare, Big Data technologies are used to analyze and process information about the health status of a population or individuals, and potentially increase the likelihood of timely identification of diseases and predispositions to them stored in the human genome. Because prevention is always cheaper than treatment, such technologies have great potential, and will soon have to be applied to every person. In most medical institutions of the future, all medical information about a person from the moment of his birth will have to be stored in an

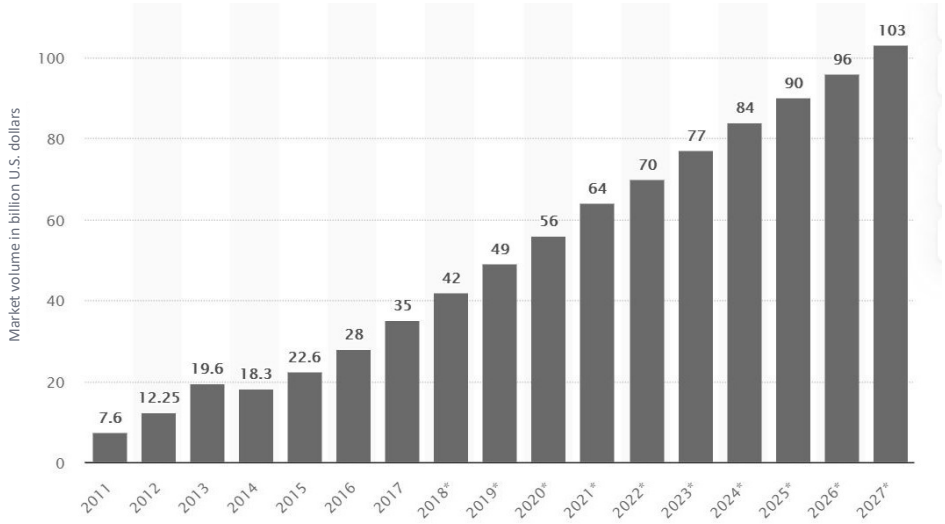

Fig. 1. Big Data market size revenue forecast worldwide from 2011 to 2027

electronic database. The issuance of forecasts, medical indications and recommendations for patients will be promptly implemented through the use of machine learning algorithms that can determine statistical correlations in an array of medical data collected from medical institutions around the world.

The initial analysis of the data and the collection of necessary information about the patient can be transferred from a person to an automated system, thanks to this, a significant part of the time can be saved. For the correct diagnosis of the disease, a large amount of data about the patient is required, namely: his purchases, diagnoses, risks of the condition, etc. Having performed this analysis, we will come to a new personalized and preventive medicine. So, thanks to the results of using Big Data, it is quite possible to predict the risks of mental or psychological diseases of a particular person and prevent diseases such as depression, psychoses, etc. These methods of individual and preventive medicine, which is based on monitoring patients remotely, will entail significant cost reductions, as well as increase the quality of life of the observed patients. For example, a medicament will not have any effect on 90 % of the observed patients, however, a certain group of patients must be offered it directly, and to another group, on the contrary, it is required to refuse this medicine, since the risk of an allergic reaction is increased. Another example is the medical examination on a certain group of people (or all, without exception) once a year, and personally, for each patient according to an individual schedule developed in connection with the risks of his condition. In the modern world, a single doctor has a large number of patients; accordingly, having received some background information, he may not notice that the patient is cutting off episodes of his life, which can be crucial. For example, in order to identify the distant spread of cancer on the bone, information about various injuries is needed, but due to the workload of the doctor and, accordingly, understatement on the part of the patient, information about the absence of injuries can be obtained, but at the same time asking a more correct and accurate question, you can find out the opposite.

Recently, the use of Big Data technologies in pharmaceutics has gained considerable importance. At the moment, Big Data tools and AI technologies are actively used in the production and marketing of medicines. Nowadays, the main 4 areas of implementation of Big Data technologies in pharmaceuticals are actively discussed, these include:

1. The development of new drugs. Big Data technology focuses on finding trends and models that otherwise would be difficult, expensive and even impossible to find using traditional methods of collecting and processing information [4].

2. Collection of clinical information about patients. Among pharmaceutical manufacturers, there is fierce competition for the right to get first access to patients medical data. Obtaining such information is characterized by transactions with large technology companies engaged in research in the field of database analysis. For example, the holding pharmaceutical company Roshe Group collaborated for a long time, and then acquired the Flatiron Health company, which is the market leader in the collection of clinical information specializing in cancer research [5].

3. Improving the quality and control of clinical trials. Corporations can increase the effectiveness of clinical trials through the use of Big Data methods. For example, searching for patients with certain parameters in a large database.

4. Detection of intolerance to medical drugs. The use of an analytical system with Big Data approaches for segregating drugs with certain medical properties can reduce costs for companies and save the lives of more patients [6].

In addition, one of the cutting-edge discoveries that contribute to monitoring the health of each person is various devices (for example, fitness trackers, smart watches). These devices contain a number of sensors that monitor the pulse, blood pressure, counting steps, in addition, through these gadgets it is possible to control the rate of breathing, as well as create a varied and balanced nutrition menu. All information from these devices can be instantly transferred to the appropriate application on the phone in order to keep all your indicators under special control, which will allow the user to draw up a daily regimen, nutrition, help monitor physical activity, water balance, and health in general. In some cases, this information can save the user's life, but if the user is healthy, the zettabytes of the collected data form continuously expanding information databases. The interdisciplinary interaction of Big Data specialists and medical professionals is facilitated by the high information potential of wearable smart devices. In the middle of 2015, two largest US corporations, Apple and IBM, decided to consolidate their efforts in the use of Big Data technologies in healthcare [7]. The companies have integrated the data received from owners of iPhone smartphones and Apple Watch smart watches with the IBM Watson Health cognitive system. The above examples of the use of smart devices for collecting medical indicators show that this area is being intensively developed by Big Data technology specialists. This is a rapidly developing area which has great potential to expand.

The provided statistics demonstrates high growth rate of the use of Big Data technologies in medical research. Large financial investments emphasize the high interest of medical and IT companies in collaborative research, the result of which is to increase the efficiency of the use of resources in the healthcare sector and improve medical statistics data.

Artificial Intelligence Algorithms for Big Data analysis

Since the size of medical data is growing rapidly, humanity comes to the logical conclusion that our health and quality of life depend on the speed and quality of their analysis. And that all this is a job for artificial intelligence. In this article, by the term AI, we understand the ability of machines to model rational behavior of people, that is, their ability to navigate in a changing context, on the basis of which they makes the most appropriate choice to achieve their goals.

To date, two AI technological processes are widely used: expert systems and neural networks. At the same time, expert systems are becoming obsolete, neural networks have flooded the market due to their ability to learn. A neural network means a sequence of neurons, which are connected to each other by synapses that serve to transmit a nerve impulse. In a computer form, artificial neural networks (ANNs) represent a graph with three or more layers of neurons. The layers are interconnected according to the selected algorithm, depending on the task. Each link has its own specific weight. These weights play a crucial role in training of an ANN. Most often, ANNs are used to solve such types of problems as classification, prediction and recognition, etc.[1]. There are quite a few types of neural networks for solving problems specific to various fields. This article describes different types of basic ANNs.

[1] Rashchenko J.V. Application of modified artificial neural networks in machine vision problems. 2018. Available: http://hdl.handle.net/11701/12215 (Accessed: 15.01.2021).

1. Shallow neural networks. This type of ANN consists of only 1 hidden layer of the perceptron. The number of neurons in the hidden layer may vary. Collaborative filtering is considered one of the examples of this type of ANN: a method of forecasting in recommendation systems. In medicine, this type of ANN can be used to create a classifier that can distinguish cancer and manage patients from mass spectrometry data [8].

2. Multilayer perceptron is a type of direct distribution ANNs. They consist of many input nodes of one or more hidden layers and one output layer of neurons. The number of input and output elements in a multilayer perceptron is determined by the conditions of the problem. This type of ANN is used in the tasks of forecasting, managing agents of classification, approximation, etc.

3. Convolutional neural network (CNN) is a type of deep ANNs, most often used for the analysis of visual images. It uses certain characteristic features of the visual cortex, in which two types of cells with different features were discovered. Simple cells tend to react to straight lines at different angles, and complex – to activate a specific set of simple cells.The network structure is unidirectional, basically multilayer. It is used in problems with image recognition, classification, analysis of medical images, natural language processing, etc. [9]. There are various CNN architectures available, which were key in the construction of algorithms that provide and should ensure the operation of AI as a whole in the foreseeable future. Some of them are listed below: LeNet, AlexNet, VGGNet, GoogLeNet, ResNet, ZFNet.

4. A recurrent neural network (RNN) is a class of artificial neural networks, where connections between nodes form a directed graph along a time sequence. This allows it to process a series of events over time. RNNs are called recurrent because they perform the same task for each element of the sequence, and the output depends on previous calculations. Recurrent neural networks add memory to artificial neural networks, but the implemented memory is short – at each step of the training, the information in memory is mixed with new and after several iterations is completely overwritten. It is applied to such tasks as non-segmented handwriting recognition with connection or speech recognition.

5. A network with long-term and short-term memory is an artificial recurrent neural network used in the field of deep learning, capable of studying long-term dependencies. Long short-term memory (LSTM) modules are designed specifically to avoid the problem of long-term dependency by storing values for both short and long periods of time. This is because the LSTM module does not use the activation function inside its recurrent components. Thus, with the method of back propagation of errors in time during network training, the stored value does not blur in time and the gradient does not disappear [10]. It is used in tasks related to image processing, video, and speech recognition.

6. Networks based on attention. The key point of the proposed attention-based model is that it takes into account the influence (interconnection) that exists between different parts or words or the whole one input sentence with another, and provides an interdependent representation of a pair of sentences that can be used in subsequent tasks. Many hierarchical patterns of attention are called transformers. These transformers are most effective for stacks running in parallel, so they provide the most innovative results with relatively minimal data and a period of time for training. This model is mainly used when working with text data.

7. Generative-competitive network is a machine-learning algorithm without a trainer, created in the composition of 2 neural networks, one of which, called a generator, creates new instances, and the other, the discriminator, evaluates them for authenticity, seeks to distinguish true (genuine) examples of samples from erroneous. Widely used in tasks related to the generation of images, video and audio. In a sense, they are robotic artists, and their result is impressive, but they can also be used to create fake media content [11].

Currently, there is a tendency to increase the number of publications on the successful use of neural networks in parallel with the development of AI technologies based on neural networks in the healthcare sector. Despite many years of attempts at application, the main directions of the use of neural networks remain unchanged and include: classification, forecasting and diagnostics. The most commonly used

type of ANNs in these tasks is a neural network with direct connection. In addition, the study revealed a significant share of hybrid models used.

The aforementioned examples of artificial intelligence usage in the medical field show the leading position of two types of neural networks: hybrid and feedback. Both of them are used to solve the main problems of data classification, recognition of medical markers on images and data arrays, and predicting the occurrence of relevant diseases.

Big Data case studies

Consequently, AI may (at least at the current stage of the technology development) not replace a doctor, but be – has already become – a useful tool, an assistant in the process of diagnosis and treatment. The examples are listed below.

One striking example is IBM Watson for Oncology (WFO), a program based on the power of the IBM Watson supercomputer that uses natural language processing and machine learning to provide recommendations for the treatment of cancer. In the course of its work, it is involved in data processing (structured and unstructured) from literature, medical records, images, all kinds of pathology reports, laboratory data, expert knowledge, etc. [12]. Each patient, with the help of IBM Watson, is offered a personalized course of treatment, which is aimed at those factors that led to the development of a cancerous tumor in a particular person, taking into account his genome. This allows increasing the chances of a successful recovery, and each new case increases the likelihood even more. The Watson Cancer application has already been used to help more than 20,000 people around the world.

A study launched in China found that different types of cancer are common various conformities as recommended by doctors. For example, recommendations for treating gastric cancer had the lowest chance of being agreed with your healthcare providers. Morbidity and pharmaceuticals may be the main causes of disagreement. In order to comprehensively and quickly apply this technology in China, WFO needs to accelerate localization [13]. At the moment, another project from IBM Watson is under development – IBM Medical Sieve – a cognitive assistant system for radiologists and cardiologists that can detect abnormalities using holistic clinical information obtained from images, text or medical data [14].

In 2016, Optellum startup was created. They are developing lung cancer prediction AI (Optellum LCP) software for the diagnosis and treatment based on computed tomography. The company's solution is based on artificial intelligence and machine learning technologies used in the world's largest clinical data set. The software provides expert-level clinical decision support, which can improve the doctor's ability to correctly diagnose nodules in the lungs [15].

Modern classifiers based on convolutional neural networks are able to systematize images of skin cancer along with dermatologists, which makes it possible to guarantee an early diagnosis of the disease. Software products that use this type of ANN are already trained using a large data set; during their work, they improve the data characteristics for systematizing skin lesions and demonstrate better efficiency with currently available limited data sets [16]. Currently, there is an O-network under development, a convolutional network that is a combination of two convolutional U-networks. These networks are combined on the coding layer and disconnected at the decoding level, which allows the convolution/reverse convolution process to be used with another core of the structure, which is aimed at searching for elements with different parameters. This type of architecture shows promising results in the field of detection of breast cancer using mammography, but needs to be improved [17].

Due to the rapid population growth over the past years, prevention and control of lung cancer has become increasingly important. The use of risk factors for lung cancer in the elderly and their quantitative analysis of the degree of influence using deep neural network models can be used as a tool for identifying risk factors, and for other types of cancer to help doctors who make decisions on cancer prevention diseases [18].

Adverse reactions to medications are becoming more common and are considered the main factor causing more than a million injuries, disabilities, anomalies and deaths per year. The best known method for identifying adverse drug reactions is to rely on spontaneous reports of them. A model has now been developed based on a fully connected neural network that uses chemical and biological information and biomedical drug data to detect adverse reactions. The purpose of this product is to identify potentially dangerous reactions to the drug and to predict possible adverse reactions to new drugs. The results of the work show that the use of this model helps to detect potential reactions to drugs, regardless of whether they were registered in the past [19].

The significance of research in the diagnosis of lung cancer is associated with social relevance, this type of disease and its leading position in the structure of cancer. An intelligent automated diagnostic system using a combined algorithm based on machine learning methods, such as deep forest and Siamese neural network, was developed in the scientific cluster of the St. Petersburg Peter the Great Polytechnic University. It is a more effective approach with a small sample of training data and optimal in terms of reproducibility. The algorithm uses the data of the St. Petersburg Clinical Scientific and Practical Center for Specialized Types of Medical Care (Oncology), which includes only computer tomograms of patients with verified diagnoses of lung cancer. The developed software is considered innovative and has already been patented. The software provides the ability to quickly recognize and detect all neoplasms without exception, which after processing the data are highlighted in a certain color, which allows the system to interpret the neoplasm as malignant or benign. Comparing information about one patient with a large amount of previously accumulated knowledge, the model discovers relationships that previous methods or doctors overlooked [20].

There is a common problem for many studies, which impedes the full realization of their potential. Such a problem for researchers working in the field of artificial intelligence is the quality of the data they use. AI can analyze and interpret scanned images or pathological images only if the data are of good quality and in large volume. Therefore, if the data quality is poor, then AI systems generate inaccurate and biased results.

The implementation of Big Data analysis technologies can significantly improve the quality and speed of diseases detection in the early stages according to medical data. This information in turn entails a reduction in the burden on doctors, providing the opportunity to redistribute efforts to other problems requiring human involvement.

Problems, difficulties of implementation, dissemination of Big Data technology in healthcare

It is important to mention, that the use of artificial intelligence algorithms in dermatology and general medicine is not fiction from distant future, but a part of modern reality. Doctors who are the first to take advantage of new opportunities will take their diagnostic skills to a whole new level and significantly increase the prestige in the eyes of patients. Once this happened with ultrasound diagnostic devices, then the turn of popularity of computed tomography came.

A Big Data system, usually characterized by size, variety, reliability and speed, goes beyond the boundaries of a data type and contains the nuances of data analysis, such as generating hypotheses, but not hypothesis testing. Big medical data has a number of features that distinguish them from Big Data from other disciplines. Big Data technology has many applications in healthcare, such as predictive modeling and clinical decision support, disease or safety surveillance, public health and research.

Integration of heterogeneous data sources (medical records, research data, laboratory test results, the Internet, biometric readings, medical images, etc.) leads to an increase in data size, incompleteness and complexity [21]. To cope with the growth of data volumes, organizations turn to a number of different technologies, such as compression, deduplication, etc. Database support requires constant funding, which creates an additional barrier to the implementation of Big Data [22]. The idea of data integration is closely related to the idea of data validation. There are cases when there are no values in the data set, and as a rule,

the easiest way is to delete incomplete information. The disadvantage of this method is the reduction in the amount of data required for analysis, which leads to inefficiency.

Big Data has a number of problems associated with the data itself, which should be taken into account when analyzing them.

The presence of a linear relationship between variables, the complexity of the model, the re-equipment of its data with large computational costs accompany the technology of Big Data and affect most quantitative analyses.

In the healthcare sector, confidentiality and security of patients data is an important aspect, because large data warehouses can attract the attention of fraudsters [23].

Another problem in introducing Big Data technologies is a lack of specialists in this field, which has led to an increase in demand for the Big Data experts [24].

The companies with experience in working with Big Data are for the most part satisfied with the results of their work, however, without exception, everyone faces difficulties in maintaining competitiveness and turning into digital enterprises [25].

Many problems, including legal, ethical, integration ones, must be overcome in order to implement large medical data as a combustible material for a continuously learning healthcare system that will help improve the condition of patients and reduce time and material costs in this area [26].

The adoption of artificial intelligence technologies in other fields of science can contribute to implementation problems solution, which leads to increase in the number of specialists in this field. The quality of medical data will be inevitably improved with the growth and spread of electronic medical systems in the world. The development of data storing and analyzing devices will lead to a reduction in the cost of this technology and an increase in the efficiency of its use in the future.

AI approaches in COVID-19 pandemic

The current global COVID-19 pandemic is the most acute topic around the world. Government authorities in all countries are trying to take measures to contain and fight the virus. A significant role in this process is taken by mathematical models, algorithms and methods in the field of technologies for human movement and identification.

To determine the COVID-19 markers, specialized software is used. For instance, Infervision, which is based on deep learning technology, using the results of computed tomography in the process. While it usually takes a few minutes for a healthcare professional to read and explain the results of a CT scan, a dedicated AI application can complete such the procedure in seconds. The system is already implemented in over 100 hospitals in China and in many countries worldwide [27].

We can also note the use of Big Data technologies in the iOmniscient's iQ-Fevercheck system, which uses cameras with temperature and recognition sensors. This software contributes to the detection of people susceptible to infection in real time in the stream, with further fixation and tracking of the face by cameras. This process takes places until they are next to a medical worker who receives a message about an infected person for further assistance. Protective medical masks do not interfere with tracking, and the system's temperature recording accuracy is up to 0.5 degrees [28].

In addition to the topic of the coronavirus, it is important to note the key factors that underlie care for healthcare providers and allied professionals. These factors are the transparency of the data and the adequacy of statistics. According to them, it is possible to determine the features of a specific extreme situation, which determines the necessary measures for their timely adoption. At the current moment, many interactive systems that provide a web interface for tracking operational information on the situation with COVID-19 have been created. For example, Yandex presented its dashboard with a global picture and detailing by regions of Russia, which uses Russian Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing (Rospotrebnadzor) reports and data from Johns Hopkins University in the USA [29].

In the Republic of South Korea, potential carriers of coronavirus infection are monitored using mobile phones and satellite technology. A similar application works in the Russian Federation, for example, on the territory of Moscow: the social monitoring application. It is used primarily by those who have been diagnosed with coronavirus infection and by people who cohabit with such patients. Patients sign a consent to receive home healthcare or an order from the Chief Sanitary doctor. In these documents, you must indicate the phone number and address at which they undertake to be isolated within 14 calendar days from the date of signing. After signing the contract, instructions for installing the application are sent to the patient's phone. In case patients have a phone, they are given a smartphone that is focused only on this application, it will not be possible to use it for other purposes, all other functions are blocked. Technical means of monitoring the location make it possible to quickly record cases when a person who is required to be isolated leaves the house and endangers other citizens of the city.

At this moment, we can note the widespread use of machine learning and other Big Data methods in the development of vaccines and drugs. For example, the AlphaFold neural network predicted the possible structure of some COVID-19 proteins using matrix population modeling technology. The synthesis of drugs for the treatment of diseases caused by coronavirus requires precise knowledge of the component and structural composition of the virus. The data obtained during the research have not been experimentally confirmed at the current stage, but based on the available materials, scientists can judge the functioning of the virus, which allows them to create hypotheses regarding the effectiveness of a particular therapy.

This pandemic has several applications for Big Data technologies, for which AI is of key importance. These examples include biomedical research, natural language processing, social networks.

Based on an epidemiological simulator with the implementation of the SEIR mathematical compartmental model, which can help outline the dynamics of the spread of the disease, Sberbank predicted an increase in the number of infected people in accordance with the use of protective measures [30]. This model calculates data both for the capital and for the whole of Russia and considers four possible scenarios when residents comply with the measures, sometimes or often violate them, or behave as usual.

Conclusion

Peter the Great St. Petersburg Polytechnic University researchers were able to predict the development of the Coronavirus epidemic using mathematical modeling methods. Periodically incoming new data give an ability to correct and refine the forecast. For the developed mathematical model, the so-called calibration of the mathematical model of the spread of the coronavirus infection is used. During the work, the specialists step by step examined and described with a high degree of accuracy the situation regarding the ongoing spread of the coronavirus infection in the Wuhan metropolitan area in China, on the Diamond Princess liner, in Italy and in the USA. This, in turn, made it possible to significantly improve the forecasting accuracy in Russia. After the adoption of measures to reduce restrictive and control measures due to the slowdown in the spread of the infection, the process of mathematical modeling and forecasting for the specialists of the NTI (National Technology Initiative) Center has become much more complicated. However, experts are closely monitoring the changing situation and adjusting forecasts taking into account its development.

REFERENCES

1. **Thibodeau P.** Gartner's Top 10 IT challenges include exiting baby boomers, Big Data. *Computerworld*, Oct. 18, 2011.

2. **Reznichenko V.A.** What is Big Data. *Problems in Programming*, 2019, No. 3, Pp. 86–100.

3. Statistical research carried out by Arne Holst, Mar 2, 2020 and published on the portal [Statistical study by Arne Holst, Mar 2, 2020 and published on Statista]. Available: *https://www.statista.com/statistics/254266/global-big-data-market-forecast/* (Accessed: 15.01.2021).

4. **McGraw B.** Reversing the Traditional Drug Development Model. Available: *https://www.perkinelmer.com/PDFs/downloads/PerkinElmer-White-Paper-Reversing-the-Traditional-Drug-Development-Model.pdf/* (Accessed: 15.01.2021).

5. **Philippidis A.** Precision oncology-focused roche to acquire flatiron health for $1.9 B. *Clinical OMICs*, 2018, Vol. 5, No. 2, Pp. 47–47.

6. **Aggarwal M., Madhukar M.** IBM's Watson analytics for health care: A miracle made true. *Cloud Computing Systems and Applications in Healthcare*. IGI Global, 2017. Pp. 117–134.

7. **Shaloha E.R.** Neural networks. *Proc. of the 73rd Student Scientific and Practical Conference*, Minsk: Belarusian Technical University, 2017, Pp. 33–35. (rus)

8. **Voronov I.V., Politov E.A., Efremenko V.M.** Review of types of artificial neural networks and methods of their training. Bulletin of KuzSTU, 2007, No. 3. Available: *https://cyberleninka.ru/article/n/obzor-tipov-iskusstvennyh-neyronnyh-setey-i-metodov-ih-obucheniya* (Accessed: 15.01.2021). (rus)

9. **Bakker B.** Reinforcement learning by backpropagation through an LSTM model/critic. *2007 IEEE International Symposium on Approximate Dynamic Programming and Reinforcement Learning*. IEEE, 2007, Pp. 127–134.

10. **Mogren O.** C-RNN-GAN: Continuous recurrent neural networks with adversarial training. *arXiv preprint arXiv:1611.09904*. 2016.

11. **Somashekhar S.P., et al.** Early experience with IBM Watson for Oncology (WFO) cognitive computing system for lung and colorectal cancer treatment. *Journal of Clinical Oncology*, 2017, No. 35, Pp. 8527–8527. DOI: 10.1200/JCO.2017.35.15_suppl.8527

12. **Zhou N., et al.** Concordance study between IBM Watson for Oncology and clinical practice for patients with cancer in China. *The Oncologist*, 2019, Vol. 24, No. 6, Pp. 812–819.

13. **Zhang L., et al.** Machine learning and visual computing. *Applied Computational Intelligence and Soft Computing*, 2017, 1–1. DOI: 10.1155/2017/7571043

14. **Baldwin D.R., et al.** External validation of a convolutional neural network artificial intelligence tool to predict malignancy in pulmonary nodules. *Thorax*, 2020, Vol. 75, No. 4, Pp. 306–312.

15. **Brinker T.J., et al.** Skin cancer classification using convolutional neural networks: Systematic review. *Journal of Medical Internet Research*, 2018, Vol. 20, No. 10, e11936.

16. **Rashed E., El Seoud M.S.A.** Deep learning approach for breast cancer diagnosis. *Proceedings of the 2019 8th International Conference on Software and Information Engineering*, 2019, Pp. 243–247.

17. **Chen S., Wu S.** Identifying lung cancer risk factors in the elderly using deep neural networks: Quantitative analysis of Web-Based Survey Data. *Journal of Medical Internet Research*, 2020, Vol. 22, No. 3, e17695.

18. **Wang C.S., et al.** Detecting potential adverse drug reactions using a deep neural network model. *Journal of Medical Internet Research*, 2019, Vol. 21, No. 2, e11016.

19. **Moiseenko B.M., Meldo A.A., Utkin L.V., Prokhorov I.Yu., Ryabinin M.A., Bogdanov A.A.** Automated detection system for lung masses as a stage of artificial intelligence development in the diagnostics of lung cancer. *Diagnostic Radiology and Radiotherapy*, 2018, No. 3, Pp. 62–68. (rus) DOI: https://doi.org/10.22328/2079-5343-2018-9-3-62-68

20. **Nepravskiy A.A.** Osobennosti vnedreniya i ispolzovaniya Big Data v razlichnykh oblastyakh deyatelnosti [Features of the implementation and use of Big Data in various fields of activity]. *Intellektualnyy Kapital v Ekonomike Znaniy* [*Intellectual Capital in the Knowledge Economy*], 2017, Pp. 56–68. (rus)

21. **Mamedova M.G.** Big Data in electronic medicine: Opportunities, challenges and perspectives. *İnformasiya texnologiyaları problemləri* [*Problems of information technology*], 2016, No. 2, Pp. 9–29.(rus)

22. **Anderson K.** *Analytical culture: From data collection to business results*. Moscow: "Mann, Ivanov and Ferber" Publ., 2017. Pp. 111–140. (rus)

23. **Paskova A.A.** Big Data technologies in the automation technological and business processes. *Nauchnoye obozreniye. Tekhnicheskiye nauki* [*Scientific Review. Technical Science*], 2018, No. 4, Pp. 23–27. (rus)

24. **Rozymov U.** Tsifrovizatsiya kak faktor konkurentosposobnosti Rossiyskikh kompaniy [Digitalization as a factor in the competitiveness of Russian companies]. *Nauchnyye i tvorcheskiye dostizheniya v ramkakh sovremennykh obrazovatelnykh standartov* [*Proceedings of the Conference on Scientific and Creative Achievements within the Framework of Modern Educational Standards*], 2019, Pp. 170–173. (rus)

25. **Lee C.H., Yoon H.J.** Medical Big Data: Promise and challenges. *Kidney Research and Clinical Practice*, 2017, Vol. 36, No. 1, P. 3.

26. **McCall B.** COVID-19 and artificial intelligence: Protecting health-care workers and curbing the spread. *The Lancet Digital Health*, 2020, Vol. 2, No. 4, e166–e167.

27. **Agbehadji I.E., et al.** Review of Big Data analytics, artificial intelligence and nature-inspired computing models towards accurate detection of COVID-19 pandemic cases and contact tracing. *International Journal of Environmental Research and Public Health*, 2020, Vol. 17, No. 15, P. 5330.

28. **Tsibikov V.A.** O pervoocherednykh merakh organov gosudarstvennoy vlasti Rossii v svyazi s rasprostraneniyem koronavirusnoy infektsii [About the priority measures taken by the state authorities of Russia in connection with the spread of coronavirus infection]. *Gosudarstvennaya sluzhba*, 2020, Vol. 22, No. 2, Pp. 42–47. (rus)

29. **Liu W., et al.** COVID-19: Attacks the 1-beta chain of hemoglobin and captures the porphyrin to inhibit human heme metabolism. Preprint revised on 12.07.2020. Available: *https://chemrxiv.org/ndownloader/files/22283226*

30. New Defence Order Strategy. Available: *//https://dfnc.ru/covid-19/kak-matematicheskoe-modelirovanie-mozhet-pomoch-v-borbe-s-koronovirusom/* (Accessed: 22.12.2020).

Received 22.12.2020.

СПИСОК ЛИТЕРАТУРЫ

1. **Thibodeau P.** Gartner's Top 10 IT challenges include exiting baby boomers, Big Data // Computerworld. Oct. 18, 2011.

2. **Reznichenko V.A.** What is Big Data // Problems in programming. 2019. No. 3. Pp. 86–100.

3. Статистическое исследование, проведённое Arne Holst, Mar 2, 2020 и опубликованное на портале Statista // URL: https://www.statista.com/statistics/254266/global-big-data-market-forecast/ (Дата обращения: 15.01.2021).

4. **McGraw B.** Reversing the Traditional Drug Development Model // URL: https://www.perkinelmer.com/PDFs/downloads/PerkinElmer-White-Paper-Reversing-the-Traditional-Drug-Development-Model.pdf/ (Дата обращения: 15.01.2021).

5. **Philippidis A.** Precision oncology-focused roche to acquire flatiron health for $1.9 B // Clinical OMICs. 2018. Vol. 5. No. 2. Pp. 47–47.

6. **Aggarwal M., Madhukar M.** IBM's Watson analytics for health care: A miracle made true // Cloud Computing Systems and Applications in Healthcare. IGI Global, 2017. Pp. 117–134.

7. **Шалоха Е.Р.** Нейронные сети // Матер. 73-й студенческой науч.-практ. конф. Минск: БНТУ, 2017. С. 33–35.

8. **Воронов И.В., Политов Е.А., Ефременко В.М.** Обзор типов искусственных нейронных сетей и методов их обучения // Вестник Кузбасского государственного технического университета. 2007. № 3. С. 38–42.

9. **Bakker B.** Reinforcement learning by backpropagation through an LSTM model/critic // 2007 IEEE Internat. Symp. on Approximate Dynamic Programming and Reinforcement Learning. IEEE, 2007. Pp. 127–134.

10. **Mogren O.** C-RNN-GAN: Continuous recurrent neural networks with adversarial training // arXiv preprint arXiv:1611.09904. 2016.

11. **Somashekhar S.P., et al.** Early experience with IBM Watson for Oncology (WFO) cognitive computing system for lung and colorectal cancer treatment // J. of Clinical Oncology. 2017. No. 35. Pp. 8527–8527. DOI: 10.1200/JCO.2017.35.15_suppl.8527

12. **Zhou N., et al.** Concordance study between IBM Watson for Oncology and clinical practice for patients with cancer in China // The Oncologist. 2019. Vol. 24. No. 6. Pp. 812–819.

13. **Zhang L., et al.** Machine learning and visual computing // Applied Computational Intelligence and Soft Computing. 2017. 1–1. DOI: 10.1155/2017/7571043

14. **Baldwin D.R., et al.** External validation of a convolutional neural network artificial intelligence tool to predict malignancy in pulmonary nodules // Thorax. 2020. Vol. 75. No. 4. Pp. 306–312.

15. **Brinker T.J., et al.** Skin cancer classification using convolutional neural networks: Systematic review // J. of Medical Internet Research. 2018. Vol. 20. No. 10. A. e11936.

16. **Rashed E., El Seoud M.S.A.** Deep learning approach for breast cancer diagnosis // Proc. of the 2019 8th Internat. Conf. on Software and Information Engineering. 2019. Pp. 243–247.

17. **Chen S., Wu S.** Identifying lung cancer risk factors in the elderly using deep neural networks: Quantitative analysis of Web-Based Survey Data // J. of Medical Internet Research. 2020. Vol. 22. No. 3. e17695.

18. **Wang C.S., et al.** Detecting potential adverse drug reactions using a deep neural network model // J. of Medical Internet Research. 2019. Vol. 21. No. 2. e11016.

19. **Моисеенко В.М., Мелдо А.А., Уткин Л.В., Прохоров И.Ю., Рябинин М.А., Богданов А.А.** Автоматизированная система обнаружения объемных образований в легких как этап развития искусственного интеллекта в диагностике рака легкого // Лучевая диагностика и терапия. 2018. № 3. С. 62–68. DOI: https://doi.org/10.22328/2079-5343-2018-9-3-62-68

20. **Неправский А.А.** Особенности внедрения и использования Big Data в различных областях деятельности // Интеллектуальный капитал в экономике знаний. 2017. С. 5–68.

21. **Мамедова М.Г.** Big Data в электронной медицине: возможности, вызовы и перспективы // İnformasiya texnologiyaları problemləri. 2016. № 2. С. 9–29.

22. **Андерсон К.** Аналитическая культура: от сбора данных до бизнес-результатов. М.: МИФ, 2017. С. 111–140.

23. **Паскова А.А.** Технологии Big Data в автоматизации технологических и бизнес-процессов // Научное обозрение. Технические науки. 2018. № 4. С. 23–27.

24. **Розымов У.** Цифровизация как фактор конкурентоспособности российских компаний // Научные и творческие достижения в рамках современных образовательных стандартов: Сб. трудов конф. 2019. С. 170–173.

25. **Lee C.H., Yoon H.J.** Medical Big Data: Promise and challenges // Kidney Research and Clinical Practice. 2017. Vol. 36. No. 1. P. 3.

26. **McCall B.** COVID-19 and artificial intelligence: Protecting health-care workers and curbing the spread // The Lancet Digital Health. 2020. Vol. 2. No. 4. e166–e167.

27. **Agbehadji I.E., et al.** Review of Big Data analytics, artificial intelligence and nature-inspired computing models towards accurate detection of COVID-19 pandemic cases and contact tracing // Internat. J. of Environmental Research and Public Health. 2020. Vol. 17. No. 15. P. 5330.

28. **Цибиков В.А.** О первоочередных мерах органов государственной власти России в связи с распространением коронавирусной инфекции // Государственная служба. 2020. Т. 22. № 2. С. 42–47.

29. **Liu W., et al.** COVID-19: Attacks the 1-beta chain of hemoglobin and captures the porphyrin to inhibit human heme metabolism. Preprint revised on 12.07.2020 // URL: https://chemrxiv.org/ndownloader/files/22283226

30. New Defence Order Strategy// URL: //https://dfnc.ru/covid-19/kak-matematicheskoe-modelirovanie-mozhet-pomoch-v-borbe-s-koronovirusom/ Internet resource (Дата обращения: 22.12.2020).

Статья поступила в редакцию 22.12.2020.

THE AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Gasanova Ilakha A.
Гасанова Илаха Адалет Гызы
E-mail: ilaha.gasanova@yandex.ru

Prelovskii Dmitrii S.
Преловский Дмитрий Сергеевич
E-mail: dimaprelovski@gmail.com

Yurkin Vladimir A.
Юркин Владимир Андреевич
E-mail: arhimondled@yandex.ru

Drobintsev Pavel D.
Дробинцев Павел Дмитриевич
E-mail: drob@ics2.ecd.spbstu.ru

Drobintseva Anna O.
Дробинцева Анна Олеговна
E-mail: anna.drobintseva@gmail.com