https://doi.org/10.23934/2223-9022-2020-9-2-231-237

Creation of MSCT-Images Collection and Clinical Data for Acute Cerebrovascular Events

F.A. Sharifullin[1, 2]*, D.D. Dolotova[3], T.G. Barmina[1], S.S. Petrikov[1], L.S. Kokov[1, 2], G.R. Ramazanov[1], Y.R. Blagosklonova[3], I.V. Arkhipov[3], I.M. Skorobogach[1], N.N. Cheremushkin[1], V.V. Donitova[4], B.A. Kobrinsky[4], A.V. Gavrilov[3, 5]

Department of Diagnostic Radiology
[1] N.V. Sklifosovsky Research Institute for Emergency Medicine
3 B. Suharevskaya Sq., Moscow 129090, Russian Federation
[2] I.M. Sechenov First Moscow State Medical University of the Ministry of Health of the Russian Federation (Sechenov University)
8 b. 2 Trubetskaya St., Moscow 119991, Russian Federation
[3] Gammamed-Soft, LLC
11 3th Samotechny Pereulok, Moscow 127473, Russian Federation
[4] Federal Research Center "Informatics and Management" of the Russian Academy of Sciences
44 b. 2 Vavilova St., Moscow 119333, Russian Federation
[5] D.V. Skobeltsyn Research Institute of Nuclear Physics, M.V. Lomonosov Moscow State University
1 b. 58 Leninskiye Gory, Moscow 119991, Russian Federation

*** Contacts:** Faat A. Sharifullin, Dr. Med. Sci., Chief Researcher, Department of Diagnostic Radiology, N.V. Sklifosovsky Research Institute for Emergency Medicine. Email: drfaat@narod.ru

BACKGROUND The use of neuroimaging methods is an integral part of the process of assisting patients with acute cerebrovascular events (ACVE), while computed tomography (CT) is the «gold standard» for examining this category of patients. The capabilities of the analysis of CT images may be significantly expanded using modern methods of machine learning, including the application of the principles of radiomics. However, since the use of these methods requires the availability of large arrays of DICOM (Digital Imaging and Communications in Medicine)-images, their introduction into clinical practice is limited with the problem of a set of representative samples. In addition, at present, collections (datasets) containing CT images of patients with stroke, which would be suitable for machine learning, are practically not available in the public domain.
AIM OF STUDY In this connection, the aim of this work was to create a collection of the DICOM -image native CT and CT-angiography in patients with different types of stroke.
MATERIAL AND METHODS The basis for the creation of the collection was the medical history of patients hospitalized in the Regional Vascular Center of the N.V. Sklifosovsky Research Institute for Emergency Medicine. We used previously developed specialized platform to enter clinical data on the incidence of stroke, attached to each case of DICOM-image studies, as well as delineate and tag (mark) 3D area of interest. For tagging, a dictionary was developed, which elements describe the type of pathological formation, location and system of blood supply.
RESULTS In the course of the work, a collection of clinical cases and images was formed, including anonymous information about 220 patients, 130 with ischemic stroke, 40 with hemorrhagic stroke, and 50 patients without cerebrovascular disorders. These clinics included information about the type of stroke, the presence of concomitant diseases and complications, the duration of hospitalization, the method of treatment and outcomes. For all patients the results of 370 studies of native CT and 102 studies of CT-angiography were entered. On each series of images, an expert physician outlined and tested areas of interest corresponding to direct and indirect signs of stroke.
CONCLUSION The generated collection of images will subsequently allow various methods of data analysis and machine learning to be used in solving the most important practical problems, including diagnosis of the type of stroke, assessment of lesion volume, and prediction of the degree of neurological deficit.
Keywords: data set, stroke, computed tomography, DICOM-images, radiomics, machine learning
For citation Sharifullin FA, Dolotova DD, Barmina TG, Petrikov SS, Kokov LS, Ramazanov GR, et al. Creation of a Collection of MSCT-Images and Clinical Data for Acute Cerebrovascular Events. *Russian Sklifosovsky Journal of Emergency Medical Care.* 2020;9(2):231–237. https://doi.org/10.23934/2223-9022-2020-9-2-231-237 (in Russ.)
Conflict of interest Authors declare lack of the conflicts of interests
Acknowledgments, sponsorship This work was financially supported by the Russian Federal Property Fund in the framework of the scientific project No. 18-29-26007mk
Affiliations

Faat A. Sharifullin	Dr. Med. Sci., Chief Researcher, Department of Diagnostic Radiology, N.V. Sklifosovsky Research Institute for Emergency Medicine; https://orcid.org/0000-0001-7483-7899, drfaat@narod.ru; 10%, organizing data collection, writing a manuscript
Darya D. Dolotova	Cand. Tec. Sci., Leading Research Assistant, Gammamed-Soft LLC; https://orcid.org/, dariadolotova@gmail.com; 10%, research design development, analysis of collected data, manuscript text writing
Tatyana G. Barmina	Cand. Tec. Sci., Senior Researcher, Department of Diagnostic Radiology, N.V. Sklifosovsky Research Institute for Emergency Medicine; https://orcid.org/0000-0003-2418-680X, barminat@inbox.ru; 10%, analysis of CT images
Sergey S. Petrikov	Corresponding Member of the Russian Academy of Sciences, Dr. Med. Sci., Director of N.V. Sklifosovsky Research Institute for Emergency Medicine; https://orcid.org/0000-0003-3292-8789, petrikovss@sklif.mos.ru; 8%, final approval of manuscript publication
Leonid S. Kokov	Corresponding Member of the Russian Academy of Sciences, Professor, Dr. Med. Sci., Head of Department of Diagnostic Radiology, N.V. Sklifosovsky Research Institute for Emergency Medicine; https://orcid.org/0000-0002-3167-3692, lskokov@mail.ru; 8%, manuscript text editing
Ganipa R. Ramazanov	Head of the Scientific Department of Emergency Neurology and Reconstructive Treatment of N.V. Sklifosovsky Research Institute for Emergency Medicine; https://orcid.org/0000-0001-6824-4114, ramazanovgr@sklif.mos.ru; 8%, data collection organization

Evgenia R. Blagosklonova	Researcher, Gammamed-Soft LLC; https://orcid.org/0000-0002-467806oH, evromsp@gmail.com; 8%, software development for the collection, data collection, analysis of data collected
Ivan V. Arkhipov	Leading Programmer, Gammamed-Soft LLC; https://orcid.org/0000-0003-4278-2285, arkhivania@gmail.com; 8%, software development for the collection and working with images
Ivan M. Skorobogach	Resident of the N.V. Sklifosovsky Research Institute for Emergency Medicine; https://orcid.org/0000-0002-5428-6687, dr.skoroboga@gmail.com; 6%, data collection
Nikolay N. Cheremushkin	Resident of the N.V. Sklifosovsky Research Institute for Emergency Medicine of the Moscow Health Department; https://orcid.org/0000-0003-0208-3033, tiforge.invest@gmail.com; 6%, data collection
Viktoria V. Donitova	Researcher, FRC Computer Science and Management of RAS; https://orcid.org/0000-0002-7838-584H, vdonitova@gmail.com; 6%, analysis of collected data priority
Boris A. Kobrinsky	Dr. Med. Sci., Professor, Head of the Department of Support Systems for Clinical Decisions, Institute of Modern Information Technologies in Medicine, FRC Computer Science and Management of RAS; https://orcid.org/0000-0002-3459-8851, evromsp@gmail.com; 6%, software development for the collection, data collection, analysis of collected data
Andrey V. Gavrilov	Cand. Tec. Sci., Head of the Laboratory of Medical Computer Systems, D.V. Skobeltsyn Research Institute of Nuclear Physics of M.V. Lomonosov Moscow State University; https://orcid.org/0000-0002-7838-584X, agavrilov49@gmail.com; 6%, research concept and design development

ACVE - acute cerebrovascular event
ASPECTS - Alberta Stroke Program Early CT Score
CT - computed tomography
DICOM - Digital Imaging and Communications in Medicine
DSS - decision support system
HS - hemorrhagic stroke
IS - ischemic stroke
MRI - magnetic-resonance imaging
MSCT - multislice computed tomography
NIHSS - National Institutes of Health Stroke Scale

INTRODUCTION

Acute cerebrovascular events (ACVE) is one of the leading causes of morbidity, mortality and disability in Russia and worldwide. More than 450 000 cases of stroke are reported in Russia annually, while this mortality in the acute period of stroke reaches 35%, increasing to 12-15% by the end of the first year; and in within 5 years after stroke 44% of patients die. Post-stroke disability occupies first place among all causes of disability and is 3.2 to 10 000 population [1-5].

One of the perspective areas of optimization of stroke diagnosis is introduction of decision support systems making (DSS), in that those on the basis of application of methods of machine learning, on the stage of interpretation of radiological images. The relevance of this direction is due to a number of factors. Although at the high saturation of medical facilities devices CT of the entire territory of Russia, it noted a deficit of personnel. Besides that, in large medical centers, where a large number of patients are examined 24/7, there is a fatigue factor and weakening of concentration of attention. In such situations, the presence of DSS might be to minimize the impact of these factors on the quality of the provision of medical assistance.

It is known, that for the construction of any DSS representative sampling is necessary. In most research projects authors have to first carry out the collection of data, which requires considerable time and organizational resources. In some cases, developers are trying to use images, collected in a medical institution for the past few years, however, work with them requires a thorough check of each clinical case on the subject of its compliance with the criteria of inclusion and registration of factors, that the protocols of conducting patients and treatment plans might change [6–7]. Sometimes, the researchers used the collection, published in the open access, but they are often just contain a small number of observations [8]. Besides that, the accompanying clinical information, which importance cannot be underestimated, may be not given at all or provided with a very short list of variables. At the present time a very limited number of datasets, containing *DICOM*-images of CT studies of patients with stroke are in open access.

In connection with the foregoing, **the aim of the study** was to create collections of MSCT images and clinical data of patients with acute stroke.

MATERIAL AND METHODS

A set of data in the collection was made from the histories of patients, hospitalized in the Regional Vascular Center of N.V. Sklifosovsky Research Institute for Emergency Medicine in 2016-2019. The inclusion criteria for patients were:

a) age older than 18 years ;
b) the main diagnosis: ischemic stroke (IS) or hemorrhagic stroke (HS);
a) *DICOM-i*mages and CT protocol, performed immediately after admission of a patient;
g) for patients with ischemic stroke (IS) - magnetic-resonance imaging (MRI), confirming the diagnosis.

Patients who met the following criteria were excluded from the medical history study:

a) the presence of heavy comorbidities (pneumonia, oncology, trauma, and other);

b) in the case of fatal outcome of the lack of information about the cause of death;

c) presence of only the decimated series of *DICOM*-images in *PACS* archive of medical institution.

For the comparison group criteria for inclusion were the absence of clinical and radiological signs of stroke, traumatic and other lesions of the brain, but also the presence of *the DICOM*-image, obtained in the result of native CT.

When creating a collection we used earlier developed platform for the collection and analysis of data and images on the basis of hardware - software complex "Gamma-Multivoks" [9]. The functionality of the system includes the following main blocks: a block for working with clinical data, a block for working with images, a block for evaluating statistical characteristics and an export block. The block for clinical data included screen forms for entering data on each clinical case, and the input formal conclusions of CT studies (Fig. 1). The following information on each patient was included:

a) main diagnosis, duration of hospitalization and outcome of treatment;

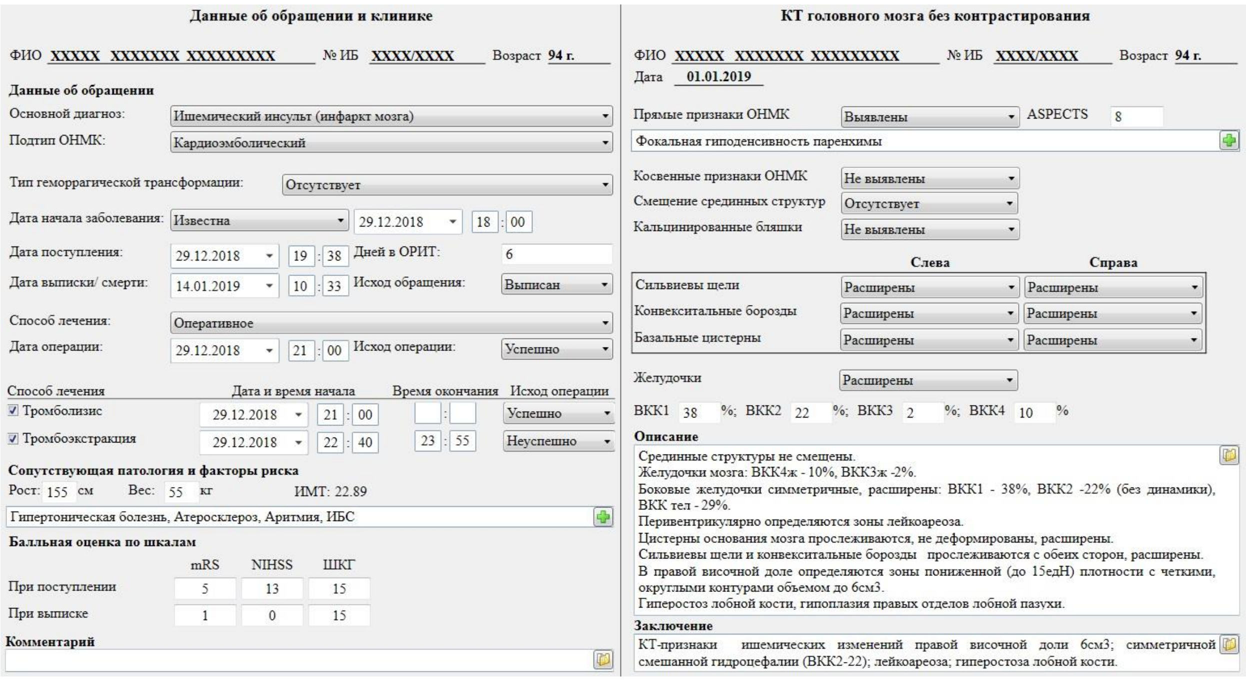
b) risk factors (arterial hypertension, atherosclerosis, receiving anticoagulants, stroke in history, compromised family history, diabetes, renal disorders, obesity);

c) data on the treatment performed (surgery and its result);

g) scoring assessment of the state upon admission and discharge from the hospital with the following scales: Modified Rankin Scale (definition of functional independence after the stroke), *NIHSS* (to assess the severity of neurologic symptoms) and Glasgow Coma Scale.

Als the platform included results of native CT and CT - angiography, performed with *Toshiba Aquilion PRIME* with slice 0.5 mm. Formal conclusions contained information about the availability of direct and indirect signs of stroke, assessment of the volume of lesions with *ASPECTS* for IS, the characteristics of cerebrospinal fluid spaces, the description of extra - and intracranial arteries, Willis circle.

The block for working with images included filtering tools, semi-automatic contouring of regions of interest and tagging. The created reference tag provided for the assignment tags to each area, describing the type of pathological mass, its location and the system of blood supply (Fig. 2).



Данные об обращении и клинике

ФИО XXXXX XXXXXXX XXXXXXXXX № ИБ XXXX/XXXX Возраст **94 г.**

Данные об обращении

Основной диагноз: Ишемический инсульт (инфаркт мозга)

Подтип ОНМК: Кардиоэмболический

Тип геморрагической трансформации: Отсутствует

Дата начала заболевания: Известна 29.12.2018 18 : 00

Дата поступления: 29.12.2018 19 : 38 Дней в ОРИТ: 6

Дата выписки/ смерти: 14.01.2019 10 : 33 Исход обращения: Выписан

Способ лечения: Оперативное

Дата операции: 29.12.2018 21 : 00 Исход операции: Успешно

Способ лечения	Дата и время начала	Время окончания	Исход операции
☑ Тромболизис	29.12.2018 21 : 00	:	Успешно
☑ Тромбоэкстракция	29.12.2018 22 : 40	23 : 55	Неуспешно

Сопутствующая патология и факторы риска

Рост: 155 см Вес: 55 кг ИМТ: 22.89

Гипертоническая болезнь, Атеросклероз, Аритмия, ИБС

Балльная оценка по шкалам

	mRS	NIHSS	ШКГ
При поступлении	5	13	15
При выписке	1	0	15

Комментарий

КТ головного мозга без контрастирования

ФИО XXXXX XXXXXXX XXXXXXXXX № ИБ XXXX/XXXX Возраст **94 г.**

Дата **01.01.2019**

Прямые признаки ОНМК Выявлены ASPECTS 8

Фокальная гиподенсивность паренхимы

Косвенные признаки ОНМК Не выявлены

Смещение срединных структур Отсутствует

Кальцинированные бляшки Не выявлены

	Слева	**Справа**
Сильвиевы щели	Расширены	Расширены
Конвекситальные борозды	Расширены	Расширены
Базальные цистерны	Расширены	Расширены

Желудочки Расширены

ВКК1 38 %; ВКК2 22 %; ВКК3 2 %; ВКК4 10 %

Описание

Срединные структуры не смещены.
Желудочки мозга: ВКК4ж - 10%, ВКК3ж -2%.
Боковые желудочки симметричные, расширены: ВКК1 - 38%, ВКК2 -22% (без динамики), ВКК тел - 29%.
Перивентрикулярно определяются зоны лейкоареоза.
Цистерны основания мозга прослеживаются, не деформированы, расширены.
Сильвиевы щели и конвекситальные борозды прослеживаются с обеих сторон, расширены.
В правой височной доле определяются зоны пониженной (до 15едН) плотности с четкими, округлыми контурами объемом до 6см3.
Гиперостоз лобной кости, гипоплазия правых отделов лобной пазухи.

Заключение

КТ-признаки ишемических изменений правой височной доли 6см3; симметричной смешанной гидроцефалии (ВКК2-22); лейкоареоза; гиперостоза лобной кости.

Fig. 1. Type of screen forms for entering clinical information and protocol of CT examination

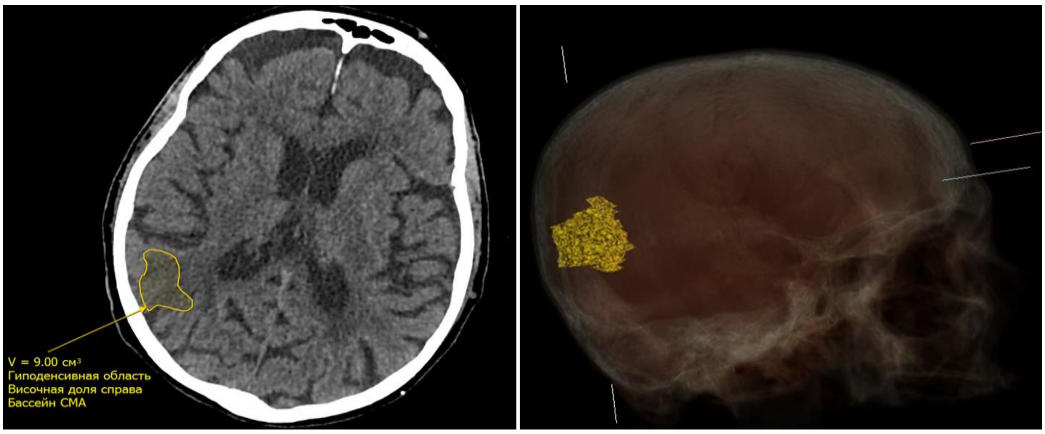


Fig. 2. Marking and tagging areas of interest on DICOM-images with 3D region of interest

All data, including *DICOM* files, have been anonymized. Storage of data has been implemented in the database *Microsoft SQL*, *DICOM*-images are stored separately in the file network storage.

Summary information for distribution of quantitative characteristics (age, *NIHSS* score*)* is represented in the form of values of the median, first and third quartile (Me [Q1; Q3]), so as the distribution characteristics were different from normal.

RESULTS

In the course of study the collection was created, which includes data on 220 patients, of whom 130 - with the IS, 40 - with HS, and also 50 people without cerebrovascular disease. A summary of the age and gender structure of each of the groups is presented in Table 1.

Table 1

The distribution of patients by age and gender

Group	Whole sample	Men	Women
Norm	*n* = 50 34 [26; 51]	*n* = 21 34 [30; 49]	*n* = 29 34 [23; 54]
Ischemic stroke	*n* = 130 69 [57; 80]	*n* = 68 65 [53; 77]	*n* = 62 77 [65; 83]
Hemorrhagic stroke	*n* = 40 58 [49; 70]	*n* = 28 57 [47; 70]	*n* = 12 67 [51; 70]

The mean score according to *NIHSS* upon admission in groups with IS and HS was 8 [3; 15] and 12 [5; 33], respectively. The mortality in the group of patients with IS was 18.6%, in the group with HS - 50%. Distribution of patients with IS on different pathogenetic subtypes according to *TOAST* was as follows: atherothromboembolic – 30, cardioembolic – 39, lacunar – 12, a stroke with other established etiology – 4, stroke other unknown etiology – 48.

Among patients with IS in 18 observations thrombolysis was performed, in 7 - trombextraction, in 5 - trombaspiration, in 2 cases - revascularization of extracranial arteries. In 17 patients with IS hemorrhagic transformation was observed (13%).

There were 330 native CT – studies totally performed: in 88 patients from IS group and in 22 patients from HS group the study was carried out twice. In all series there was markings of identified direct and indirect signs of stroke (hypodense region with ischemia, hyperdense middle cerebral artery, intracerebral and in intraventricular bleeding). Areas with cystic and glial changes, corresponding to manifestations of "old" stroke, were delineated as well. In patients with IS CT - angiography was performed in 78.5% of cases (*n* = 102).

DISCUSSION

The introduction of decision support system in the clinical practice, which significantly accelerates and increases the efficiency of the provision of medical aid during stroke, is an important task. The automatic analysis of neuroimaging data allows in the shortest possible time for carrying out an early differential diagnosis, prediction of the possible outcome of the disease and giving recommendations to the most effective method of treatment individually for each patient. The presence of a representative sample, consisting of a large number of structured and true data, is the basis for the implementation of various methods of analysis of medical images, in fact those with the use of machine learning. The efficiency and accuracy of the models directly depends also on the quality of the original data (learning sample)and requiring tons their thorough preliminary processing.

When conducting research the need to search cases arises, relevant for selected criteria, in the local storage of medical institutions with subsequent laborious process of manual markup. This is why the samples used for training systems most often contain less than 100 diagnostic series [10].

Despite the relevance of the automation of the process of diagnosis of stroke, the presence of the respective collections of images in the open access to the world, not numerous, but in Russia such projects are absent, that inevitably leads to loss of precision work models, received at the foreign population data. The most common purpose of the organization of public data sets in the main is the promotion teams in the creation and improvement of algorithms for automatic segmentation of the volume of lesions, and often collections are presented with data of diagnosing HS [6, 11], or images from the already developed, large ischemic areas (does not include cases in the most acute period), and practically do not contain clinical information about the patient. Some datasets do not include marking the foci of lesions, or marked with the help of algorithms for automatic segmentation without the participation of experts, making their use in a training sample impossible without preliminary processing. Undoubtedly, systems of emergency and early diagnosis of patients with IS are mostly relevant, but at the moment it is difficult to find data for their development. More than that, such large projects, dedicated to ischemic lesions, like *ENIGMA the Stroke of Recovery* [12], *ATLAS* [13], as well as *ISLES* [8], conducted in 2016-2018, are focused exclusively on the collection of data of magnetic resonance imafing and some types of functional studies of cerebral brain. Images of these modalities, no doubt, have a high information content for taking clinical decisions, but not are widely prevalent methods of diagnosis of stroke. Another disadvantage of universally accessible collections of images is their compressed form or pre-conversion to formats with loss of quality and metadata, which considerably limits the researchers in selecting methods of analysis.

In this article we described approach to the formation of collection MSCT images and clinical data during acute cerebrovascular events, which allows considerably expand the circle of problems, solved by the DSS on the basis of the analysis of medical images. Specificity collected the first clinical data, the presence of these neuroimaging for one patient in dynamics, counting direct and indirect signs of lesions, their clear identification of the type and localization will allow to reveal a statistically significant relation between the biomarkers of images, clinical data and the effects of treatment. Placing the collection in the open access will facilitate the involvement of specialists of different profile to participate in the creation of large-scale systematic domestic bank data, allowing to find and use a sample with a necessary set of parameters during the training and testing of algorithms for segmentation and analysis of images in such significant social diseases like stroke.

CONCLUSION

On the results of this work we created collection of multislice computed tomography and clinical data of patients with acute cerebrovascular events. The collection includes anonymized data of 130 patients with ischemic stroke, 40 patients with hemorrhagic stroke and 50 patients without cerebrovascular disease. *DICOM* – images of 330 native CT studies and 102 CT angiography studies. In 110 patients computer tomography study was performed over time. All direct and indirect signs of acute disorders of cerebral blood flow identified on images were delineated by experts, each three-dimensional region of interest are had tags, describing the type of pathological formation, its location and the system of blood supply. The collected data can be used when building a system to support the adoption of medical solutions, including methods of machine learning and analyzing biomarkers images, in addressing such important tasks, as the differential diagnosis of types of acute disorders of cerebral circulation, automatic determination of the volume of the area of lesions, evaluation risks of hemorrhagic transformation, prognosis of the outcome of a clinical case and the degree of neurological deficit. In the future, it is planned to publish the collected collection in the public access.

REFERENCES

1. Kornienko VN, Pronina IN. (eds.) *Diagnosticheskaya neyroradiologiya*. Moscow: Izdatel'stvo:T. M. Andreeva Publ.; 2006. (In Russ.)
2. Suslina ZA, Piradov MA. (eds.) *Insul't: diagnostika, lechenie, profilaktika*. Moscow: MEDpress–inform Publ.; 2008. (In Russ.)
3. Piradov MA, Krylov VV, Belkin AA, Petrikov SS. Insul'ty. In: Gel'fanda BR, Zabolotskikh IB (eds.) *Intensivnaya terapiya*. 2nd ed., rev. and exp. Moscow: GEOTAR-Media Publ.; 2017. Ch. 2. pp. 288–309. (In Russ.)
4. Suslina ZA, Varakin YuYa. Clinico-epidemiological studies a perspective direction for investigation of cerebrovascular disorders. *Annals of Clinical and Experimental Neurology*. 2009;3(3):4–11. (In Russ.)
5. Vereshchagin NV, Piradov MA, Suslina ZA. Printsipy diagnostiki i lecheniya bol'nykh v ostrom periode insul'ta. *Atmosfera. Nervnye bolezni.* 2002;(1):8–14. (In Russ.)
6. Chilamkurthy S, Ghosh R, Tanamala S, Biviji M, Campeau NG, Venugopal VK, et al. *Development and Validation of Deep Learning Algorithms for Detection of Critical Findings in Head CT Scans*. arXiv:1803.05854 [cs.CV]. 2018. Available at: https://arxiv.org/abs/1803.05854 [Accessed 21 May 2020].
7. Hajimani E, Ruano MG, Ruano AE. An intelligent support system for automatic detection of cerebral vascular accidents from brain CT images. *Comput Methods Programs Biomed*. 2017;146:109-123. PMID: 28688480 https://doi.org/10.1016/j.cmpb.2017.05.005
8. ISLES: Ischemic Stroke Lesion Segmentation Challenge. Available at: http://www.isles-challenge.org/ [Accessed 21 May 2020].
9. Dolotova D, Donitova V, Arhipov I, Sharifullin F, Zagriazkina T, Kobrinskii B, et al. A Platform for Collection and Analysis of Image Data on Stroke. *Stud Health Technol Inform*. 2019;262:312–315. PMID: 31349330 https://doi.org/10.3233/SHTI190081
10. Mikhail P, Le MGD, Mair G. Computational Image Analysis of Nonenhanced Computed Tomography for Acute Ischaemic Stroke: A Systematic Review. *J Stroke Cerebrovasc Dis*. 2020;29(5):104715. doi:10.1016/j.jstrokecerebrovasdis.2020.104715
11. Hssayeni M. Computed Tomography Images for Intracranial Hemorrhage Detection and Segmentation (version 1.3.1). *PhysioNet*. 2020. https://doi.org/10.13026/4nae-zg36.
12. *Enigma stroke recovery*. Available at: http://enigma.ini.usc.edu/ongoing/enigma-stroke-recovery/ [Accessed 21 May 2020].
13. Liew S, Anglin J, Banks N, et al. A large, open source dataset of stroke anatomical brain images and manual lesion segmentations. *Sci Data*. 2018;5:180011. https://doi.org/10.1038/sdata.2018.11

Received on 19.11.2019

Accepted on 24.03.2020