INFORMATION SYSTEM FOR RADIOBIOLOGICAL STUDIES

I.A. Kolesnikova[1,2], A.V. Nechaevskiy[1], D.V. Podgainy[1], A.V. Stadnik[1,2], A.I. Streltsov[3], O.I. Streltsova[1,2]

[1] *Joint Institute for Nuclear Research, Dubna, Russia*

[2] *Federal State-Funded Educational Institution of Higher Education of Moscow Region "Dubna University", Dubna, Russia*

[3] *SAP SE, Germany*

E-mail: podgainy@jinr.ru

The article discusses the concept of building an information system (IS) for radiobiological studies underway at the Laboratory of Radiation Biology of JINR. The information system under development should have the following properties: to provide storage and access to experimental data and methods of its processing, to contain a set of methods for experimental results systematization and for the detection of hidden patterns that appear in the response of biological systems to the effects of damaging factors, to ensure data presentation in a form convenient for complex statistical analysis, to provide opportunities for research automation based on machine and deep learning methods and neural network approaches, as well as a comfortable environment for interaction and collaboration of different research groups. The implementation of this system will significantly enhance the efficiency of radiobiological studies.

Keywords: information systems, neuromorphology, machine learning, data analysis, neural networks, behavioral responses, tracking.

Inna Kolesnikova, Andrey Nechaevskiy, Dmitry Podgainy, Alexey Stadnik, Alexej Streltsov, Oksana Streltsova

1. Introduction

Modern information technologies and nuclear medicine are essential components of advances in medical technologies. Studies in this field are impossible without high-performance computing complexes, adequate mathematical support and software. The general rapid development of technologies and research in the field of the neural network approach and deep learning leads to the emergence of new developments for the automation of medical and biological studies. Technologies of machine learning and deep neural networks for creating systems of automated diagnosis of diseases have the most practical application [1-5].

At present, the use of proton beam therapy for tumor treatment is the most promising direction in radiation oncology. To make up an optimal plan for proton therapy of small deep-seated brain tumors, it is highly important to detect possible morphological and functional changes in normal and tumor tissues of the central nervous system (CNS) that can be caused by exposure to ionizing radiation. It is related both to the rather compact special arrangement of many brain structures (particularly in the limbic system) and to the set of a number of important functions that they possess.

This kind of the analysis of the damaging effect of ionizing radiation on tissues and organs is also applicable to the assessment of risks of human space flights. Recent studies by Russian and international experts indicate that radiation exposure of brain structures to heavy charged particles (HCP) and high-energy protons of cosmic origin can result in cognitive impairment. This, in turn, entails a partial or complete loss of the operator's functions of spacecraft crew members. A new strategy for planning further experimental work on modeling the biological effects of space types of radiation and on assessing the risks of their damaging effects in the condition of human interplanetary flights is the organization of complex neuro-radiobiological studies on the effect of heavy charged particles on the CNS [6].

All the risks described above make investigations aimed at studying the pathogenesis in different body tissues after exposure to ionizing radiation extremely relevant. The development and study of new methods of pharmacological protection against radiation damage, the testing of radiation-resistant materials and the elaboration of prevention methods can be the solution to the existing problems. The process of organizing this kind of studies is complex and rather laborious, for example, it takes 7-10 days from the moment of collecting tissues for research to receiving a report from a pathologist. At least two groups of animals, 10 rodents each, participate in experiments within scientific studies, i.e. at least 20 drugs need to be analyzed. With such a method of data processing, the human factor plays a key role (individual characteristics, pathomorphologist's competence).

In the light of the above, the automation of the entire process of working with data from the moment of carrying out an experiment to the detection and visualization of the obtained patterns and models of the ongoing processes is of particular relevance.

2. Structure and methods for the implementation of the information system

One of the reasons for the laboriousness of automation is the complexity in the analysis of heterogeneous experimental data, which may include: morphological data (images of slices of different biological tissues), behavioral data (video data of experimental animals) and others obtained by different research groups. A complete understanding of the process of exposure and a qualitative picture of the consequences of exposure to ionizing radiation on biosystems require the systematization and simultaneous processing of a significant amount of data related to different aspects of the manifestation of exposure.

A schematic diagram of the radiobiological research workflow and the corresponding data flows is shown in the figure below.
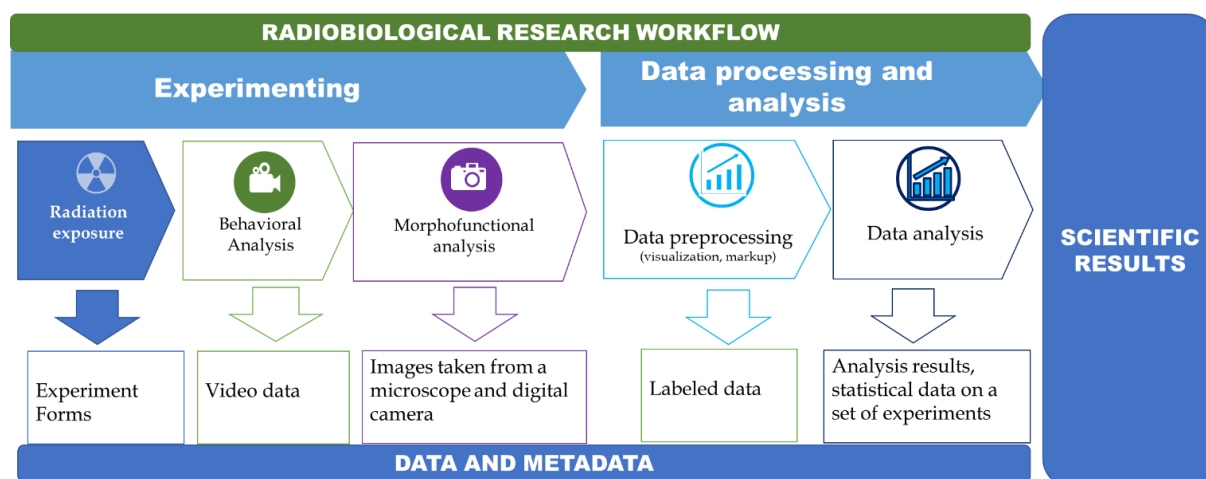

Fig. 1. Radiobiological research workflow schema

The heterogeneity of experimental data defines the creation of a developed subsystem for acquiring, storing and systematizing experimental data, which is capable of working with data with the same efficiency for both conducted and current experiments when studying the effects of ionizing radiation and other factors on biological objects.

A complete picture and a model of the ongoing processes require the development of a qualitative set of algorithms for experimental data processing based on machine and deep learning methods for the tasks of pathomorphology and behavioral analysis in the study of the effects of ionizing radiation and other factors.

Close interaction of different research groups and the requirement to ensure information security when accessing data and research results entail the application of a maximum of modern IT solutions, including web technologies, reliable modern means of authentication and hierarchical access management, as well as components for convenient operation and visualization of data analysis results.

The first component of the information system, ensuring work with data, should correspond in its structure to the logic of conducted radiobiological experiments. It dictates both the hierarchical structure of data storage and the corresponding organization of an interface for interacting with the user of the information system. Schematically, such an organization can be illustrated as follows:

Experiment is the major element inside of which all the contents of the experiment will be stored. It has a standard set of attributes: name, creator's name, creation date, description; and a special attribute such as exposure (for example, gamma radiation).

Group is an auxiliary element related to a specific experiment, inside of which data on "objects" of research (experimental animals) will be stored. In addition to the standard attributes, it has special ones: organ under study, dye, microscope magnification, drug.

"Object" of research (experimental animal) is an auxiliary element related to a specific group; metadata of the "objects" of the experiment (photos and video files) will be stored inside.

The client-side (frontend) is the graphical interface that the user sees on the page. It is responsible for the appearance of the components (stylization) and their position on the page (layout). With the help of the frontend, the user can interact with the server-side (backend) and the database using HTTP requests to the API. From the user's side it happens unnoticed and feels like, for example, pressing a button, entering data into a form, clicking on a page

element. The following technologies are used as means for implementing the frontend of the service: HTML5/SCSS, JavaScript, WebPack, React.js, Redux.

The component of the information system related to the algorithmic part can be divided into two large subgroups:

1. Subgroup of computer vision methods associated with morphological data processing (images of slices of different biological tissues);

2. The second subgroup of methods comprises methods for analyzing video sequences in the study of behavioral patterns of experimental animals (video data of experiments on animal behavior).

Modern tasks of computer vision are conventionally divided into:

Classification – classification of an image by the type of object;

Semantic segmentation – identification of all pixels of objects of a certain type or background in the image. When objects of the same class overlap, their pixels are not separated from each other;

Object detection – detection of all objects of the specified classes and definition of the size and position of the rectangular image area containing them;

Instance segmentation – identification of pixels belonging to each object of each class separately.

The tasks of semantic and instance segmentation are in demand for the tasks of the first subgroup of methods.

Image processing is based on a set of classical computer vision algorithms (image filtering, statistical analysis methods) and techniques for using deep convolutional neural network architectures, among which the following should be considered:

1. Mask R-CNN: convolutional neural network architecture for instance segmentation in images. It extends the Faster R-CNN architecture by adding another branch that predicts the position of the mask covering the found object, thus solving the task of instance segmentation. The mask represents a simple rectangular matrix, in which ones in the cells indicate that the corresponding pixel belongs to the object of the specified class, and zeros respectively mask the pixels that do not belong to the object. In the architecture of this network one conventionally distinguishes a separate convolutional neural network for calculating image features, the so-called backbone, and a head, i.e. the union of parts responsible for predicting the position and size of the object, classifying the object and defining its mask.

The Mask R-CNN shows high results in instance segmentation and object detection [7].

2. U-Net: convolutional neural network architecture designed for image segmentation, which is originally developed for biomedical imaging. The network architecture is a sequence of convolution + pooling layers, which first reduce the spatial resolution of the image and then increase it, having previously combined it with image data and passed through other convolutional layers, which turns the neural network into a kind of complex filter that carries out instance segmentation. The U-Net convolutional network architecture performs very well in machine learning competitions. In this case, it can be used not only for segmentation, but also for object detection in images [8].

3. Xception: compact deep neural network. It extends the Inception architecture, the idea of which is to eliminate the choice of the size of the convolution kernel by taking several options and using them, while combining the results. This increases the number of operations that need to be performed to calculate the activations of one layer, therefore, before each convolutional block, a convolution with a 1x1 kernel is done, which reduces the dimension of the signal fed to the input to convolutions with large kernel sizes.

The architecture has proven itself well in its versatility and performance [9].

In the second subgroup of techniques, the study of video data entails additional methods of work using classical computer vision methods, such as the multivariate Gaussian distribution for assessing the static background of a scene, the analysis of the optical flow to evaluate and predict movement in frames, algorithms for interframe tracking of an experimental animal to assess the trajectory of its movement and its significant parameters. Some methods can be taken from the open computer vision library OpenCV [10].

The development of these algorithms is planned to be performed on the "ML/DL Ecosystem", which provides wide opportunities both for the development of mathematical models and algorithms and for carrying out resource-intensive calculations and data analysis on the basis of the HybriLIT platform [11]. The created ecosystem has two components: the first component is designed for the development of models and algorithms based on JupyterHub, i.e. a multi-user platform for working with Jupyter Notebook (known as IPython with the ability to work in a web browser); the second component is intended for performing resource-intensive, massively parallel calculations, for example, for training neural networks using NVIDIA graphics accelerators.

The ecosystem enables the development of services based on ML\DL algorithms, the debugging process of the corresponding software and provides visualization tools for the results of experimental data analysis, as well as allows implementing different modern approaches for data analysis, image and video processing. The services developed on the basis of the ecosystem allow users to access the computing resources of the "Govorun" supercomputer for conducting massively parallel calculations [12].

The third component of the information system is used for the tasks of implementing the information system based on modern IT solutions, including web technologies, modern solutions for inference and components for the visualization of data analysis.

The choice of the client-server architecture for the tasks being solved is determined by the enumerated features of the project. Using the web service, users will be able to interact with the database (to add data, delete or edit), to analyze images and video materials on high-performance computing resources and obtain the results of processing in a convenient form. To implement the web service, a technology stack based on the Node.js platform and the React.js framework and the MySQL database management system to store data have been chosen. Node.js is a cross-platform backend JavaScript runtime environment that executes backend JavaScript scenarios. Node.js provides the ability to implement one's own web application server and write a REST API, i.e. a special interface through which data is exchanged between the client and the server via the HTTP protocol, and interaction with the database and the file system of the server itself is performed. This allows one to organize the acquisition and storage of experimental data and store the entire amount of downloaded digital data directly in the server file system. React is a JavaScript library for user interface development. Using HTTP requests to the REST API, the client receives data from the server, and the data is processed and used to build an interface. React enables the implementation of a convenient web interface and, with its help, the provision of user access to the REST API.

Conclusion

At present, the team of authors is actively developing the information system in accordance with the methods and approaches considered in the previous sections. The reader can find a detailed description of what has already been done in each of the indicated directions in the articles of the present collection.

The designed and implemented IS will make it possible to perform a comprehensive analysis of heterogeneous experimental data, including from different research groups, and to

automate most of the work within data analysis and results presentation, which, as a result, will accelerate the acquisition of qualitatively new results.

In conclusion, it is noteworthy that the information system under development can be used not only for medical and research purposes, but also for education. The IS will enable further testing of specialists on histology of nervous tissue to confirm their qualification, moreover, the IS can also be used as a training system for training new personnel in the field of radiobiology.

References

[1] R.A. Tomakova, S.A. Philist, S.A. Gorbatenko, N.A. Shvetsova, Analysis of histological images using morphological operators synthesized on the basis of the Fourier transform and neural network modeling. // Automatic analysis and image recognition, no.3 (9), 2010, pp. 54-60, in Russian. https://cyberleninka.ru/article/n/analiz-gistologicheskih-izobrazheniy-posredstvom-morfologicheskih-operatorov-sintezirovannyh-na-osnovepreobrazovaniya-furie-i/viewer

[2] M.M. Lukashevich, V.V. Starovoitov, Technique for counting the number of cell nuclei in medical histological images. // System analysis and applied informatics, 2, 2016, pp. 37-42, in Russian. https://cyberleninka.ru/article/n/metodika-podscheta-chisla-yader-kletok-na-meditsinskih-gistologicheskih-izobrazheniyah/viewer

[3] Eulenberg, P., Köhler, N., Blasi, T. et al. Reconstructing cell cycle and disease progression using deep learning. Nat Commun 8, 463 (2017). https://doi.org/10.1038/s41467-017-00623-3

[4] AlphaFold: Using AI for scientific discovery. https://deepmind.com/blog/alphafold/

[5] Masayasu Toratani, Masamitsu Konno et al. A Convolutional Neural Network Uses Microscopic Images to Differentiate between Mouse and Human Cell Lines and Their Radioresistant Clones. DOI: 10.1158/0008-5472.CAN-18-0653 Published December 2018

[6] O. Taranina, RAS Council on Space. New concept of risk. // Joint Institute for Nuclear Research weekly, № 51-52, 25.12.2017. http://jinrmag.jinr.ru/2017/51/kr51.htm

[7] Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick, Mask R-CNN. https://arxiv.org/abs/1703.06870

[8] Olaf Ronneberger, Philipp Fischer, Thomas Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation. https://arxiv.org/abs/1505.04597

[9] François Chollet, Xception: Deep Learning with Depthwise Separable Convolutions. https://arxiv.org/abs/1610.02357

[10] Open Source Computer Vision Library. Electronic resource: https://opencv.org/

[11] HybriLIT heterogeneous computing platform. Electronic resource: http://hlit.jinr.ru/ecosystem-for-ml_dl_bigdataanalysis-tasks

[12] Gh. Adam et al., IT-ecosystem of the HybriLIT heterogeneous platform for high-performance computing and training of IT-specialists // CEUR Workshop Proceedings, Selected Papers of the 8th International Conference «Distributed Computing and Grid-technologies in Science and Education» (GRID 2018), Dubna, Russia, September 10-14, 2018, http://ceur-ws.org/Vol-2267/638-644-paper-122.pdf