N.P. BEHTEREVA INSTITUTE OF THE HUMAN BRAIN OF THE RUSSIAN ACADEMY OF SCIENCES

As manuscript

POLINA LEONIDOVNA ANDROPOVA

DIAGNOSIS OF ISCHEMIC MIDDLE CEREBRAL ARTERY STROKE BY COMPUTED TOMOGRAPHY USING AUTOMATED IMAGE ANALYSIS SYSTEMS

Scientific specialty 3.1.25. Diagnostic radiology

THESIS WORK
For a Candidate Degree
in Medical Science
Translation from Russian

Scientific supervisor:
Corresponding Member of the Russian Academy of Sciences,
Doctor of Medical Science, Professor
Tatyana Nikolaevna Trofimova

St. Petersburg – 2023

CONTENT

INTRODUCTION

Research Rationale

According to experts from the World Health Organization (WHO), ischemic stroke is an extremely important health and social problem because it is responsible for a high proportion of morbidity and mortality in the population, as well as a significant proportion of temporary work loss and primary disability [62]. According to the Federal State Statistics Service of the Russian Federation, cerebrovascular diseases occupy the second place in the list of causes of death from cardiovascular diseases (39%), and the share of Cerebrovascular Accidents (CVA) is 21.4%. In the acute phase of ischemic stroke, mortality reaches 35%, and in the first year after occurrence, 50% of patients die. In the period from 2008 to 2016, stroke mortality in the Russian Federation has decreased by 45% and amounted to 123 cases per 100 thousand inhabitants, in 2019 this indicator reached extremely low values – 88.2 cases per 100 thousand inhabitants. However, in the wake of the new coronavirus epidemic, this indicator for 2020 has increased to 92.4 cases per 100 thousand inhabitants and continues to grow further [12, 13].

To date, neuroimaging has taken a leading position in the diagnosis of CVA. According to clinical recommendations adopted in the Russian Federation, all patients with suspected ischemic stroke are recommended to undergo urgent noncontrast CT or MRI of the brain. The results of the examination should be available within 40 minutes of admission to the in-patient facility in order to make a differential diagnosis of the CVA form and determine treatment tactics [12, 23]. Taking into account a number of reasons (time of the examination, absence of absolute contraindications, availability), computed tomography (CT) is the leading method of neuroimaging in the diagnosis of CVA, which should first answer this main question – is it the ischemic brain injury and/or intracerebral hemorrhage [12].

Detecting CT signs of ischemia in the early stages is a complex clinical issue. To provide a uniform approach to the diagnosis of ischemic stroke, the Alberta stroke program early CT score (ASPECTS), a semiquantitative scale that assesses the prevalence of early ischemic changes in the middle cerebral artery (MCA) system according to a 10-point scale, was developed in 2000. A score of ≤7 points indicate a more pronounced volume of brain tissue injury in the MCA system and correlates with a worse functional outcome and a higher risk of hemorrhagic conversion [3, 28].

The ASPECTS has been positively evaluated by many researchers as a reliable diagnostic method. However, the use of this 10-point scale has a number of limitations, including the lack of standardization, which leads to high variability in the assessment of ischemic changes by experts using ASPECTS, which may influence the further course of the patient's treatment [3, 75].

To partially solve the problem of subjectivity in the application of the ASPECTS, it is proposed to introduce automated CT analysis systems as a method to support medical decision making. To date, these systems are being developed for analysis of noncontrast CT images, CT angiography, and CT perfusion. Their application aims to automatically determine the score on the ASPECTS, quantify the stroke core, penumbra, and collateral blood flow status, and localize arterial occlusions [3, 5, 113].

The use of artificial intelligence (AI) algorithms means more effective detection of ischemic changes and a reduction in variability between experts when evaluating CT images of patients requiring emergency medical care. Developing the ability to combine collective thinking and predictions from automated CT analysis systems can have a profound impact on the organization and management of medical care [49, 113].

Therefore, it is important to validate the developed artificial intelligence algorithms and apply them in clinical practice to support medical decision making and standardize the interpretation of data from CT, which could improve patient treatment tactics and functional outcomes of ischemic stroke.

Extent of Previous Research

The time-critical nature of stroke care (Time=Brain) requires accurate and rapid tools for stroke diagnosis [12, 63]. In recent years, one of the trends in the development of clinical medical practice and the research topic of the international scientific community is the development and implementation of algorithms for automated CT evaluation of brain images based on artificial intelligence [31]. Several automated CT systems have been presented to determine the distribution volume of ischemic changes in the MCA territory according to the ASPECTS, the size of the penumbra and stroke core according to CT perfusion, and to detect cerebral thrombosis according to CT angiography [51, 53]. Several foreign IT companies announce the introduction of commercially available automated CT and semi-automated CT software for acute stroke diagnosis into a standard workflow (Aidoc®, Apollo Medical Imaging Technology®), Brainomix®, inferVISION®, RAPID®, JLK Inspection®, Max-Q AI®, Nico.lab®, Olea Medical®, Qure.ai®, Viz.ai® and Zebra Medical Vision®).

Currently, in accordance with the approved National Strategy for the Development of Artificial Intelligence for the Period up to 2030, approved by Decree No. 490 of 10 October 2019, automated CT analysis systems are being actively developed in the Russian Federation, including for use in healthcare and, in particular, in neuroradiology [9, 21]. As an instrument for regulating this activity, the Subcommittee "Artificial Intelligence in Healthcare" was established in 2019 on the basis of the Center for Diagnostics and Telemedicine Technologies of the Moscow Health. Under the activities of this subcommittee, national standards [17, 18] and clinical recommendations for testing software based on intelligent technologies [15] are being developed in the Russian Federation. Diagnosis of ischemic stroke is one of the areas where medical decision support systems are to be introduced in the Russian healthcare system. The main task for the developers of artificial intelligence algorithms is to achieve with their systems the thresholds of efficiency defined not only for clinical recommendations [15], but also for the effectiveness of assessment comparable to that performed by young professionals such as radiologists with a short-proven experience (up to 3 years).

The use of automated CT analysis systems for CT images could reduce the number of cases of underdiagnosis and overdiagnosis and reduce variability between experts in the detection of ischemic stroke by balancing the human factor [3, 89, 113]. Despite the good prospects for their practical application, there are no publications available dealing with the independent evaluation of such systems in the CT diagnosis of ischemic stroke.

Considering the above issues, there is a need to investigate the diagnostic capabilities of artificial intelligence systems through analytical and clinical validation and to study the characteristics of the interaction between a radiologist and automated CT image analysis systems.

Research Goals

Improving the diagnosis of middle cerebral artery ischemic stroke by X-ray computed tomography using automated CT image analysis systems.

Research Tasks

To achieve this goal, the following tasks have been defined:

1. Creation of a database of anonymized computed tomography studies of patients with proven middle cerebral artery stroke and no pathologic brain changes.
2. Evaluation of indicators of diagnostic effectiveness of radiologists with different experience and expertise in emergency medicine when diagnosing the middle cerebral artery ischemic stroke, and multidisciplinary concordance of emergency neuroradiology specialists in the assessment of ischemic changes by ASPECTS.
3. Evaluation of the feasibility of automated CT image analysis systems as a method for detecting ischemic stroke in the middle cerebral artery system.

4. Development of an optimal algorithm for the use of automated CT image analysis systems in a radiologist's office as a method for detecting ischemic stroke in the middle cerebral artery system.
5. Justification of recommendations for the choice of a model of interaction between a radiologist and an automated CT image analysis system.

Scientific Novelty

The thesis work proves the dependence of indicators of diagnostic effectiveness in the detection of middle cerebral artery ischemic stroke on the years of practice of radiologists and their experience in emergency medicine.

Low indicators of reproducibility of the ASPECTS were found in specialists of regional vascular centers, regardless of professional experience.

This thesis work has demonstrated the importance of choosing a model for the use of automated CT image analysis systems in the joint evaluation of ischemic changes by radiologists.

It is proven that, despite the low accuracy rates (less than 0.8) according to clinical recommendations [15], the automated CT image analysis system contributes to an increase in the diagnostic effectiveness of radiologists in joint evaluation.

In young specialists with up to three years of experience, a positive correlation was found with the introduction of an automated CT image analysis system to reduce variability in the assessment of ischemic changes in the middle cerebral artery territory with the ASPECTS.

Research Theoretical and Practical Relevance

The thesis work confirmed the direct dependence of diagnostic effectiveness in detecting ischemic middle cerebral artery stroke by computed tomography on the experience and expertise of radiologists in emergency medicine. We found that there was little agreement among experts in the assessment of ischemic changes according to ASPECTS by radiologists with different years of experience who specialized in the diagnosis of ischemic stroke. It is advocated that an automated CT image analysis system should be considered as a second opinion for radiologists with up to three years of experience. Based on the study, we developed practical recommendations for the choice of a model of interaction between a radiologist and an automated CT image analysis system for the joint evaluation of ischemic changes in the middle cerebral artery territory by computed tomography.

Research Materials and Methods

The thesis research was conducted in several stages. In the first stage, we performed a detailed analysis of publications on this problem. The analysis of publications is based on 118 sources, of which 23 are domestic and 95 are foreign.

In the second stage, a database with the results of the CT scans of 150 patients with the clinical picture of stroke the middle cerebral artery was created and registered (Certificate of Database Registration RU 2022620850). The database included 100 patients with ischemic middle cerebral artery stroke confirmed by CT angiography and CT perfusion and 50 patients excluded with ischemic stroke on the basis of dynamic CT observation and CT angiography. Two collections of native CT images were formed for further testing of radiologists and automated CT image analysis systems. In addition, three options of collections were compiled for further research stages: 1A and 2A for the third stage, 1B, 1C, 2B and 2C for the fifth stage of the research (Figure 1).

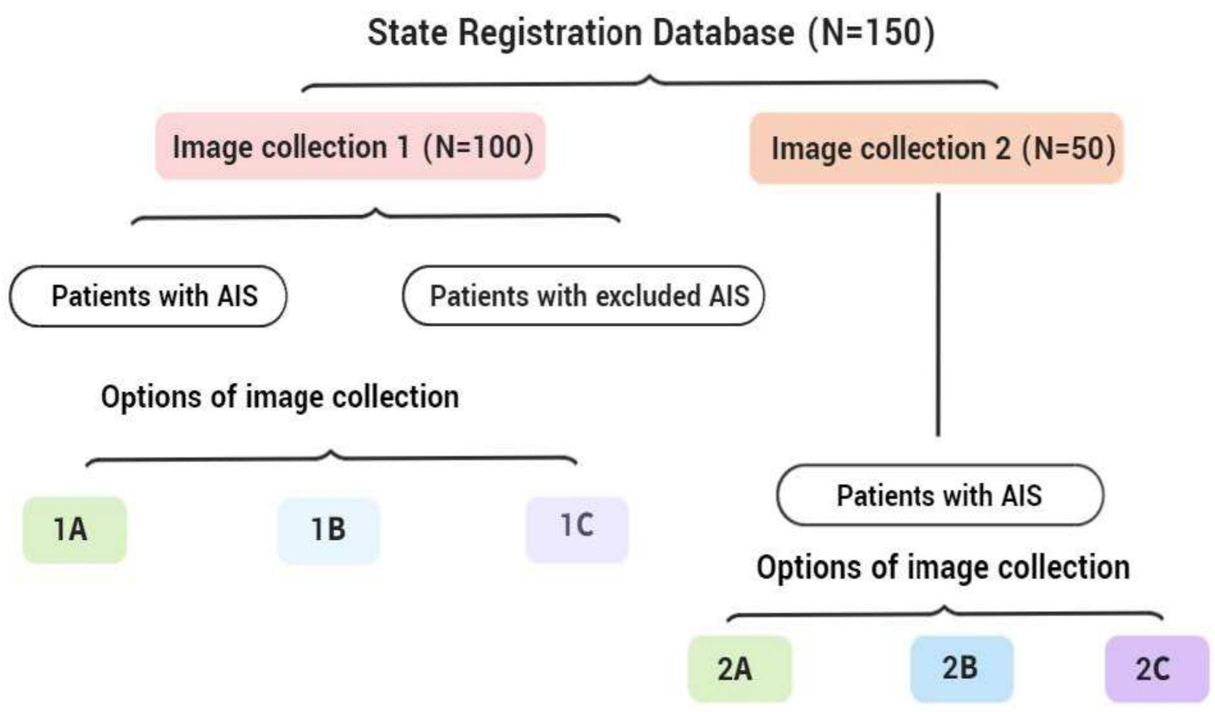

Figure 1 – Schematic representation of the second stage of the research with developing collections of images (1A, 2A for the third stage of the research, 1B, 1C, 2B and 2C for the fifth stage of the research)

In the third stage of this research, radiologists with different experience in stroke diagnosis were tested to determine their strength in determining of middle cerebral artery stroke by computed tomography and to assess inter-rater reliability with respect to CT signs of ischemic stroke (hyperdense middle cerebral artery (HAS), loss of gray matter and white matter differentiation, sulcal effacement, and decrease in CT density of brain matter). The degree of inter-rater reliability was also evident when determining the prevalence of ischemic changes according to ASPECTS by specialists with different professional experience and specialization in emergency neurology (Figure 2).

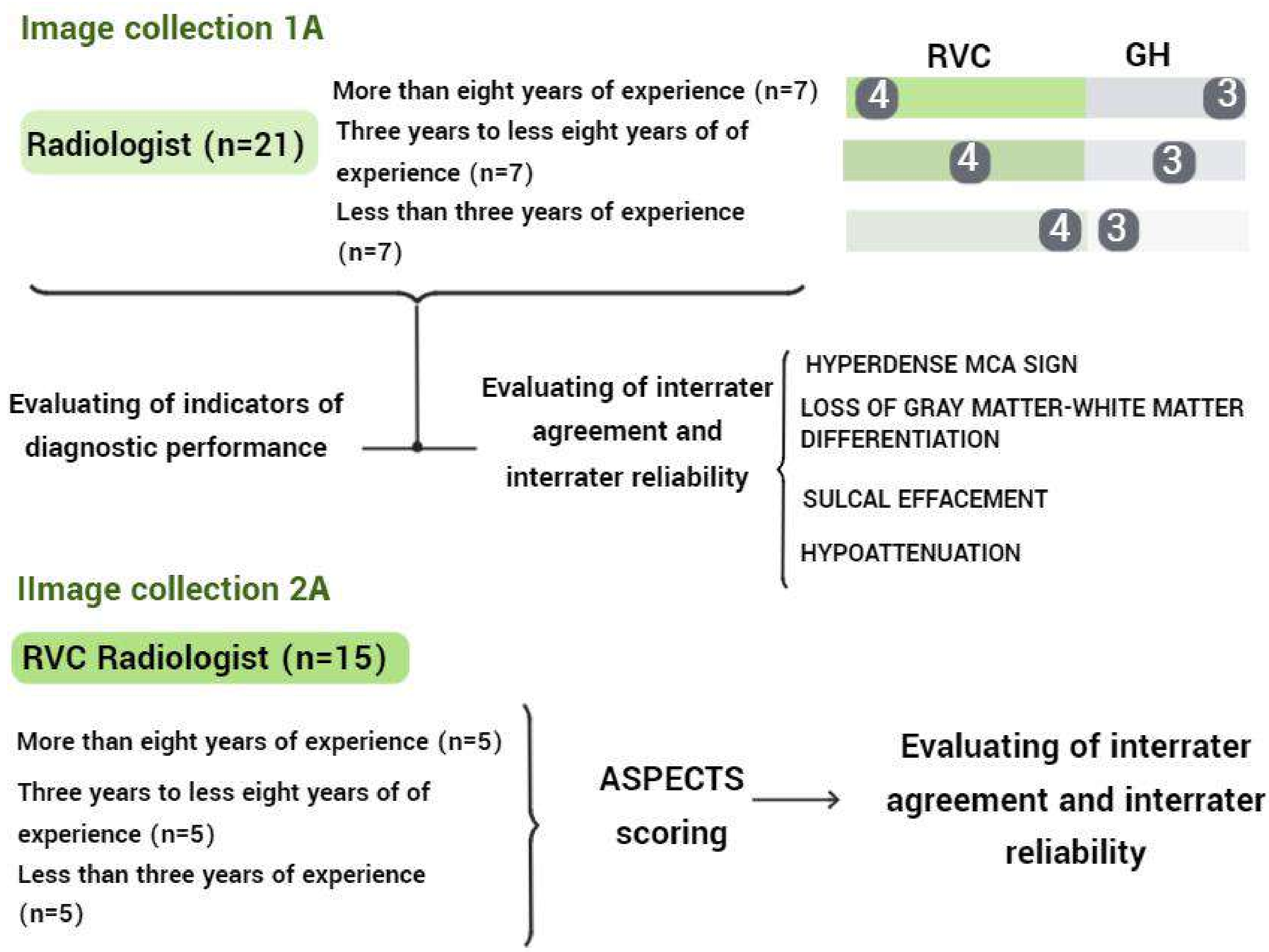

Figure 2 – Schematic representation of the third phase of the research performed on 23/05-03/06/2022, with images from collections 1A and 2A, with tests performed by radiologists from regional vascular centers (RVC) and from general hospitals (GH), who have different years of experience

In the fourth stage, the automated CT image analytical systems were tested with the analytical validation method and the indicators of their diagnostic accuracy were determined. After the program was selected for further research, it was tested with a collections of CT images from patients with confirmed stroke in the middle cerebral artery system to determine the volume of ischemic changes by ASPECTS (Figure 3).

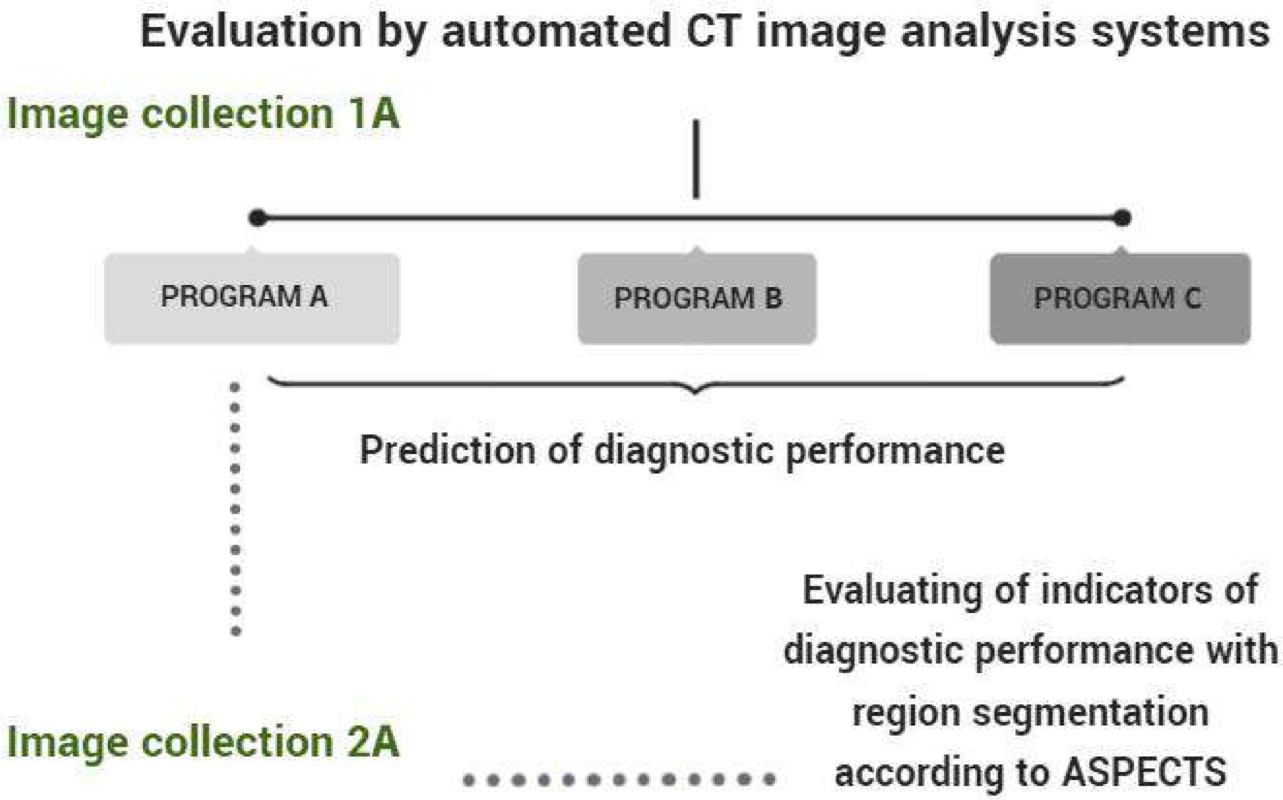

Figure 3 – Schematic representation of the fourth stage of the research with testing of anonymized automated CT image analysis systems (artificial intelligence software A, B, C)

The fifth stage explored possible options for implementing automated image analysis systems of computed tomography images as a tool to support medical decision making by radiologists with less than 3 years of experience by testing the modeling of two different options of joint evaluation: the first (parallel) review model and the second review model (Figure 4).

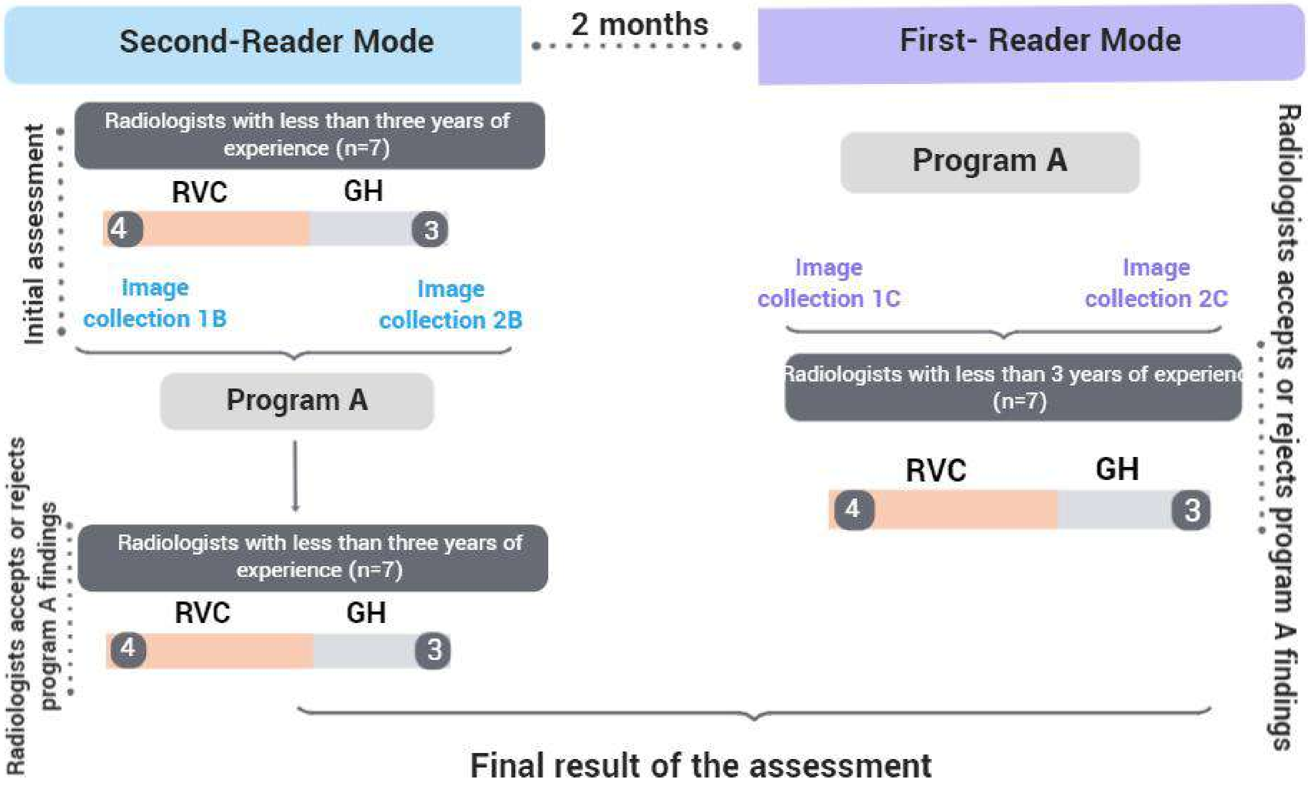

Figure 4 – Schematic representation of the fifth stage of the research performed on the images of collections 1B and 2B on 05/03-05/05/2023 and on the images of collections 1C and 2C on 07/01-07/04/2023

Statistical analysis was performed with the Statistical Package for the Social Sciences (SPSS) program, version 19.0 (SPSS Inc, Chicago, IL, USA). Strength in detecting ischemic middle cerebral artery stroke (true-positive, false-positive, false-negative, true-negative results, sensitivity, specificity, accuracy) and inter-rater reliability (kappa statistics: Cohen's kappa and Fleiss's kappa) were assessed. The ROC analysis was performed and the area under the ROC curve was calculated.

Provisions to Be Defended

1. The ASPECTS has low reproducibility rates in the assessment of ischemic changes by radiologists, regardless of their years of work or experience in emergency medicine.

2. Automatic analysis of computed tomograms with specialized software allows for greater diagnostic strength in radiologists with less than 3 years of experience in a first- reader mode.

3. Automatic analysis of computed tomograms with specialized software allows higher reliability in radiologists with less than three years of experience, in the evaluation of ischemic changes in the middle cerebral artery territory with ASPECTS using a first- reader mode.

Degree of Credibility and Evaluation of Results

The degree of credibility of the research results is determined by a sufficient number of clinical observations, a representative sample size, the use of modern examination methods by radiologists with different years of work and professional experience (detection of signs of stroke in the middle cerebral artery and evaluation of ischemic changes by ASPECTS), testing of automated analysis systems and joint testing of radiologists with similar years of work and automated analysis systems, and processing of the obtained data by appropriate methods of mathematical statistics.

Dissertation Materials Presented at Conferences

The main results obtained as part of the work on the thesis were presented at the VIII Congress of the National Association of Phthisiatricians (St. Petersburg, 25-27 November 2019), XII International Congress *Nevsky Radiological Forum – 2021* (St. Petersburg, 7-10 April 2021), All-Russian National Congress of Radiation Diagnosticians and Therapists *Radiology – 2021* (Moscow, 25-27 May 2021), European Congress of Radiology (Vienna, 2-6 March 2022), European Congress of Radiology (Vienna, 13-17 July 2022), Joint Workshop *Machine Learning Methods and Statistical Models in Medicine* St. Petersburg State University (SPSU) – Huazhong University of Science and Technology (HUST) (St. Petersburg, 29 September 2022), XIV International Congress *Nevsky Radiological Forum – 2021* (St. Petersburg, 7-8 April 2023), XVII All-Russian National Congress of Radiation Diagnosticians and Therapists *Radiology – 2023* (Moscow, 30 May – 1 June 2023), Conference *Computational*

Biology and Artificial Intelligence for Personalized Medicine – 2023 (online, 9-11 August 2023), VI Congress of the National Society of Neuroradiologists (Sochi, 29-30 September 2023).

Research Practical Implementation

The research results and developments are implemented in the practical work of the X-ray Department of the St. Petersburg State Medical Institution Elizavetinskaya Hospital, the X-ray Department of the Ochapovsky Regional Hospital No. 1, the X-ray department of the Samara Regional Clinical Hospital named after V.D. Seredavin, in the comprehensive medical clinic Scandinavia for children and adults, in St. Petersburg. The theoretical and practical results of the thesis are used in the educational activities of the N.P. Behtereva Institute of the Human Brain, St. Petersburg Research Institute of Phthisiopulmonology and Kuban State Medical University.

Publications

Six publications on the thesis topic were published, including three papers in publications recommended by the Higher Attestation Commission of the Russian Ministry of Education and Science for publication of the results of thesis works, two papers in publications associated with Scopus, and one database was registered (Registration Certificate No. 2022620850). The scientific publications adequately reflect the content of the thesis work and the author’s abstract.

Author’s Personal Contribution in Results Generation

The topic and plan of the thesis, its main ideas and content were developed together with the supervisor on the basis of a comprehensive study of the relevant publications. The author independently worked out and justified the relevance of the topic of the thesis, the goals, tasks and stages of scientific research. The thesis defender personally studied the publications and developed a research methodology with the supervisor. She contributed to the development of a database of radiographic images used for analytical validation of automated analysis programs and testing of radiologists.

The author personally performed all studies included in the database, as well as tests of all selected programs for image analysis with subsequent interpretation of the obtained results, collected and analyzed complex computed tomography data. The author's personal contribution to the study of the publications, collection, generalization, analysis, statistical processing of the obtained data and writing of the dissertation is 100%. The author personally wrote the text of the thesis work.

Correspondence of Thesis Work with Specialization Certificate

The work corresponds with the specialization certificate 3.1.25 "Diagnostic Radiology"

i.1 Diagnostics and monitoring of physiological and pathological conditions, diseases, injuries and malformations (including antenatal) by evaluating the qualitative and quantitative parameters obtained using the methods of diagnostic radiology.

i.11 The use of digital technologies, artificial intelligence and neural networks for the diagnostic and monitoring of physiological and pathological conditions, diseases, injuries and malformations (including antenatal) using the methods of diagnostic radiology.

Scope and Structure

The thesis works is designed to be 132 pages, Times New Roman, font size 14. The paper consists of the following parts: Introduction, overview of reference sources, description of research methods and materials, chapters with own research, conclusions, practical recommendations, references. The paper contains 28 tables and 18 illustrations (Figures).

CHAPTER 1. POSSIBILITIES AND PROSPECTS OF CLINICAL IMAGING OF ISCHEMIC STROKE (REVIEW OF PUBLICATIONS)

1.1. Computed tomography in the diagnosis of ischemic middle cerebral artery stroke

According to estimates from the World Health Organization (WHO), stroke is the second leading cause of death in the world. An estimated 9.6 million strokes are predicted annually, with incidence increasing as the population ages, with ischemic stroke accounting for 85% of cases [62].

In the Russian Federation, from 2008 to 2016, mortality from stroke decreased by 45% and amounted to 123 cases per 100 thousand population; in 2019, this figure reached extremely low results - 88.2 cases per 100 thousand population. However, against the background of the epidemic of a coronavirus disease (COVID-19), this figure for 2020 increased to 92.4 cases per 100 thousand population and continues to grow [12].

Mortality rates in patients with stroke largely depend on the conditions of treatment in the acute period. The early 30-day mortality rate due to stroke reaches 35%. At the same time, about 24% of patients die in hospitals, 43% die at home, and 50% of patients die by the first year from the development of the disease [1, 23].

Along with high mortality, the socially significant consequences of stroke include disability in surviving patients after a primary stroke, as well as an increased risk of developing recurrent strokes. According to the National Register, approximately 60% of patients who have suffered a stroke remain disabled and are able to care for themselves, 19-35% need outside help to care for themselves and become dependent on others, and only 15-20% remain active. According to the epidemiological study of stroke using the territorial-population registry method, the ratio of ischemic to hemorrhagic strokes was 5:1. The average age of stroke onset is 66.7 years (63.7 years in men and 69.4 years in women). The absolute number of strokes in patients under 67 years of age is higher in men, and at older ages it is higher in women [23].

It is necessary to note the medical and economic significance of the problem of stroke. According to WHO, the cost of treatment and rehabilitation of one patient with stroke is 55,000-73,000 dollars/year, which indicates the enormous economic damage from this disease [63].

Analysis of the above data shows that the treatment of patients with cerebrovascular pathology is one of the most pressing and complex problems in medicine and social care, due to the rapidly growing incidence, as well as extremely severe consequences leading to high disability.

The most commonly affected intracerebral artery in ischemic stroke is the middle cerebral artery (MCA). It branches directly from the internal carotid artery and consists of four main segments: M1, M2, M3 and M4. These vessels supply blood to most of the frontal, temporal and parietal lobes of the brain, as well as (lenticulostriate arteries) deeper structures of the hemispheres, including the basal ganglia [14, 85], which causes a worse functional outcome and a higher risk of mortality if inadequately treated [4, 85]. That is why improving the diagnosis of ischemic stroke in the middle cerebral artery is of such high importance.

In the diagnosis of stroke, the leading place belongs to such imaging methods as CT and MRI [22, 60, 86, 117]. CT is a highly informative method with which it is possible to identify stroke and its nature (ischemic, hemorrhagic, mixed type) [55]. The method makes it possible to determine the location, size (volume) and nature of the lesion with high accuracy, which in the study of certain types of cerebral circulatory disorders reaches 90-100% [36]. Following CT scan makes it possible to trace the evolution of pathological changes over time.

According to clinical guidelines [12], all patients with signs of stroke are urgently recommended to undergo a non-contrast CT or MRI of the brain with the results of the study (conclusion) obtained within 40 minutes from the patient's admission to the hospital for differential diagnosis of stroke in order to determine treatment tactics [4, 10, 56, 103, 104, 110, 118]. Reducing the time interval from the moment of admission to the hospital to the start of a neuroimaging study of the brain helps to reduce the time before the start of therapy and, accordingly, a better clinical outcome of the disease.

In accordance with Decrees of the Government of the Russian Federation of December 27, 2007 No. 1012 and No. 186 of March 2, 2009, the implementation of measures aimed at improving medical care for patients with vascular diseases has begun in the Russian Federation. As part of regional vascular centers and primary vascular departments, units have been created for the treatment of patients with acute stroke, with round-the-clock radiation diagnostic units (CT and/or MRI rooms), as well as a department (office) for x-ray surgical diagnostic methods, treatment and an operating room for x-ray endovascular diagnostic and treatment methods. One of the unforeseen consequences of this initiative was the hospitalization of almost all patients with impaired consciousness of various etiologies with a diagnosis of stroke in the vascular center by emergency medical personnel. The consequence of this approach of emergency medical care was the massive admission of patients to primary vascular departments and regional vascular centers with suspected stroke, causing an increase in the load on specialists involved in the emergency diagnosis of stroke at the inpatient stage (neurologists and radiologists), as well as an increase in the use of serious material resources [95]. Also, patients with sudden onset stroke who are in hospitals providing medical care for the underlying disease undergo a CT scan to exclude other etiologies of the clinical presentation and to justify transfer to a vascular center.

In the absence of division of radiologists by specialization, following the example of foreign medical practice, each radiologist specialized on computed tomography in the Russian Federation must demonstrate high level diagnostic efficiency in detecting ischemic stroke, which is impossible without certain competencies, experience and practical skills in working with the specified patient population.

Early CT diagnosis of stroke makes it possible to identify and evaluate ischemic changes, which, in turn, determines treatment tactics and prognosis of complications [4, 80].

There are a number of CT signs that indicate the presence of ischemic stroke: a sign of acute arterial occlusion (hyperdensity of the middle cerebral artery sign), signs of early ischemic changes (loss gray/white matter differentiation, sulcal effacement on the affected side) and hypoattenuation [4, 98].

Hyperdensity of the middle cerebral artery, as a highly specific sign (90-100%) of acute ischemic stroke in the same area, was first described by Gacs G. et al. in 1983 [47]. MCA hyperdensity sign is manifested by an increase in CT density of the proximal MCA in a fairly wide range from +40 HU to +80 HU, and is associated with thrombosis of the MCA M1 (and often M2) segment. The same mechanism is detected in the distal parts of the MCA with the formation of a pattern of pointwise increase in MCA density ("point sign") [4, 68, 114].

MCA hyperdensity sign is often the only diagnostic CT sign at the early stage of ischemic stroke [4].

However, we should not forget that despite its high specificity, MCA hyperdensity sign can occur with increased hematocrit value (in contrast to this sign, an increase in CT density will be observed in all intracranial arteries and veins), with viral lesions [4, 76], There are also known references to MCA hyperdensity sign during artery dissection [4, 66]. A decrease in the density of the brain parenchyma surrounding the artery (infectious/neoplastic lesion) can also create a picture of a pseudohyperdense MCA sign [4, 74].

Koo C.K. et al. [76] established objective criteria for MCA hyperdensity by measuring the absolute density of affected and unaffected vessels: the absolute density should be at least +43 HU with a ratio with the density of other arteries <1.2 [4].

Edema is a common response to various forms of brain injury. In ischemic stroke, there are three types of edema: cytotoxic, ionic and vasogenic. Each of these types has its own dominant imaging pattern, the identification of which, in combination with additional imaging data and clinical history, often provides clues to the correct interpretation of changes and influences further treatment [4, 39].

Cytotoxic edema results from disruption of the adenosine triphosphate (ATP)-dependent transmembrane sodium-potassium and calcium pumps, and is usually caused by cerebral ischemia or excitotoxic (secondary to excessive neurotransmitter stimulation) brain injury. This leads to intracellular fluid accumulation in neurons and neuroglia. The gray matter is primarily affected due to its high metabolic activity and higher density of astrocytes [4, 98]. At this stage of development of the ischemic process, CT is not a sensitive method for detecting it, and the visualization pattern can only be detected using MRI [4].

Further, after all compensatory mechanisms have been exhausted, water begins to flow along the osmotic gradient into the depleted intercellular space, but without damaging the blood-brain barrier, simulating ionic edema. This process ultimately involves both gray and white matter with a corresponding loss of their differentiation on CT [4, 39].

Loss grey -white differentiation, first of all, includes the "insular ribbon sign" (loss of definition of the gray-white interface in the lateral margin of the insular cortex) and the disappearance of the boundaries (contrast) of the basal ganglia, and subsequently, a change in the differentiation of the cerebral cortex [74]. Early involvement of the basal nuclear region in the pathological process during thrombosis of the proximal MCA is due to the fact that this anatomical region is supplied by perforating (lateral lenticulostriate) arteries that branch directly from the M1 segment of the MCA [4].

Severe or recurrent damage overloads transmembrane ion pumps, causing cell death with disruption of the blood-brain barrier, resulting in vasogenic edema. Late complications include neuronal apoptosis, atrophy and gliosis [4].

Vasogenic edema is modeled by disruption of the endothelial tight junctions that form the blood-brain barrier, secondary to either physical disruption or release of vasoactive mediators. As a result, intravascular proteins and fluid enter the extracellular space [4, 39]. An increase in the fluid content in the intercellular space by 1% results in a CT attenuation decrease of 2.5 HU (Hounsfield Units). Given that vasogenic edema leads to disruption of the blood-brain barrier, increased vascular permeability can lead to hemorrhagic stroke transformation and extravasation of contrast agent after intravenous thrombolytic therapy and/or endovascular intervention. And the larger the area of change, the higher this probability. It was for the purpose of assessing the risk of such complications that the ASPECTS was introduced (is a 10-point quantitative topographic CT scan score used for middle cerebral artery (MCA) stroke patients), with the help of which, based on all the above-described signs, it is possible to predict the functional outcome of a stroke, as well as the results thrombolytic therapy and thrombus extraction [4, 57].

The sulcal effacement is described as a consequence of mass effect, which may indicate irreversible damage to the brain parenchyma [4, 36]. However, Haussen D.C. et al. in their study confirmed that despite the generally accepted interpretation of the meaning of this sign, isolated sulcal effacement without an associated decrease in the density of the brain substance, with preservation of the differentiation of the cortex, cannot be interpreted as a sign of ischemic changes, and also cannot be taken into account when assessing according to ASPECTS, since it is a potentially reversible phenomenon, due to increased leptomeningeal collateral blood flow [4, 105]. Taking into account the lack of a unified approach to defining this sign, according to foreign literature, it can cause variability in the results of assessing CT signs of ischemic stroke [4].

In the article by Wardlaw J. M. et al. [116] conducted a systematic review of publications issued between 1990 and 2003 containing studies assessing inter-rater agreement in identifying signs of early ischemic changes (including follow-up ASPECTS scores), and determining the correlation between diagnostic test results, clinical outcome, as well as treatment (thrombolytic therapy). Across 15 studies, inter-rater agreement (an average of 30 CT studies and six experts) for the presence of AIC ranged between 0.14 and 0.78 for any AIC (kappa statistic). The mean sensitivity and specificity for detecting early ischemic changes were 66% (range 20–87%) and 87% (range 56–100%), respectively. In studies that included neuroradiologists with more experience and expert qualifications, the accuracy of detecting AIC increased. In articles where they tried to increase diagnostic efficiency due to homogeneous and general theoretical training of experts, the results did not improve after the experiment. The authors concluded that further research is needed to determine the most reliable tools that can help detect AIC, which is likely to influence thrombolysis decisions and clinical outcome [4, 116].

In a study by Dippel D. W. J. et al. [108] it was suggested that the level of correct verification of CT signs of ischemic stroke is an important indicator of its outcome and the effect of thrombolytic therapy. The authors studied these signs prospectively in 260 patients with ischemic stroke in the anterior circulation. The presence of signs of AIC was confirmed by dynamic observation of patients, in the form of CT examinations 1 and 12 weeks after the patient's admission. The reviewers were experienced clinicians specializing in the treatment and diagnosis of stroke (neurologists, neuroradiologists and endovascular surgeons). The mean time between stroke onset and initial CT examination was 3.2 hours. In more than half of the patients (52%), a control CT scan revealed an extensive infarction in the middle cerebral artery territory. The interobserver odds ratio (chi statistic) for any early sign of stroke was 0.27 (95% confidence interval: 0.15 to 0.39). Agreement regarding the prevalence of ischemic changes in the territory of the middle cerebral artery was insignificant: 0.37 and 0.35, respectively [4, 108].

In a study by Grotta J.C. et al. [26] found a level of agreement between neuroradiologists with extensive experience in identifying stroke and other clinicians regarding the presence of signs of AIC. Seventy initial CT examinations were evaluated by 16 experts, including neurologists from regional vascular centers (specializing in the diagnosis and treatment of stroke), neurologists without stroke's diagnosis experience, other emergency medicine physicians, and emergency medicine neuroradiologists. Kappa values for interrater agreement (95% confidence interval [CI]) ranged from 0.20 (-0.20 - 0.61) (weak agreement) to 0.41 (0.37 - 0.45) (moderate agreement). The authors concluded that even among experienced clinicians, there are low levels of agreement regarding the recognition and quantification of signs of early ischemic changes [4, 26].

Based on foreign literature data, recognizing signs of thrombosis of the anterior circulation arteries and acute ischemia is a complex diagnostic task. There are conflicting data regarding the ability of radiologists, including radiologists with little experience, to correctly interpret CT signs of ischemic stroke. In the context of the annual renewal of hospital staff by young specialists, the presence of a period of their adaptation to the specifics of the work of a neuroradiologist, combined with a shortage of radiologists for the prompt interpretation of CT studies in the constituent entities of the Russian Federation, a natural increase in the level of variability in the interpretation of radiation images is predicted between doctors. And, as a consequence, a decrease in the quality of medical care [4].

1.2. The Alberta Stroke Program Early CT score: issues of inter-rater reliability in clinical practice

In order to form a unified approach to the diagnosis of ischemic stroke, ASPECTS (Alberta stroke program early CT score) was developed in 2000 - a quantitative topographic CT scan score used for middle cerebral artery (MCA) stroke patients. ASPECTS involves the assessment of ten areas on CT slices at standard levels (basal ganglia and rostral structures) (Figure 5): the caudate nucleus, insula, lenticular nucleus, internal capsule, and six other cortical areas (M1-M6). Areas M1–M3 are located at the level of the basal ganglia, areas M4–M6 are at the level of the ventricles directly above the basal ganglia, the border is the head of the caudate nucleus [3, 90].

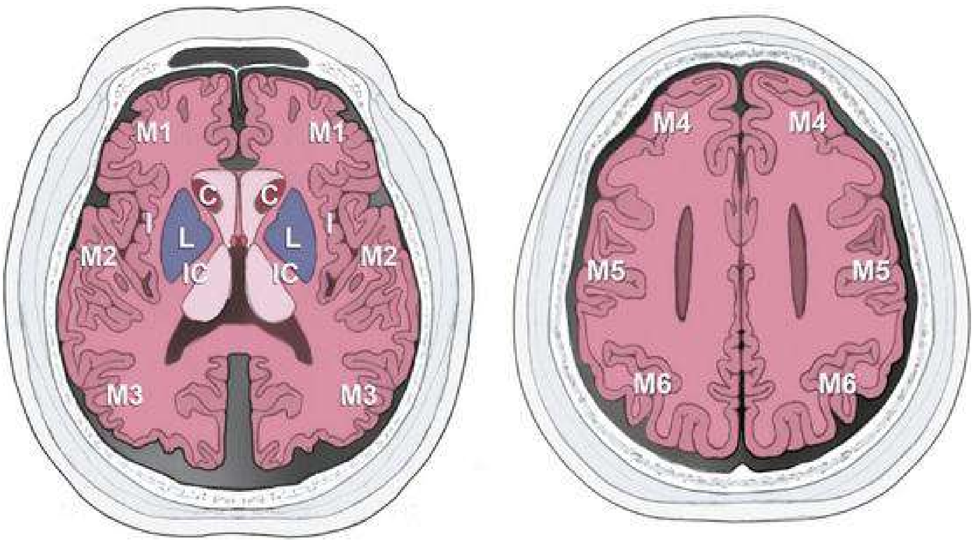

Figure 5 – Schematic representation of ASPECTS (Alberta stroke program early CT score) [96]

According to the scale, each of the ten regions is sequentially assessed. The ASPECT score is made by 1 point is deducted from the initial score of 10 for every region involved. In this case, pathological changes must be visualized on at least two consecutive sections to exclude the effect of volumetric averaging. Thus, a CT scan without pathological changes corresponds to an ASPECT score of 10. A score of 0 points indicates diffuse damage to the entire territory of the middle cerebral artery. A score of ≤7 points correlate with a worse functional outcome and a higher risk of hemorrhagic transformation of stroke [3, 57].

The ASPECTS scale has been praised by many researchers as a reliable diagnostic method. For example, Baek J. H. et al. [91] concluded that ASPECTS is a better method for predicting functional outcome in patients with acute ischemic stroke receiving thrombolytic therapy compared with other scoring systems [3, 91]. This conclusion is confirmed by other studies [111].

However, the use of the scale also has limitations [69]:

— the ASPECTS scale is used in the MCA region (although at the moment a modification of the scale for assessing ischemia in the vertebrobasilar (VB) system is also known - posterior circulation ASPECTS – pc-ASPECTS). Therefore, its use is incorrect for occlusion of the internal carotid artery, including the fetal type of structure of the posterior cerebral artery [3, 106];

— ASPECTS assessment is difficult at the level of the M2 territory due to artifacts from the bones of the skull base [3];

— the presence of subcortical and age-related periventricular changes in white matter can lead to incorrect assessment according to ASPECTS [3];

— low-quality images, for example, with motion artifacts, can also lead to erroneous ASPECTS scores [3].

Also, a significant limitation in the use of this scale is the level of thrombosis of the middle cerebral artery. Since the scale covers the entire territory of the MCA, its use is correct only in case of occlusion of the proximal parts of the artery. If the vessel is thrombosed more distally, then scoring all 10 ASPECTS areas becomes inaccurate (Figure 6).

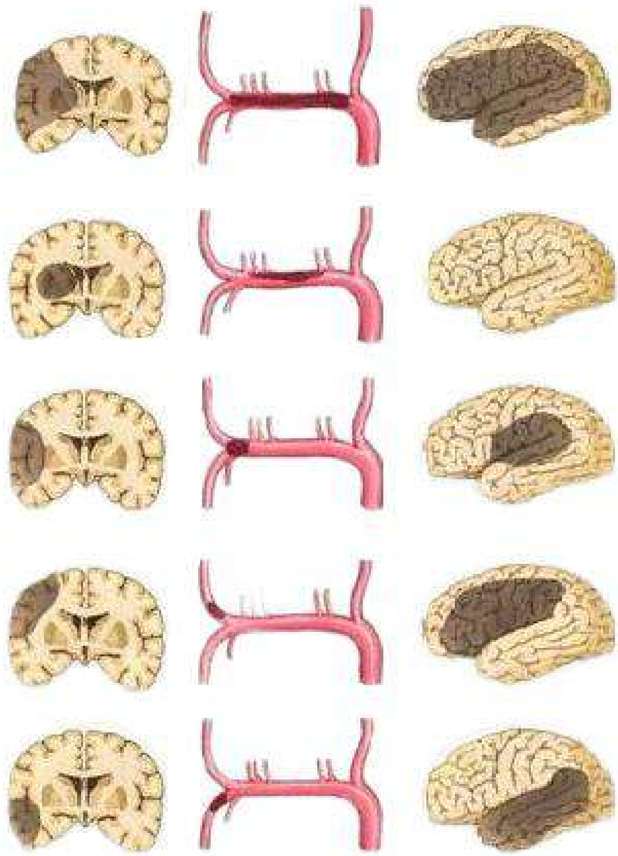

Figure 6 – Variants of localization of middle cerebral artery occlusion associated with the volume of the affected territory of middle cerebral artery

It is also important to note that the scale is not completely standardized ASPECTS areas are indicated schematically. For example, the assessment on a scale at the “watershed boundaries” causes some difficulties. The location of cortical border zones may differ due to the development of leptomeningeal collaterals. Also, in the original ASPECTS study, only two native CT slices were used for evaluation, but modern medical practice overwhelmingly uses all available scan slices [3, 115].

Another source of intrasubject variability is the characteristics of the AIC that are used to calculate the scores. Loss of grey-white differentiation, according to some authors, is associated with edema and irreversible damage. Under experimental conditions, ischemia in the form of impaired differentiation of the cortex is reversible only within a few minutes after the onset of stroke [3, 73]. In clinical settings, impaired cortical differentiation with a focal decrease in cortical density highly specifically represents irreversibly damaged brain tissue (i.e., infarct core) [3, 54]. Since, against the background of edema in these areas, a narrowing of the cerebrospinal fluid spaces occurs, the sulcal effacement is an associated sign of impaired differentiation of the cortex. However, in connection with recent pathophysiological studies, the possibility of visualization of isolated effacement of the grooves (without disruption of differentiation) has been identified, which is not considered a reliable sign of ischemia taken into account in ASPECTS, since it is a potentially reversible phenomenon [3, 50, 58].

An additional criticism of ASPECTS is that some areas - such as the internal capsule - are much smaller in volume than other areas, but the structure is equally weighted in the score distribution; thus, two patients with the same ASPECT scores may have different degrees of severity and extent of AIC [3, 101]. For example, ASPECTS was criticized in a study by Phan T. G. et al [107]. The authors noted that in the ASPECTS scoring system, the striatocapsular region (putamen, dorsolateral parts of the caudate nucleus, anterior femur of the internal capsule) has a disproportionate “weight.” Four experienced neuroradiologists assessed individual ASPECTS regions on CT images of nineteen patients with MCA stroke, comparing the final volume of ischemic changes on CT studies over time (days 5–10). The extent of ischemic changes was determined by manual segmentation on CT images. Linear regression was used to estimate the regional volume associated with each ASPECTS region. ASPECTS areas had unequal weight, with the striatocapsular area accounting for 21% of the AIC volume in the MCA territory with total damage to the blood supply territory [3, 107].

The most common criticism of this scale is that ASPECTS was introduced to reduce inter-rater variability, but studies have shown fairly heterogeneous results in the level of agreement between experts [3].

In a study by Farzin B. et al. [52] examined the level of inter- and intra-rater agreement of fifteen specialists when assessing CT examinations using the ASPECTS in thirty patients with ischemic stroke. The expert group included: six neurologists specializing in the diagnosis and treatment of patients with stroke (four with more than ten years of experience), five x-ray endovascular surgeons (three with more than ten years of experience) and four neuroradiologists (two with more than ten years of experience). Each specialist underwent two blind assessments at least three weeks apart. Additionally, the experts were provided with the following clinical data: gender, age, presence/absence of aphasia and hemiparesis, as well as the NIHSS (National Institutes of Health Stroke Scale). Interrater agreement was weak to moderate, with no significant difference between specialties (neuroradiologist, endovascular surgeon, neurologist) and experience. Interobserver reliability was determined using the weighted kappa statistic (Cohen's κ). The agreement coefficient was considered as moderate if moderate (k>0.4–0.6), significant (k>0.6–0.8) and as complete agreement (k>0.8–1.0). None of the kappa values reached a moderate level (0.6) among all experts. Even when dichotomized ASPECTS into two categories (0–5/5–10 points), inter-rater agreement did not reach a significant level (k>0.561), which means that at least 5 out of 15 experts will give a different opinion in 15% of cases. Internal (intra-rater) agreement (rater ratings on first and repeat viewings) ranged from 0.599 to 0.943. Radiologists had the same rate of agreement on repeat review in 40% of cases [52]. The researchers also reviewed articles assessing interrater agreement for the use of ASPECTS, published from 2000 to 2015. The methodologies of the studies reviewed differed on several characteristics, including whether clinicians were provided with clinical data when assessing studies, whether there was a time limit, access to all CT scans, the ability for experts to set their own window parameters. The authors made a general conclusion that these factors play a significant role in the identified subjectivity of expert assessment, which reflects the high degree of variability in inter-rater agreement [3, 52].

In a study by Kobkitsuksakul C. et al. [75] examined the level of inter- and intra-rater agreement when assessing CT examinations of forty-three patients with ischemic stroke according to the ASPECTS by two neuroradiologists (with more than ten years of experience), a fellow in the field of neuroradiology (a foreign version of one of the stages of postgraduate education) and a resident. Agreement between the two neuroradiologists and the fellow regarding the dichotomization of ASPECTS scores, as assessed by Cohen's kappa, was mostly moderate (0.486 - 0.678). Agreement between two neuroradiologists and a resident on the total ASPECTS score ranged from weak to moderate (0.198 - 0.491) [3, 75].

Coutts S. B. et al. [72] concluded that ASPECTS varied between prospective and retrospective CT scans. At higher ASPECTS (>7), the prospective examiner tended to underscore ischemic changes. With lower ASPECTS (<3), the present tense expert, on the contrary, showed a tendency to exaggerate the volume of ischemic changes by almost 2 points [3, 72]. The reasons for this likely reflect a combination of factors, including the tendency of the human visual system to overestimate the boundaries of affected areas [3, 41]. It should be noted that Coutts S. B. et al. also pointed out the influence of the human factor (the desire not to perform thrombolysis in patients with concomitant pathology and a burdened medical history), when assigning lower scores to the patient on the ASPECTS scale [3, 72].

In a study by Alexander L. D. et al. [78], which included fifty-five patients diagnosed with subacute ischemic brain injury, had CT images obtained two days or more after disease onset retrospectively assessed using ASPECTS by three neurologists. The authors noted that interrater agreement was almost complete with Cohen's k value of 0.90 [3, 78].

Van Seeters T. et al. [97] conducted a prospective study involving one hundred and five patients with acute neurological deficits and suspected acute ischemic stroke in an extended (9 hours after symptom onset) “stroke window”. All patients underwent non-contrast CT examination of the brain, CT perfusion and CT angiography of cerebral vessels upon admission. All images were assessed twice according to two parameters: the presence of AIC with their assessment on the ASPECTS, as well as using the “one third” rule of the middle cerebral artery basin (an assessment scale of the prevalence of ischemic changes dividing the MCA territory into three parts; if more than ⅓ is affected, thrombolysis is not was carried out). Four neuroradiologists assessed the CT images of these patients twice for inter- and intra-rater agreement using the kappa statistic and intraclass correlation coefficient. As a result, interrater agreement for the 1⁄3 rule of MCA and ASPECTS ranged from moderate for non-contrast CT scans, low for CT angiography images, and was complete for all CT perfusion maps. The researchers concluded that CT perfusion is a more reliable method for detecting ischemic stroke using computed tomography in terms of interrater agreement [3, 97].

Pexman J. H. et al. [111] surveyed six physicians regarding their interpretation of ASPECTS. Half of the researchers included isolated swelling of the cortex (without disruption of its differentiation) in the assessment. Despite this, interrater agreement for ASPECTS was higher than interrater agreement for the 1/3 MCA rule [3, 111].

In their study, Mc Taggart R. A. et al. [27] compared the level of inter-rater variability in the assessment of CT and MR ASPECTS in the early stages of stroke. Seventy-four patients who underwent CT, MRI, and thrombectomy within 12 hours of stroke onset were prospectively analyzed. Two experts assessed the presence and extent of early ischemic changes on non-contrast CT images using ASPECTS, as well as on diffusion-weighted images (DWI). Inter-rater agreement for CT ASPECTS and DWI ASPECTS was 0.579 and 0.867, respectively. DWI-ASPECTS correlated with functional outcome ($p=0.004$), whereas CT-ASPECTS did not ($p=0.534$). Both DWI ASPECTS and CT ASPECTS correlated with the amount of ischemia detected on DWI. Performance analysis showed that DWI-ASPECTS was superior to CT-ASPECTS in terms of the time interval between symptom onset and prediction of good functional outcome. Also, based on the results of this study, it was found that the agreement between experts when assessing using DWI-ASPECTS was higher than when assessing using CT-ASPECTS. DWI-ASPECTS assessment of RII was also superior to CT-ASPECTS in predicting functional outcome (90 days) [3, 27].

In a study by Barber P. A. et al. [115] analyzed two hundred and three CT studies to establish the relationship between the quantitative assessment of scale scores on CT images and the prediction of the outcome of stroke during thrombolytic therapy. Baseline ASPECTS was inversely correlated with NIHSS stroke severity ($r=-0.56$, $p<0.001$). ASPECTS predicted functional outcome and the occurrence of hemorrhagic transformation of stroke ($p < 0.001$, $p = 0.012$, respectively). The sensitivity of ASPECTS for functional outcome was 0.78 and specificity was 0.96. Inter-rater agreement for ASPECTS was quite high (kappa 0.71–0.89). The researchers concluded in their article that the ASPECTS score is a reliable method for predicting functional outcome after thrombolytic therapy [3, 115]. Actually, the introduction of the scale as an assessment technique in the diagnosis of ischemic stroke was consistent, according to the results of this study, with the beginning of the international use of thrombolytic therapy [3].

Another reason for the variability of assessment between experts on this scale may be the heterogeneity of the groups of specialists participating in the studies [3]. Wilson A. T. et al. [81] indicate that experience, level of training and medical specialization may also influence the assessment. Most studies testing reliability and interrater agreement involved experienced neurologists and/ or neuroradiologists; a limited number of articles mention residents and junior specialists [3, 21, 45]. However, even among experienced clinicians, there may be differences in the interpretation of ASPECTS in different countries [69]. Other important but unresolved reasons for variability in ASPECTS scores include the following factors: software and hardware, location of the expert, time of day, time pressure, and personal qualities [3, 69].

Despite the fact that the ASPECTS is a method that is widely used in modern clinical practice, its main disadvantage is the possibility of variability in expert assessments, including among doctors with little experience, since high rates of inter-rater agreement were achieved mainly in groups with experienced radiologists. The pronounced diversity of results and the low level of inter-expert agreement do not currently allow this scale to be considered a truly reliable option for a standardized assessment and may affect the further treatment process [3, 5].

In this regard, the introduction into clinical practice of methods of semi-automatic and automatic processing of CT images using artificial intelligence systems, which in the future can improve the standardization of assessment.

1.3. Potential of automated CT image analysis systems in the diagnosis of the middle cerebral artery ischemic stroke and the applications of the ASPECTS

Artificial intelligence is a set of technological solutions that imitate human cognitive functions (including self-learning and searching for solutions without a predetermined algorithm), allowing, when performing tasks, to achieve results that are at least comparable to human intellectual activity. This concept includes information and communication infrastructure and software that use machine, representative and deep learning methods, as well as data processing and decision-making processes [6, 9]. AI is one of the rapidly growing fields of computer science and has significant implications for radiology. The most developed and used class of artificial intelligence methods is machine learning. It is used for partial or complete automation of solving complex professional problems based on accumulated data. Algorithms develop as the volume of available databases increases, improve their performance with experience, and also learn to give specific answers by processing large amounts of information [6, 100]. Before evaluating machine learning models, specification of medical diagnostic tasks is necessary as the models must be trained accordingly. This is achieved using the main typical tasks: supervised learning, unsupervised learning and semi-supervised learning [6]. These types of case-based learning tasks are fundamental strategies applied depending on the available data. In supervised learning, artificial intelligence extracts information from a certain part of trained samples with verified and labeled pathology in order to predict the results of unknown data [6, 44]. Conversely, unsupervised learning compares the “norm” with unlabeled pathology databases to implement a clustering task, which is to group objects into clusters using data on pairwise similarity of objects. In partial training, a combination of the above two methods takes place and only a small portion of the labeled training data is required. Unlabeled images are also used in training [6, 102]. It is necessary that the labeled data be reliable. This is why the concept of “ground truth” was introduced, which defines the testing of machine learning results for accuracy and is fundamental to testing the effectiveness of AI programs. The “ground truth” can be considered data that is reliably verified (confirmed by the “gold standard”) [6, 43]. Deep learning is a collection of machine learning techniques and the leading focus of most AI tools for image interpretation. Deep learning is algorithms represented as several layers of information processing that are interconnected. This multi-layer system of nonlinear filters is used to extract features with transformations, which means that each subsequent layer receives the output data of the previous layer as an input. Such multilayer algorithms form large artificial neural networks [6, 102]. Artificial neural networks must be “trained” using standardized data sets. In diagnostic imaging, they usually consist (at least initially) of manually labeled images used by algorithms for training (segmentation). Once the network is trained, it needs to be tested using another dataset designed to evaluate whether the learning model matches the required output [6]. At this stage, model overfitting is often observed. Yamashita R. et al. [42] describe overfitting as a situation “when a model learns statistical patterns specific to the training set, that is, ends up learning irrelevant noise instead of learning from the data, and therefore performs worse on a subsequent new data set. The consequence of overfitting is that the network will not generalize to data it has not previously encountered and will begin to make more errors [6, 42]. There is a direct correlation between the volume and quality of primary data and the model’s compliance with the required result. As network performance gradually improves due to the ability to train in multiple stages and test on heterogeneous data sets, its accuracy and generality are assessed before the algorithm is released for general use. Another solution is the so-called "data augmentation," which means modifying the training data by adding some new information [6, 42]. As a rule, the "deeper" the network (more layers) and the more training cycles, the higher the network performance. The fundamental area of application of AI in neuroradiology is the automatic segmentation of brain lesions, which makes it possible to relieve the radiologist from the labor-intensive function of performing manual segmentation [6, 29]. In clinical applications, radiologists primarily perform manual segmentation, which is subjective, time-consuming, and poorly reproducible with repeated procedures. Automatic segmentation is completely controlled by the algorithm without human intervention. The segmentation speed is high and the results are reproducible [6, 25]. Another application of AI in radiology is radiomics. This is the ability to represent complex objects in the form of a set of quantitative characteristics [6].

Software products based on radiomics make it possible to extract features from diagnostic images, the final product of which is the determination of a parameter of a specific pathology [6]. Radiological analysis can extract a large number of features from a region of interest in a CT or MRI study and correlate these features with each other and with other data [61]. Machine learning systems are currently applied to the diagnosis and classification of brain tumors [70], certain mental disorders [37], epilepsy [24], neurodegenerative disorders [99] and demyelinating disease [84]. Machine learning algorithms have also been developed to help diagnose and individualize treatment for acute ischemic stroke. One of the most important clinical criteria for successful thrombolysis with tissue plasminogen activator in acute ischemic stroke is its implementation within the first 4.5 hours from the onset of symptoms, but the onset of stroke is usually unknown [6]. To solve this problem, But K. S. et al. [39] developed a deep learning algorithm based on an autoencoder architecture to extract imaging features from MR perfusion images (PWI) to determine the time elapsed since stroke onset. Chen L. et al. [38], based on data from seven hundred and forty-one patients and a deep learning model that included two neural networks, developed an algorithm for segmenting brain lesions in stroke using DWI images. Measuring perfusion-diffusion mismatch and calculating the probability of infarct core zone formation using MRI-based approaches to assess tissue at risk can be used to guide decisions about the type of stroke treatment [6]. Bouts M. J. et al. [53] analyzed the ability of five algorithms to identify potentially viable brain tissue from MR images of rats subjected to right middle cerebral artery (MCA) occlusion without subsequent reperfusion, spontaneous reperfusion, or thrombolysis-induced reperfusion. The highest accuracy in identifying ischemic tissue likely to recover after subsequent reperfusion was observed using the generalized linear model (Dice similarity coefficient = 0.79 ± 0.14). Similarly, Huang S. et al. [67] used support vector machine (SVM) to predict infarction in rats on a pixel-by-pixel basis using cerebral blood flow velocity (CBF) and measured diffusion coefficient (ADC) images from MRI. Another application of machine learning systems in acute ischemic stroke is to predict factors that will contribute to the deterioration of neurological status and increase the likelihood of developing cerebral edema [6]. Chen Y. et al. [35] proposed definitions and measurements of cerebrospinal fluid volume over time, as it may represent a sensitive biomarker of the development and progression of cerebral edema. The initial cohort consisted of one hundred and fifty-five CT scans. Preprocessing was performed using a generalized estimating equations model to calculate age-adjusted cerebrospinal fluid volume in the brain. The results of the study showed that the decrease in cerebrospinal fluid volume over time correlated with the volume of the infarction, the presence of cerebral edema and the degree of displacement of the midline structures [6]. For comparison, Dhar R. et al. [30] presented an automated method for segmenting liquor spaces using a random forest machine learning ensemble with geometric active contour segmentation. In 38 patients, the cerebrospinal fluid spaces of the brain were marked within 6 hours from the onset of stroke, and then 24 hours from the onset of the disease. This approach made it possible to accurately track narrowing of the subarachnoid spaces of the brain as a correlative indicator of CSF volume. Pearson correlation coefficients between changes in CSF volume and normal values were statistically significant. The developed algorithms represent potential for future research and may serve as a biomarker for the severity of cerebral edema. The outcome of patients with acute ischemic stroke depends on the quality and timeliness of treatment, so the risks of complications should be taken into account when deciding on a specific type of therapy. Yu Y. et al. [92] developed a method for predicting the location and degree of hemorrhagic transformation in stroke, as the most severe complication after reperfusion therapy. Perfusion and DWI images of one hundred sixty-five patients receiving reperfusion therapy were analyzed using five machine learning approaches, with spectral kernel regression demonstrating an accuracy of 83.7 ± 2.6%. In a multicenter retrospective study [6, 82], researchers assessed the predictive ability of hemorrhagic transformation using MR perfusion data. MR perfusion imaging data were collected from 263 patients from four medical centers and served as input to linear and nonlinear predictive models with an average accuracy of >85% in predicting IS [6]. Nielsen A. et al. [94] conducted a study using a neural network they created with 9 biomarkers as input to calculate lesion volume in patients receiving intravenous thrombolytic therapy (IVT). The baseline data of 35 patients receiving intravenous IVT and 29 patients from the control group were compared. This model predicted final infarct volume with 88% accuracy. Bentley P. et al. [93], based on data from one hundred and sixteen patients (computed tomography and clinical severity scores) receiving IVT, developed a system to predict the risk of symptomatic intracerebral hemorrhage after intravenous thrombolysis therapy [6]. In their study, the SLM-based system provided better prediction than traditional prediction tools based on expert judgment data such as intracerebral hemorrhage after thrombolysis, early signs of infarction, hyperdense cerebral artery sign, age, and NIHSS scores [6].

Machine learning algorithms can also help predict motor disorders in stroke patients. Forkert N. D. et al. [83] used 12 support vector machine classification models to process MRI images to calculate the appropriate Modified Rankin scale (mRS) score for ischemic stroke patients over 30 days, using parameters that included different lesion involvement. brain regions, stroke laterality, and other additional characteristics such as infarct volume, NIHSS at presentation, and patient age. Superior prediction of neurological disorders by mRS was observed by integrating additional features and providing localization information, with multi-value mRS prediction accuracy of 56% and dichotomous mRS prediction accuracy (0–2 vs 3–5) of 85% [6]. In a study by Rondina J. M. et al. [48] created a model for predicting upper extremity motor deficits in 50 stroke patients, based on structural MRI data instead of functional MRI. Thus, there are currently successful attempts to apply all the capabilities of artificial intelligence systems to evaluate neuroimaging data in stroke: early detection by diagnostic imaging methods, assessment of the time of onset of the disease, lesion segmentation, analysis of the presence and possibility of cerebral edema, as well as prediction of complications and outcomes treatment. However, there are a number of limitations for the further development of artificial intelligence systems [6]. The first limitation is the sample size. Deep learning algorithms using medical imaging often require a significant amount of data, which, due to its specific nature, may not be available. For example, a machine learning algorithm that demonstrated superior performance in differentiating between malignant and benign skin lesions when compared to peer review by 21 dermatologists was trained on a dataset of nearly 130,000 images [6, 49]. A dataset of this size for public use does not currently exist. Obstacles in sharing data between institutions, as well as the lack of funding to properly pre-process and curate these images and restrictions on hosting such a dataset, are responsible for some of the delays in the creation of this repository [6].

Another limitation encountered in neuroimaging-based machine learning methods is the need to label regions of interest or “gold standard” findings in images. In other words, in addition to collecting images, their marking, identification and segmentation are necessary. For example, to train an algorithm that evaluates the presence or absence of hyperdense SMA, preliminary marking by a “teacher” is required. Taking into account the fact that the human expert resource is quite limited, there is a request to reduce the need for its use in training artificial intelligence models. It should be noted that most of the results of the algorithms presented by the authors have not currently been independently evaluated in clinical practice. The medical community will have to undergo extensive clinical testing of the developed artificial intelligence systems on independent data sets. Randomized studies are required to evaluate the long-term results of artificial intelligence systems, which may change the diagnostic accuracy of these algorithms. Also, based on independent clinical testing, it is possible to recommend specific software products for medical practice. A separate point is the study of the interaction between a doctor and machine learning systems, as well as the impact of this collaboration on the decision-making process, indicators of the quality of medical care, and the duration and patients' quality of life [6].

To partially solve the problems of subjectivity of assessment on the ASPECTS scale, a number of authors propose the use of artificial intelligence systems. To date, automated image analysis programs have focused on the analysis of non-contrast CT, CT angiography, and CT- or MRI-based perfusion imaging. They are aimed at identifying and quantifying the stroke core, penumbra, collateral blood flow status and localization of arterial occlusion in an automatic mode. One of the software options for diagnosing ischemic changes was presented by Wolff L. et al. [113] compared its performance with the analytical abilities of medical specialists. Their study obtained specificity of 89–89%, sensitivity of 41–57%, and accuracy of 0.750–0.795. The authors concluded that the diagnostic accuracy of this system was comparable to that of participating physicians and could assist radiologist in detecting early ischemic changes [6, 113].

Also, a number of studies have focused on improving inter-rater agreement through the implementation of automatic analysis systems. Thus, Delio P. R. et al. found that the use of artificial intelligence algorithms increases the agreement between experts from 72 to 78% [6, 34].

Also, Culbertson C. J. et al. concluded that using automatic segmentation it is possible to increase inter-rater agreement among experts with little experience in diagnosing ischemic stroke [6, 109].

Several clinical software products are currently being tested to assist radiologists in image interpretation in acute ischemic stroke, capable of automatically assessing CT data and assigning an ASPECTS score. A limitation of these studies is the small number of data included in the study. Also, the studies do not analyze the impact of various reconstruction algorithms on the diagnostic performance indicators of current system options. But the first results show that the use of these algorithms improves inter-rater agreement when assessing the ASPECTS scale [6, 34, 40]. At the same time, the algorithms of automatic analysis systems are not intended for use as stand-alone diagnostic tools. They can help clinicians obtain more accurate and standardized interpretation of CT and MRI findings, which can improve patient management and functional outcome [6, 31].

In this regard, the introduction into clinical practice of methods of semi-automatic and automatic processing of CT images using artificial intelligence systems, which, according to the results of the first studies, improve the standardization of assessment. But for the full adoption of such systems into clinical practice, their clinical testing on independent sets of different data is necessary [6, 33].

According to Obuchowski N.A. et al. [88], there are four options for using AI algorithms in medical imaging: (1) a first or parallel reader mode, where the AI algorithm first provides interpretation of the CT images and then human gives interpretations incorporating AI results, (2) a second reader mode, where the human gives initial interpretations without AI results, then AI provides interpretation and human gives final interpretation incorporating AI results, (3) a triage mode, where the algorithm sorts cases according to the presence of the suspected pathology and then human gives interpretation incorporating for a prioritized worklist of cases (4) a pre-screening mode is applied to a set of images to identify the norm with the generation of a clinical report, and further assessment of the remaining (classified as “unknown”) cases by a radiologists (Figure 7).

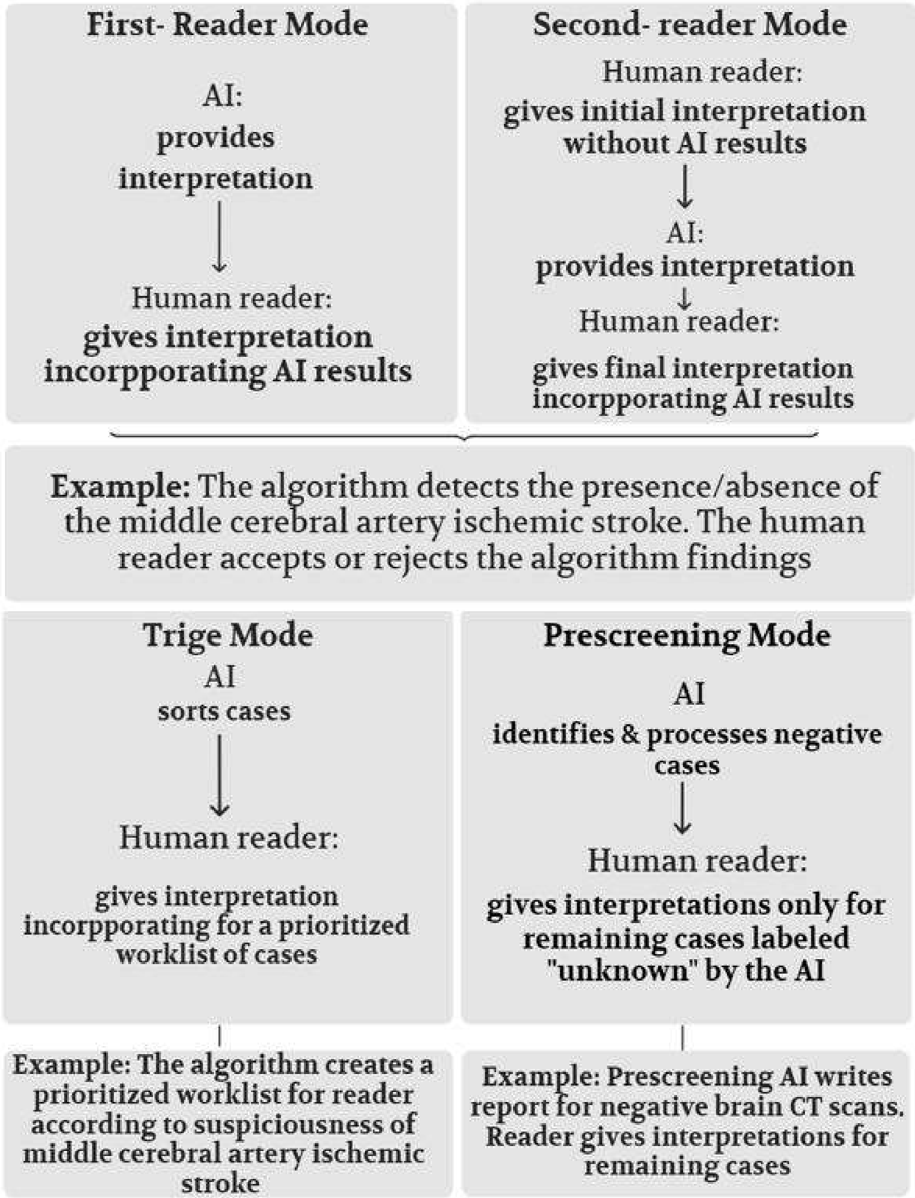

Figure 7 – Illustration of four use cases for AI [88]

These modes can only be used if pathological changes are correctly diagnosed by artificial intelligence algorithms. In this case, the programs must achieve and prove the following parameters: high accuracy, specificity and sensitivity. However, despite a certain Obuchowski N. A. et al. [88] the order of interaction between a doctor and artificial intelligence models, it still remains unclear how these types of complementary assessment will affect the diagnostic efficiency of doctors, which model is the most effective, and whether such collaboration will cause negative dynamics in the correct interpretation of radiation images.

Also, there are a number of restrictions for the implementation of artificial intelligence algorithms. These include not only practical considerations such as the compatibility of these technologies with other systems, including PACS and electronic health records, but also the need to verify the accuracy of the programs, since most of the results of the presented algorithms have not currently been independently evaluated in clinical practice. According to the increasing complexity and variety of artificial intelligence tools, the technical characteristics of these applications are not always obvious to the neuroradiologist, especially since developers very rarely provide complete information about their products. Thus, the doctor is faced with the dilemma of whether or not to base his decision on the results of artificial intelligence algorithms, without knowing in detail what is happening inside the “black box” of these systems. This dilemma is aggravated by the fact that the accuracy of the algorithms directly depends on the characteristics, quality and structure of the databases (tomograph, slice thickness, presence of technical and dynamic blur artifacts) used for training, as well as the qualifications of specialists acting as “teachers” of the software product. Also, often, when training an automatic analysis system, publicly available databases are used without proper verification of pathology, which cannot increase the level of confidence of doctors in artificial intelligence algorithms.

An extremely important factor in the implementation of artificial intelligence systems is the ergonomics of each individual automatic analysis system. With this integration, it must be proven that their operation not only takes a small amount of time, but is also convenient to use. A critical aspect remains the change in the radiologist’s workflow regulations.

At the same time, the lack of a structured approach to the implementation of artificial intelligence technologies and significant differences in the ways of using such systems [32] cannot but affect the quality of medical care, the functional outcome of the disease, and requires detailed study with the provision of the most effective model of complementary assessment various pathological changes.

Thus, only if the competence and effectiveness of the implementation of automatic analysis systems has been proven in all aspects, improving the working conditions of the radiologist and increasing the qualifications of his assessment, will there be further discussion of the successful artificial intelligence implementation in healthcare.

CHAPTER 2. GENERAL DATA ON CLINICAL MATERIAL AND RESEARCH METHODS

2.1. Development of a database for CT examinations of patients with a clinical presentation of middle cerebral artery stroke

To study the quality of interpretation of CT by radiologists, as well as the diagnostic effectiveness of automated CT image analysis systems, at the second stage of the dissertation research, a database was developed [19], consisting of non-contrast CT of the brain of 150 patients with the clinical presentation of middle cerebral artery stroke admitted to the regional vascular center of St. Petersburg in the period from December 1, 2020 to December 30, 2021(according to the Los Angeles Motor Deficiency Scale - LAMS).

The clinical presentation of all patients included middle cerebral artery syndrome (contralateral to the lesion hemiplegia or hemiparesis, hemihypesthesia, hemianopsia). In 43 patients (35%), hemiparesis was more pronounced in the upper extremities. In 15 patients (1%) patients, gaze paresis in the direction of the lesion was detected. 84 patients (56%) were diagnosed with various types of aphasia - efferent and afferent motor aphasia, sensory aphasia, and their combination [12]. In 10 patients (6%), the clinical picture included severe focal hemispheric symptoms of the hemitype, including total aphasia or anosognosia.

All patients underwent a comprehensive clinical and neurological examination, which included a thorough collection and analysis of complaints, medical history, taking into account concomitant somatic pathology, and an objective and neurological examination. The patients were examined according to the standard procedure for examining a neurological patient. The degree of neurological deficit and severity of stroke was assessed at admission and over time using the NIHSS scale, activity in daily life - using the Barthel and Rivermead Mobility Index, intellectual-mnestic disorders -

using the MMSE (Mini-mental State Examination) scale, functional capacity - using the Rankin scale on the day of admission and upon discharge from the hospital.

Concomitant somatic pathology was diagnosed in 120 patients (80% of patients): arterial hypertension in 101 patients (67%), diabetes mellitus in 23 patients (15%), atrial fibrillation in 20 patients (13%), myocardial infarction in 3 patients (2%).

In 100 patients, middle cerebral artery stroke was confirmed, in 50 patients' ischemic stroke was excluded. In 79 (79%) patients, the cardioembolic stroke was diagnosed, in 21 (21%) - atherothrombotic, according to the TOAST classification (Trial of Org 10172 in Acute Stroke Treatment). For these patients, the diagnosis of ischemic stroke was established by a neurologist in accordance with the recommendations of the Ministry of Health of the Russian Federation and verified using CT angiography and CT perfusion data [12]. Patients with ischemic stroke excluded also underwent CT angiography, CT perfusion and a following CT (after 24 hours), which did not reveal pathological changes.

CT were obtained using a GE Revolution EVO 128 (148 CT) and a Toshiba Aquilion 64 (2 CT) CT scanner. CT were carried out in a standard position - lying on the back, arms along the body, in a headrest, without holding the breath. Scanning was carried out from the convex to the level of the atlantoaxial joint. The matrix size was 512x512, slice thickness 5 mm, with the possibility of reconstruction up to 1.3 mm (reconstruction algorithms - convolution Kernel), pitch factor 0.53, DFOV (display field of view) along the X/Y axis was 180-324 mm.

To confirm or exclude stroke, patients additionally underwent CT angiography and CT perfusion. CT angiography was also performed in the standard (described above) setup, with the introduction of a nonionic contrast agent (Omnipaque 300 mg/ml) in a volume of 80 ml using an automatic syringe injector. When using it, the injection rate was 2-3 ml/s with a delay in the start of scanning of 50-60 s. The tomography range is from the aortic arch to the cerebral convex. The administration of a contrast agent was performed for health reasons without taking into account general contraindications (serum creatinine >1.5 mg/dL (>130 μM/L) and hyperthyroidism). The matrix size was

512x512, slice thickness was 0.6 mm, DFOV (display field of view) along the X/Y axis was 180–412 mm.

CT perfusion was performed with intravenous contrast with a non-ionic contrast agent (Omnipaque) with an iodine concentration of 300 mg/ml, a volume of 50 ml, bolus with an automatic syringe-injector into the cubital vein at a rate of 5-7 ml/second. Data recording began 4 seconds after the start of contrast agent administration and continued for 50 seconds at 1 second intervals. Tomography parameters: slice collimation 4-8 mm, tube voltage – 80 kV, 200 mA, tube rotation time – 1 s. Scanning was carried out without holding your breath, the slice block (8.5 mm) was positioned to cover the largest territory of middle cerebral artery. Post-processing of the dynamic scanning series data was carried out in a specialized software application. The effective radiation dose for the entire complex of CT studies averaged 20 mSv.

All data was depersonalized using the Launch DicomAnonymize program.

The database, after verification by CT angiography and CT perfusion, included 100 patients with confirmed thrombosis of the middle cerebral artery and signs of ischemic changes in the most acute and acute stages in this pool, as well as 50 patients with a diagnosis of MCA stroke was not confirmed (including during the following non-contrast CT study after 24 hours). Database registration certificate RU 2022620850 [19].

The database combines: data from non- contrast CT of the brain (DICOM formats), CT angiography of cerebral vessels, brachiocephalic arteries (DICOM format), CT perfusion (presence of penumbra area and the mismatch of the stroke core, DICOM and JPEG formats); information about the pathological process - the presence of signs of acute and acute cerebrovascular accident, the presence of changes in the control CT study (hemorrhagic transformation of stroke, absence of formed ischemia) (Figure 8).

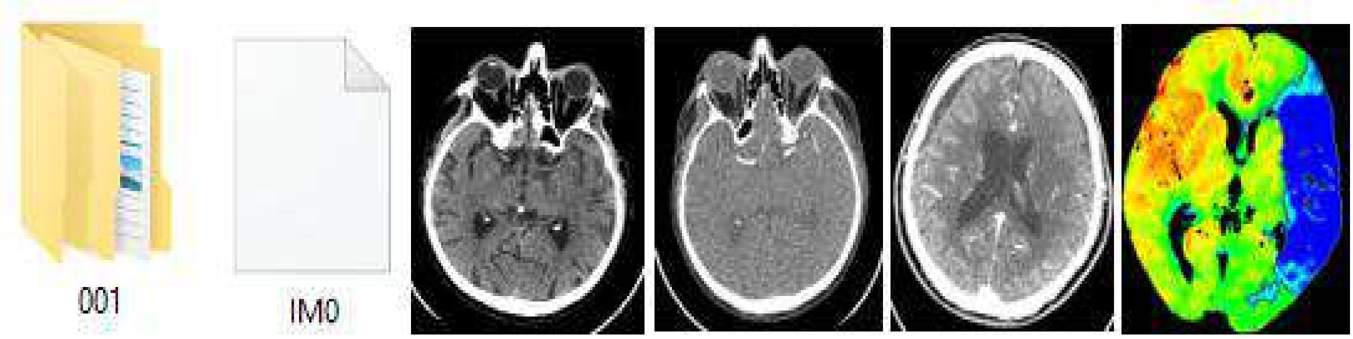

Figure 8 – An example of data from a database of CT images of patients with acute ischemic stroke clinical presentation

The database also includes information such as the gender and age of patients (Table 1), the period of CT scanning from the onset of the disease (up to/more than 6 hours), the stroke territory (left/right middle cerebral artery), the segment of the middle cerebral artery in which thrombosis was detected (Table 2), features of the development of the circle of Willis (posterior/anterior trifurcation of the ICA/closure), the presence of signs of ischemic stroke (assessment was made based on the consensus of three neuroradiologists with more than 10 years of experience in emergency medicine, as well as based on CT perfusion data [46]), scores on the ASPECTS scale (Table 3), CT perfusion data (presence of penumbra, percentage of nucleus to penumbra ratio), data on thrombus extraction, presence of a control non-contrast CT study (scores on ASPECTS/data according to the Heidelberg classification, ECASS criteria), outcome (Table 4), as well as technical data (tomograph/presence of dynamic artifacts/slice thickness (non-contrast CT/CTA)/reconstruction algorithms (convolution Kernel)).

Table 1 – Distribution of data on gender and age of patients included in the database

<table>
<tr><th>Patient group</th><th>Number of people</th><th>Average age</th><th>Female</th><th colspan="2">%</th><th>Male</th><th colspan="2">%</th></tr>
<tr><td>Patients with confirmed diagnosis of AIS</td><td>100</td><td>72,7 (36 - 95)</td><td>52</td><td>52</td><td rowspan="2">51,3</td><td>48</td><td>48</td><td rowspan="2">48,7</td></tr>
<tr><td>Patients with excluded diagnosis of AIS</td><td>50</td><td>66,8 (23-96)</td><td>25</td><td>50</td><td>25</td><td>50</td></tr>
</table>

Table 2 – Distribution of data on the period of CT scanning from the onset of the disease, stroke area, occluded segment of the middle cerebral artery in patients included in the database

<table>
<tr><th rowspan="2">Patient group</th><th rowspan="2">Number of people</th><th colspan="3">Time to CT scan imaging</th><th colspan="2">Sides of hemispheres</th><th colspan="2">Thrombosis of the MCA segment</th></tr>
<tr><th>Less than 6 hours</th><th>More than 6 hours</th><th>More than 24 hours</th><th>Left MCA</th><th>Right MCA</th><th>M1</th><th>M2</th></tr>
<tr><td>Patients with confirmed diagnosis of AIS</td><td>100</td><td>69</td><td>31</td><td>71</td><td>52</td><td>48</td><td>79</td><td>21</td></tr>
<tr><td>Patients with excluded diagnosis of AIS</td><td>50</td><td>50</td><td>0</td><td>28</td><td>-</td><td>-</td><td>-</td><td>-</td></tr>
</table>

Table 3 – Distribution of data on signs of ischemic stroke and ASPECTS rating in patients included in the database

Patient group	Number of people	Signs of ischemic stroke		ASPECTS		
		HAS	EIC	10	от 9 до 6	от 5 до 0
Patients with confirmed diagnosis of AIS	100	79	67	19	51	30
Patients with excluded diagnosis of AIS	50	0	0	50	0	0

Table 4 – Distribution of CT perfusion, thrombus extraction, and outcome data

Patient group	Number of people	Perfusion			ITE	Patient Outcome Data		
		Penumbra	Stroke core ratio			Hemorrhagic transformation of stroke	Negative affective of ASPECTS rating within 24 hours	Death
			less than 40%	more than 40%				
Patients with confirmed diagnosis of AIS	100	100	82	18	73	33	49	20
Patients with excluded diagnosis of AIS	50	0	0	0	0	0	0	0

The base is intended for training radiologists from offices and departments of computed tomography with different experience and experience in identifying signs of ischemic changes in the most acute and acute stages in the middle cerebral artery basin, testing the qualifications of radiologists, testing automated CT analysis systems of images obtained by computed tomography.

Based on the database, collections of images 2 of non-contrast CT of the brain was generated. The first collections of images included non-contrast CT of 50 patients with confirmed stroke and 50 patients in whom this diagnosis was excluded (pathology/normal ratio - 1:1), using CT perfusion, CT angiography, and also conducting a following CT scan every other day. Three options of collections of images were prepared for the third (1A) and fifth stages (1B and 1C) of the dissertation research with a changed order of brain CT. The collections of images №2 included non-contrast CT of 50 patients with confirmed MCA stroke, as well as occlusion of the M1 segment of the middle cerebral artery (for further more correct assessment on the ASPECTS scale), according to CT angiography and CT perfusion (Table 5). Three options of collections of images were prepared for the third (2A) and fifth stages (2B and 2C) of the dissertation research with a changed order of brain CT scans (Table 6).

Table 5 – Distribution of patients in database-based samples

<table>
<tr><th>Patient group</th><th>Number of people</th><th colspan="2">Collection of images 1</th><th colspan="2">Collection of images 2</th></tr>
<tr><td>Patients with confirmed diagnosis of AIS</td><td>100</td><td>50</td><td rowspan="2">100</td><td>50</td><td rowspan="2">50</td></tr>
<tr><td>Patients with excluded diagnosis of AIS</td><td>50</td><td>50</td><td>0</td></tr>
</table>

Table 6 – Stages of testing radiologists with correlation with the options of CT collections of images used

Research scale			Collection of images
Third stage of the study	Assessment in CT examinations by radiologists (n=21) with different experience and expertise in emergency medicine	Presence/absence of MCA stroke, side of stroke	1A
		Side of stroke, final ASPECTS rating	2A
Fifth stage of the study	Assessment in CT examinations by radiologists (n=7) with three years of experience and different expertise in emergency medicine	Presence/absence of MCA stroke, side of stroke	1B
		Side of stroke, final ASPECTS rating	2B
	Assessment in CT examinations by radiologists (n=7) with three years of experience and different expertise in emergency medicine	Presence/absence of MCA stroke, side of stroke	1C
		Side of stroke, final ASPECTS rating	2C

The inclusion criteria in collections of images №1 for patients with confirmed MCA stroke is the presence of acute ischemic stroke, verified using CT angiography and CT perfusion [4, 5].

Exclusion criteria for this group of patients:

1) Recurrent middle cerebral artery stroke;
2) Ischemic posterior circulation stroke and anterior cerebral artery stroke;
3) Cerebral venous and sinus thrombosis;
4) Chronic occlusion of arteries of the anterior circulation;
5) No change on CT perfusion maps.

Criteria for inclusion in the study in collections of images №1 for patients in whom the diagnosis of stroke was excluded:

1) absence of CT signs of ischemic changes according to non-contrast CT in the middle cerebral artery territory, with confirmation of the absence of thrombosis of the brachycephalic and cerebral arteries and veins using CT angiography, as well as CT perfusion;
2) absence of CT signs of ischemic changes in the middle cerebral artery territory according to a following non-contrast CT performed after 24 hours.

Exclusion criteria for this group of patients:

1) stroke and other pathological changes detected by CT.

The inclusion criteria in collections of images № 2 was the presence of middle cerebral artery stroke (with thrombosis of the M1 segment), confirmed by CT angiography and CT perfusion maps.

Exclusion criteria:

1) Recurrent middle cerebral artery stroke;
2) Ischemic posterior circulation stroke and anterior cerebral artery stroke;
3) Cerebral venous and sinus thrombosis;
4) Chronic occlusion of arteries of the anterior circulation;
5) No change on CT perfusion maps.

2.2. Problems and issues of inter-rater reliability among radiologists with different experience and expertise in emergency medicine in the diagnosis of middle cerebral artery ischemic stroke

At the third stage of the thesis work, two types of testing were carried out.

The first testing was attended by 21 radiologists (Table 7), working either in regional vascular centers (n=12), or in other medical institutions (general hospitals) not related to urgent medicine (n=9) and without permanent experience in stroke assessment. Doctors were divided into groups according to their experience: less than three years (n=7), three years to less eight years (n=7) and more than eight years (n=7).

Table 7 – Data of radiologists with different experience and expertise in emergency medicine at the first stages of research

Hospital	RVC			GH		
Years of experience	<3 years	3 - 8 years	> 8 years	<3 years	3 - 8 years	> 8 years
Number	4	4	4	3	3	3

For the evaluation, the collections of images 1A of depersonalized data from the generated database was used, which included 50 CT of patients with verified ischemic stroke, and 50 CT studies of patients with an unconfirmed diagnosis of stroke.

CT of patients with confirmed middle cerebral artery stroke from the first collections of images were consensually assessed by three radiologists with expert qualifications in urgent neuroradiology and more than 10 years of experience. The radiologists jointly came to the decision that all CT with confirmed stroke showed signs of early ischemic changes (with ASPECTS 9 and lower) in the territory of the occluded middle cerebral artery.

During the first testing, in the process of analyzing non-contrast CT, specialists filled out a form with answers, interpreting CT images as normal or pathological, as well as the presence of such signs as the side of the lesion, the presence of a hyperdense artery sign, a sign of loss white- gray matter differentiation (also including includes the "insular ribbon sign" and the disappearing basal ganglia sign), as well as sulcal effacement and a hypoattenuation.

After providing non-contrast CT to radiologists, the CT scans were viewed using the RadiAntDICOM Viewer program (with the ability to construct MPR image reformation), as a universal tool installed on any computer. Images were viewed on HP EliteDisplay S340c monitors with a resolution of 3440x1440 megapixels, Windows 10 operating system.

Window settings included both standard values of window width of 80 HU and window level of 40 HU, and non-standard values with narrow window width and window level components of approximately 35-45 HU for width and 35-45 HU for window level, respectively, for evaluation according to the ASPECT scale.

The experts independently assessed each study and marked the signs if present as “1”, if absent as “0”. Next, all assessments were combined by an independent researcher into a common table for further statistical processing of the results.

Case 1 – Noncontrast brain CT image of the brain with a hyperdense artery sign. The absolute CT density of the right middle cerebral artery is +63 HU (should be at least +43 HU), with a left MCA ratio of 1.2 (Figure 9).

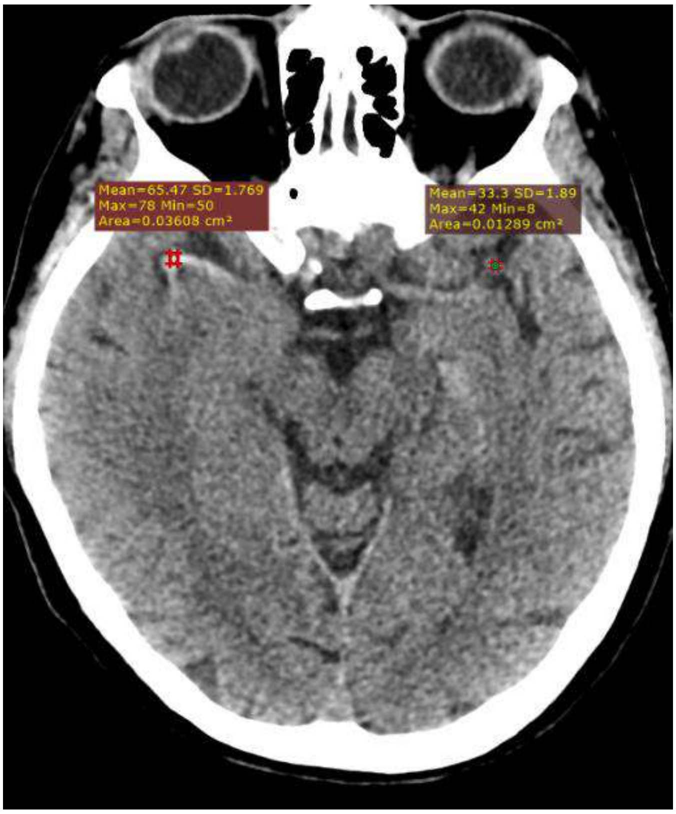

Figure 9 – Right middle cerebral artery ischemic stroke. CT examination data, Case N1

Case 2 – Noncontrast brain CT image with loss of basal ganglia (lentiform nucleus) sign of the right hemisphere (Figure 10).

Figure 10 – Right middle cerebral artery ischemic stroke. CT examination data, Case N2

Case 3 – Noncontrast brain CT image with a sign of sulcal effacement of the left frontal lobe (Figure 11).

Figure 11 – Left middle cerebral artery ischemic stroke. CT examination data, Case N3

Case 4 – Noncontrast brain CT image (Figure 12), with a hypoattenuating brain tissue - Insular Ribbon sign (up to +20 HU).

Figure 12 – Left middle cerebral artery ischemic stroke. CT examination data, Case N3

The second testing involved 15 radiologists (Table 8), working only in the RVC, divided into equal groups according to years of experience: less than three years (n=5), three years to less eight years (n=5), more than eight years (n=5). Taking into account the specifics of their work and the routine use of the ASPECT scale to assess the prevalence of ischemic changes in the territory of the middle cerebral artery, radiologists were asked to rank the ASPECTS score according to their subjective opinion. For the evaluation, the collections of images 2A of depersonalized data from the generated database was used, which included 50 CT of patients with verified ischemic stroke (with identified acute occlusion of the M1 segment of the MCA).

Table 8 – Data of radiologists with different years of experience of the second stage of testing

	RVC		
Years of experience	<3 years	3 - 8 years	<8 years
Number	5	5	5

Descriptive statistics methods were used to carry out the statistical analysis. The analysis was performed using SPSS Statistics 19 and the Python programming language.

Based on the results of the evaluation of the first test, calculations were made of such parameters as sensitivity, specificity, accuracy for assessing the presence/absence of MCA stroke (ROC analysis was carried out and the area under the ROC curve was calculated) [4, 77].

When assessing CT signs of ischemic stroke, statistical analysis included determining the coefficient of inter-rater agreement (Fleiss kappa) regarding CT signs of ischemic stroke [64]. The coefficient of interrater agreement (Fleiss's kappa) was assessed as follows: slight agreement, 0.00 to 0.20; fair agreement, 0.21 to 0.40; moderate agreement, 0.41 to 0.60; substantial agreement, 0.61 to 0.80; or almost perfect agreement, 0.81 to 1.00.) [75].

Based on the results of the assessment of the second test, statistical analysis included the determination of inter-rater agreement coefficients (Cohen's kappa - by alternately assessing the agreement of experts with each other and Fleiss's kappa) [59]. The results of assessing interrater agreement using Cohen's kappa (k) are presented as a range of values from minimum to maximum (min k - max k), as well as the mean value (μ). The use of two coefficients to measure inter-rater agreement is due to different options for statistical analysis in the literature and is necessary for correct comparison of the obtained data with the data of other researchers. In addition, statistical analysis was performed by dichotomizing ASPECTS scores: ≤ 6 and > 6 and ≤ 7 and > 7, as the literature differs in its discussion of the cutoff ASPECTS score correlated with worse functional outcome as well as the risk of intracerebral hemorrhage [112]. The coefficients of interrater agreement (Cohen's kappa and Fleiss's kappa) were assessed as follows: slight agreement, 0.00 to 0.20; fair agreement, 0.21 to 0.40; moderate agreement, 0.41 to 0.60; substantial agreement, 0.61 to 0.80; or almost perfect agreement, 0.81 to 1.00 [75].

2.3. Testing of automated CT image analysis systems

At the fourth stage of the dissertation research, we selected for testing three software programs based on convolutional neural networks, positioning themselves as a software device for computer-aided detection, used by a radiologist to interpret the nature and distribution of pathological changes in brain tissue by assessing the ASPECT scale based on data CT.

Below are given the testing selection criteria:

1. Availability of a test online access;
2. The description of the software program indicates the function of rating areas with early ischemic changes in the acute period (ASPECTS);
3. The declared accuracy in diagnosing middle cerebral artery stroke is more than 75% (taking into account the results of first stage of the thesis work, described in Chapter 3).

According to the inclusion criteria, for software program were selected (two systems of domestic developers and one foreign system). Since the research was aimed at the general assessment of diagnostic indicators of currently available systems and not at the assessment of a particular product, all programs in the research were disguised as A, B, C, D. The accuracy, sensitivity and specificity of automated CT image analysis programs declared by the developers are presented in Table 9.

Table 9 – Stated indicators of diagnostic effectiveness of software products participating in the study

Automated CT analysis systems	Accuracy	Sensitivity	Specificity
A	97,0	99,0	94,0
B	83,0	82,0	82,0
C	80,0	90,0	70,0

Program A is positioned by the developers as software based on artificial intelligence (AI) technology designed for processing CT images of the brain in cases of suspected stroke for the purpose of early detection of urgent pathology (ischemic stroke) and notification of it. The requirements for CT images of the brain to ensure optimal performance of program A are presented in Table 10.

Table 10 – Recommended image acquisition parameters for algorithmic processing purposes

Scan area	In CT examinations recommended the area from the first and second cervical vertebrae level to cerebral convexity is scanned
Image matrix size	512x512
Slice thickness	0,625-5 mm
DFOV	180-324 mm

In accordance with the ASPECTS recommendations, 10 regions of interest are identified for each cerebral hemisphere, which are outlined in the image with a yellow outline. Regions of interest identified by the ASPECTS algorithm as having EIC are outlined in red and appear in the ASPECTS reporting panel to the right of the viewport. The overall ASPECTS score is displayed on the reporting panel and reflects the number of areas out of 10 identified as unaffected (shown in green), while the number of areas out of 10 identified as involved in the disease process are shown in red. Additionally, the user can turn image overlay AOIs on or off and manipulate the incoming non-contrast CT image by prompting for the "ASPECTS: ON/OFF" annotation in the viewing area (Figure 13).

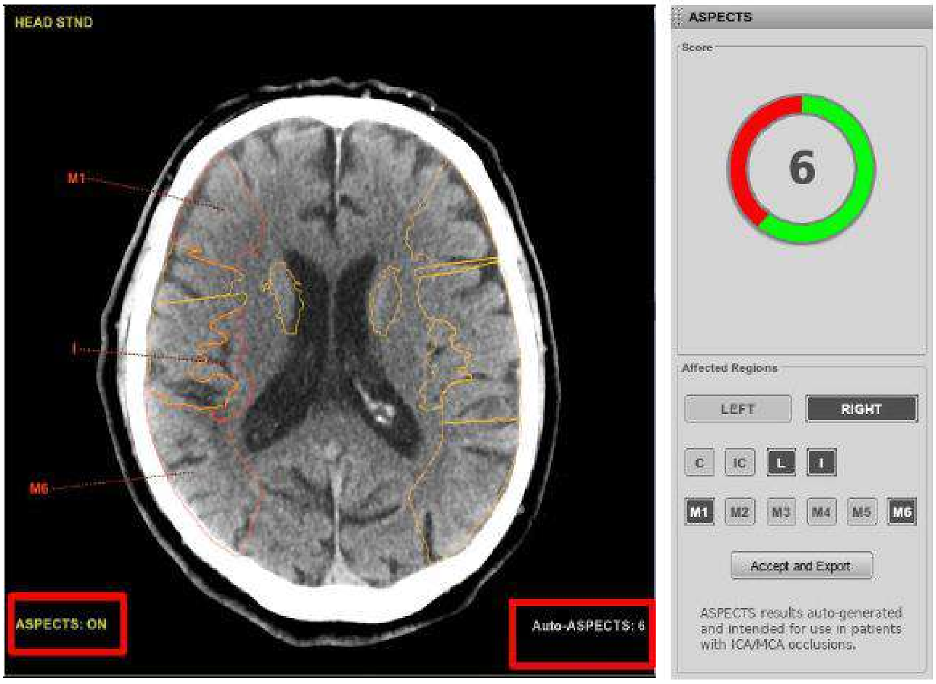

Figure 13 – Interface and result of non-contrast CT processing - brain research with Artificial intelligence software A

Similarly, Program A scores 10 (without EIC), all areas in the reporting panel are shown in gray, and all area overlays in the viewport are shown in yellow.

Program A was trained on a database consisting of CT scans obtained from the Prove-IT multicenter clinical trial (ClinicalTrials.gov identifier: NCT02184936) involving patients aged 22 years and older. For analytical validation of Program A, a data set of 200 patients was used, from two clinical trials and two randomized controlled trials: Prove-IT (N = 40, ClinicalTrials.gov identifier: NCT02184936), INTERRSeCT (N = 59, JamaNetwork.com identifier: 2702146), ESCAPE (N = 16, ClinicalTrials.gov Identifier: NCT01778335), ESCAPE-NA1 (N = 85, ClinicalTrials.gov Identifier: NCT02930018). Test data was obtained from medical institutions in various geographical regions (Canada, USA, EU, Asia), performed on different computed tomographs (GE, Siemens, Philips, Toshiba). CT images of patients included in the analytical validation database were representative in nature with a wide range of clinical severity grades (ASPECT score range 0–10, median ASPECTS score = 8; NIHSS score range 0–30, median NIHSS score = 17) and time from onset of symptoms to CT < 360 minutes.

Program B ((Registration No.58.29.32-001-14161592-2022) is software designed to process and maintain the quality of description of medical diagnostic images of patients with acute cerebrovascular accident. The program determines the presence of foci of ischemia and segments them, highlighting areas in the form of a mask (red color corresponds to areas of acute ischemia, blue - chronic ischemia). The system also allows you to mark the territories of the middle cerebral artery basin, determined on CT when assessed according to the ASPECTS scale, in the form of a mask (Figure 14).

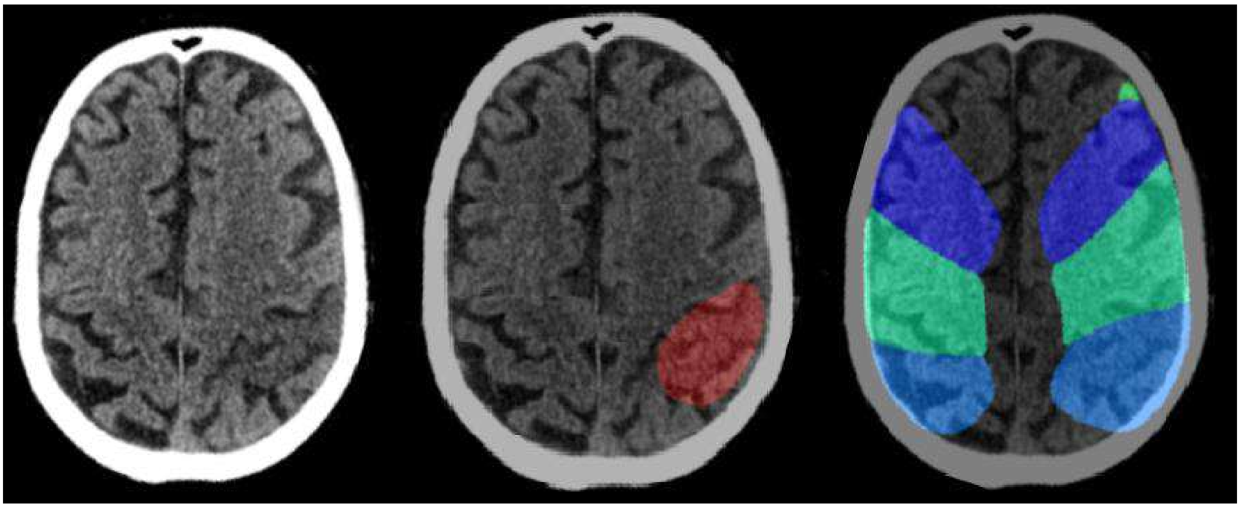

Figure 14 – An example of identifying area ASPECTS with early ischemic CT signs in the territory of the left middle cerebral artery using artificial intelligence software B

Each color corresponds to one region:

• red – caudate nucleus (C),

• pink – lentiform nucleus (L),

• yellow – internal capsule (IC),

• orange – insular cortex (I),

• green – anterior part of the cortical region of the MCA (M1),

• purple – cortical region of the territory of middle cerebral artery, lateral to the insular cortex (M2),

• turquoise – posterior part of the MCA cortex (M3),

• blue – anterior area of the territory of middle cerebral artery, located immediately above and rostral to M1 (M4),

• mint – lateral area of the territory of middle cerebral artery, located immediately above and rostral to M2 (M5),

• blue – posterior area of the territory of middle cerebral artery, located immediately above and rostral to M3 (M6).

The program was trained and validated on a data set of 600 brain CT of men and women over 18 years of age, generated in accordance with GOST R 59921.5-2021 [18].

Program C (Registration No. 2013616688) is intended to support medical decision-making for the work of radiologists and neurologists in assessing the degree of brain damage in ischemic stroke using images obtained on computed tomographs (CT).

The program provides:

- registration and visualization of CT images of the patient's brain in the DICOM standard;
- automatic segmentation of CT hypodense areas in native CT images of the head;
- marking of focal corresponding to areas of the ASPECT scale with calculation of the total score characterizing the identified changes;
- forming a conclusion with an assessment of the volume of the hypoattenuating area and the score on the ASPECTS scale;
- representation of the affected areas ASPECTS in the form of a schematic image.

Recommended slice thickness for CT studies is >1 mm.

Areas of the brain in which pathological changes are suspected are highlighted in light blue in the form of a mask. If they are located in areas corresponding to ASPECTS areas, the total score on this scale is assessed. An additional service marks in beige hypoattenuating areas that are highly likely to be cystic atrophic changes (Figure 15). (Figure 15).

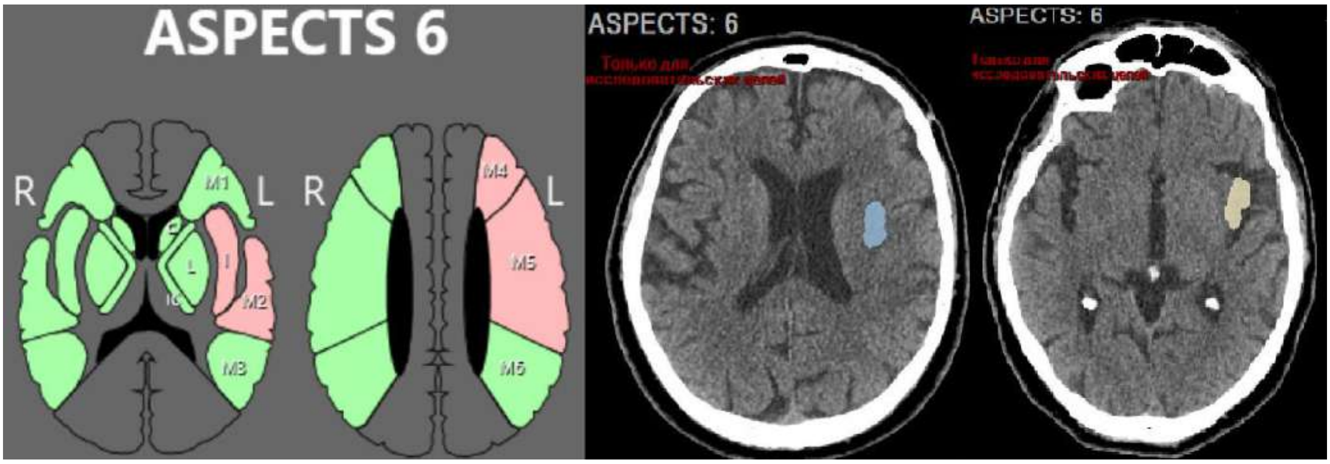

Figure 15 – An example of highlighting a zone of interest by Artificial intelligence software C, with the calculation of affected areas according to ASPECTS in the territory of middle cerebral artery, as well as cystic-atrophic changes in the left insula

The developers indicate that the service's conclusion contains a probabilistic assessment of the presence of ischemic stroke and a score on the ASPECTS scale.

Limitations of using the service include CT artifacts at the scanning level, surgical interventions performed on the brain, the presence of concomitant pathology (oncology), computed tomography slice thickness of more than 3 mm, as well as technical research artifacts that do not relate to the patient (related to a malfunction of the tomograph). It is worth noting that software product C did not interpret sixteen CT scans of the brain from collections of images № 1, due to an unrecoverable technical error when loading CT images.

Testing of software products was carried out using the method of analytical validation - assessing the effectiveness of an artificial intelligence system by using reference data with confirmation of the program's ability to reproducibly and reliably generate the intended technical results of calculations from input data [15, 18].

For testing, collections of images №1 of images was used. The software products provided the results of CT image analysis in the form of ASPECT score and side of the lesion. Since all CT studies with confirmed stroke were previously consensually assessed by three doctors with expert qualifications in urgent neuroradiology (with experience in diagnosing stroke), as a result of which they jointly came to the decision that all CT studies with confirmed stroke had signs of early ischemic changes (with ASPECTS 9 and below), then the answer to the programs “ASPECTS 10” was interpreted as the absence of MCA stroke, the answer “ASPECTS 9 and below” as the presence of MCA stroke.

Statistical analysis was performed using SPSS Statistics 19 and Python programming language. The performance indicators for detecting ischemic stroke in the middle cerebral artery basin on computed tomograms were assessed (sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, number of true positive, false positive, false negative and true negative responses, as well as accuracy). To compare these indicators with the results of doctors tested at the third stage of the dissertation research, for RVC radiologists with less than three years of experience and more than 8 years, as well as their GH colleagues with less than three years of experience, additional diagnostic efficiency indicators were calculated (positive likelihood ratio result, the likelihood ratio of a negative result, the number of true positive, false positive, false negative and true negative responses). In addition, graphs were constructed to evaluate the quality of binary classification - characteristic curves (ROC curves) [4, 7]. During the analysis, all metrics were assessed in the range of 0–1: <0.6 - unsuitable; 0.61–0.8 - requires improvement; > 0.81 - can be accepted for clinical validation [15].

2.4. Joint evaluation of CT examinations by an automated CT image analysis system and radiologists

At the fifth stage of the dissertation research, we studied possible options for introducing automated CT image analysis systems as a method for diagnosing middle cerebral artery stroke in the clinical practice of a radiologist [88].

Program A was selected for clinical validation. The program was tested on two collections of images of CT. Radiologists with less than three years of experience and different skills in assessing ischemic stroke from the third stage of the dissertation research (described in Chapter 3) were also involved. Eleven months passed between the third and fifth stages; during the previous testing, the radiologists were not told the correct answers. Before starting this testing, radiologists were familiarized with the diagnostic performance indicators of the automated CT analysis system used. The distribution of radiologists between groups 1 and 2 is described in Table 11.

Table 11 – Distribution of radiologists by groups

	With experience in RVC	No experience in RVC
1 group	4	-
2 group	-	3

A joint analysis of test collections of images used for testing by radiologists and an automatic analysis system was carried out with modes of two options (time interval between tests 2 months):

1. Primary interpretation of non-contrast CT images by a radiologist, followed by providing him with interpretation data from the automatic analysis system and the specialist’s possible adjustment of his answer.

2. Primary interpretation of non-contrast CT images by an automated CT analysis system and subsequent assessment, taking into account the program interpretation data, by a radiologist.

In the first option, specialists in the process of analyzing collections of images 1B filled out a form with answers, interpreting CT images as normal or pathological, as well as the presence of the affected side. Experts independently assessed each study and marked them in the presence of stroke as “1”, in the absence as “0”. After providing interpretation data to the automated CT analysis system, they could adjust their answer. Similarly, radiologists completed a form when evaluating CT scans from collections of images 2B to assess side of stroke, areas of early ischemic change, and final ASPECTS score. The ASPECT score was made by subtracting 1 point from 10 for each significant sign of early ischemic changes (loss gray/white matter differentiation, sulcal effacement, and hypoattenuation) in each region. Next, radiologists were informed of the results of the evaluation of the artificial intelligence algorithm (zones according to ASPECTS, final score and side of the lesion). Radiologists could accept or rejects AI findings (ASPECTS area).

In the second version of testing, carried out 2 months later, when assessing CT from collections of images 1C, radiologists were initially informed of the program interpretation data on the presence/absence of MCA stroke and the affected side. Next, the radiologists either agreed with the artificial intelligence, or offered their own version of interpretation, and also filled out the answer form again, marking studies in the presence of stroke as “1”, in the absence as “0”. Similarly, in the evaluation of the collections of images 2C, radiologists were provided with data from the system's assessment of stroke side, areas of early ischemic change, and final ASPECTS score. The radiologist could both agree with the interpretation of the software product and form his final opinion on the side of the lesion, the extent of involvement of the territory of the blood supply of the middle cerebral artery in the ischemic process according to ASPECTS (Table 12).

Table 12 – Stages of testing radiologists with correlation with the options of CT collections of images used, as well as the time interval of the study

Research stage		Collections of images	Time interval
Assessment in CT examinations by radiologists (n=21) with different experience and expertise in emergency medicine	Presence/absence of MCA stroke, side of stroke	1A	23.05-03.06.2022
	Side of stroke, final ASPECTS rating	2A	
Assessment in CT examinations by radiologists (n=7) with three years of experience and different expertise in emergency medicine	Presence/absence of MCA stroke, side of stroke	1B	03.05-05.05.2023
	Side of stroke, final ASPECTS rating	2B	
Assessment in CT examinations by radiologists (n=7) with three years of experience and different expertise in emergency medicine	Presence/absence of MCA stroke, side of stroke	1C	01.07-04.07.2023
	Side of stroke, final ASPECTS rating	2C	

The CT images was viewed using the RadiAnt DICOM Viewer program (with the ability to construct MPR image reformation), as a universal tool installed on any imaging computer. Images were viewed on HP EliteDisplay S340c monitors with a resolution of 3440x1440 megapixels, Windows 10 operating system.

Statistical analysis of the obtained data was carried out using IBM SPSS Statistics 19 and the Python programming language. According to the fact that the distribution of most quantitative parameters did not obey the law of normal distribution, nonparametric tests were not used. Given the lack of a significant difference between the results of the agreement coefficients of Cohen's kappa and Fleiss's kappa at the third stage of the study, the assessment was carried out only with the determination of Fleiss's kappa. Statistical analysis to determine agreement regarding ASPECTS ischemic stroke scoring included determination of the interrater agreement coefficient (Fleiss's kappa) [64]. The coefficients of interrater agreement (Cohen’s kappa and Fleiss’s kappa) were assessed as follows: slight agreement, 0.00 to 0.20; fair agreement, 0.21 to 0.40; moderate agreement, 0.41 to 0.60; substantial agreement, 0.61 to 0.80; or almost perfect agreement, 0.81 to 1.00.

CHAPTER 3. RESULTS OF CT IMAGING OF MIDDLE CEREBRAL ARTERY STROKE BY RADIOLOGISTS DEPENDING ON THEIR EXPERIENCE AND EXPERTISE

In order to study the diagnostic performance of radiologists and their interrater agreement regarding ASPECTS scores, two tests were conducted in the third stage of the dissertation research.

Twenty-one experts took part in the first test. Radiologists were initially divided into two groups: radiologists working in region vascular center (RVC), and their colleagues working in general hospitals (GH), who had little experience in diagnosing ischemic stroke in the most acute and acute stages. Within the group, experts were divided depending on experience in clinical practice: more than eight years, three years to less eight years, and less than three years - the ratio of RSC and SOP doctors was 4/3 in each group, divided by work experience [4].

Experts were asked to assess the presence/absence of stroke of ischemic type based on non-contrast CT images of the brain. Also, radiologists had to establish the presence of such signs as a hypoattenuation, sulcal effacement, loss gray/white differentiation, as well as hyperdense middle cerebral artery sign and "point" symptom [4].

The results of the first test assessing the diagnostic indicators of radiologists with different experience in clinical practice and expertise in urgent neuroradiology are presented in Table 13.

Table 13 – The effectiveness of identifying ischemic changes depending on their experience and expertise

Years of experience	Specialization	Sensitivity, %	Specificity, %	Accuracy, %
>8 years	RVC	90.5	97.0	93.8
	GH	85.3	79.3	82.3
3-8 years	RVC	92.5	95.0	91.3
	GH	83.3	86.7	85.0
<3 years	RVC	84.0	92.5	88.3
	GH	57.3	92.7	75.0

According to our study, there is a direct dependence of the degree of diagnostic strength in the detection of ischemic changes in the middle cerebral artery territory by computed tomography on the years of practice of radiologists and their experience in the regional vascular center. For radiologists from the GH, this correlation did not hold. It is worth noting that there is a jump in the number of correct interpretations of CT images among experts in each group with more than three years in clinical practice. Based on these statistical results, we can conclude that the degree of diagnostic strength of radiologists depends on experience and expertise in emergency medicine [4].

The level of over- and under-diagnosis among RVC radiologists more than three years of experience is lower than among RVC specialists with less than three years of experience and GH doctors in general. It is worth noting that all radiologists participating in the testing were more prone to overdiagnosis; this trend is more pronounced among GH specialists with less than three years of experience. Also, based on the results described above, the experience of specialists who do not have skills in emergency medicine does not affect their diagnostic effectiveness [4]. To assess the relationship between sensitivity and specificity, ROC curves were constructed to calculate the area under the curve (AUC) for all analyzed groups (Figure 16).

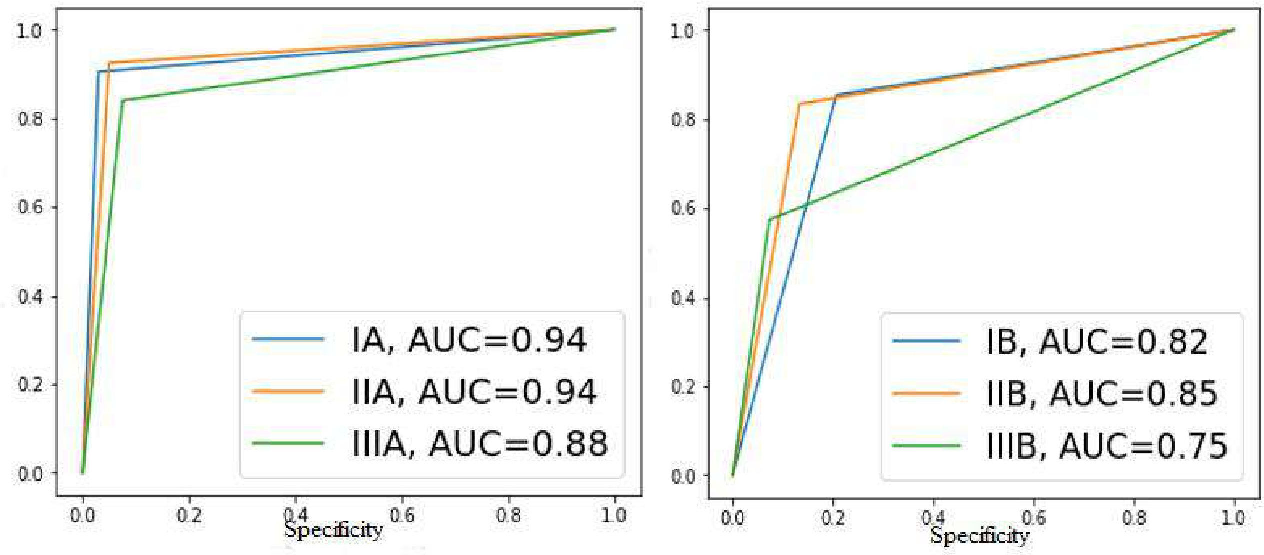

Figure 16 – ROC curve to assess the diagnostic efficiency of radiologists at the regional vascular center (A), and general hospitals (B) with different years of experience: I - more than 8 years, II - 3-8 years, III - less than three years

To assess the agreement between the study groups of specialists regarding CT signs of ischemic stroke in the MCA territory, Fleiss's kappa was calculated (Table 14).

Table 14 – Agreement between radiologists with different experience and expertise in emergency medicine regarding CT signs of ischemic stroke in the middle cerebral artery territory

<table>
<tr><td colspan="2">Years of experience</td><td colspan="2">> 8 years</td><td colspan="2">3- 8 years</td><td colspan="2">< 3 years</td></tr>
<tr><td colspan="2">Specialization</td><td>RVC</td><td>GH</td><td>RVC</td><td>GH</td><td>RVC</td><td>GH</td></tr>
<tr><td rowspan="4">Fleiss's Kappa</td><td rowspan="2">Hyperdense middle cerebral artery sign</td><td>0,750</td><td>0,406</td><td>0,697</td><td>0,631</td><td>0,591</td><td>0,095</td></tr>
<tr><td>РСЦ</td><td colspan="3">0,684</td><td>СОП</td><td colspan="2">0,318</td></tr>
<tr><td rowspan="2">Loss of grey-white differentiation</td><td>0,671</td><td>0,519</td><td>0,626</td><td>0,530</td><td>0,599</td><td>0,114</td></tr>
<tr><td>РСЦ</td><td colspan="3">0,629</td><td>СОП</td><td colspan="2">0,378</td></tr>
</table>

Fleiss's Kappa	Sulcal effacement	0,689	0,254	0,416	0,454	0,433	-0,020
		РСЦ	0,525		СОП	0,244	
	Hypoattenuation	0,535	0,495	0,493	0,530	0,479	0,077
		РСЦ	0,529		СОП	0,449	

The overall agreement of the RVC specialists on all criteria ranged from moderate to substantial, the agreement between the GH doctors was lower - from slight to moderate.

Interrater agreement regarding the hyperdense middle cerebral artery sign was substantial among RVC radiologists with more than three years of experience (with higher results for more than eight years in clinical practice) and moderate among experts with more than three years of experience. Among GH radiologists, interrater agreement was greatest among radiologists with three years to less eight years of experience and was substantial. Among doctors with more than eight years of experience, agreement was moderate, and among doctors with less than three years of experience, it was slight [4].

Agreement in groups stratified by experience was significant among experts with more than three years of experience, both from the RVC and from the GH, and fair among experts with less than three years of experience.

Interrater agreement regarding the sign of loss gray and white differentiation was similar to the hyperdense middle cerebral artery sign, except that radiologists without expertise in emergency medicine and three years to less eight years of experience showed moderate agreement regarding this sign (compared to a substantial level regarding the hyperdense middle cerebral artery sign) [4].

Regarding the sign of hypoattenuation, the agreement among all specialists, with the exception of doctors without expertise in emergency neuroradiology and less than three years of experience, was moderate. Among GH specialists with less than three years of experience, consistency was slight. It is worth paying attention to the substantial agreement among all RVC specialists on this sign and moderate agreement among GH doctors, which indicates the influence of only relevant experience on the correct interpretation of specific signs [4].

The sulcal effacement was the most ambiguous sign when assessing interrater agreement. Substantial agreement was observed only among RSC specialists with more than eight years of experience. Moderate agreement was recorded between doctors with three years to less eight years of experience in both groups and among RVC specialists with less than three years of experience. Fair agreement was found among GH doctors with more than eight years in clinical practice; agreement among their colleagues with less than three years of experience was random (<0) [4].

Based on the results of the first testing, it can be concluded that doctors working in RVC have higher level of diagnostic efficiency in identifying middle cerebral artery strokes and greater agreement in assessing early ischemic changes, compared with colleagues from the GH.

The further course of the thesis work was aimed at determining inter-expert variability between doctors directly specializing in urgent neuroradiology. In the second test, 50 anonymized CT were presented to 15 radiologists with one to ten years in clinical practice in an emergency hospital setting. The study involved five doctors with more than eight years in clinical practice (Group I), five doctors with three years to less eight years of experience (Group II) and five doctors with less than three years in clinical practice (Group III). Radiologists independently assessed each examination and determined the side of the lesion and the ASPECTS score. The results were dichotomized ASPECTS (≤ 7 and > 7 and ≤ 6 and > 6) for further statistical analysis (Table 15).

Table 15 – Statistical analysis of inter-rater agreement (Cohen's k and Fleiss's k) in assessed according to dichotomized ASPECTS (≤6 and >6) by RVC radiologists

ASPECTS 6				Sight	Fair	Moderate	Substantial	Almost perfect agreement
Years of experience	Fleiss's Kappa	min k – max k (Cohen's kappa)	μ (Cohen's kappa)	0–0,2	0,21–0,4	0,41–0,6	0,61–0,8	0,81–1
> 8 years	0,366	0,125–0,633	0,36	20 %	40 %	30 %	10 %	0
3- 8 years	0,452	0,31–0,625	0,46	0	30 %	60 %	10 %	0
< 3 years	0,462	0,31–0,646	0,46	0	30 %	60 %	10 %	0
Overall coefficient of interrater agreement (Cohen's kappa)								
	0,391	0,011–0,789	0,389	15,2 %	34 %	40 %	10,5 %	0

The coefficient of inter-rater agreement (Cohen's kappa, k) with dichotomized ASPECTS (≤ 6 and > 6) in group I ranged between 0.125–0.633 (it was more heterogeneous than in groups II and III) and to a greater extent (40%) was fair, and also included the percentage (20%) of Cohen's kappa results corresponding to slight agreement (such low results were not recorded in groups II and III). In group II, Cohen's kappa ranged between 0.31–0.625, averaged 0.46, and was more consistent with moderate agreement. The results of group III (0.31–0.646) were almost identical to the results of group II [5]. Also, Cohen's kappa between raters ranged from 0.011 (slight agreement) to 0.789 (substantial agreement). Among all specialists, agreement was predominantly either moderate or fair [5].

The coefficient of inter-rater agreement (Cohen's kappa, k) was higher among groups II and III and on average corresponded to a value of 0.48, which is a moderate indicator. The worst agreement was observed among groups I and III, Kohn's kappa was 0.26, which is fair agreement (also including a large percentage of weak interrater agreement) (Table 16) [5].

Table 16 – Interrater agreement (Cohen's kappa) among groups of specialists to dichotomized ASPECTS (≤6 and >6)

ASPECTS 6			Sight	Fair	Moderate	Substantial	Almost perfect agreement
Radiologists	min k – max k	μ	0–0,2	0,21–0,4	0,41–0,6	0,61–0,8	0,81–1
Between I and II groups							
1	0,107–0,789	0,47	20 %	20 %	20 %	40 %	0
2	0,12–0,559	0,34	20 %	40 %	40 %	0	0
3	0,312–0,545	0,41	0	40 %	60 %	0	0
4	0,011–0,286	0,13	80 %	20 %	0	0	0
5	0,336–0,719	0,47	0	40 %	40 %	20 %	0
		0,37					
Between II and III groups							
6	0,351–0,545	0,46	0	20 %	80 %	0	0
7	0,429–0,694	0,54	0	0	60 %	40 %	0
8	0,483–0,634	0,52	0	0	80 %	20 %	0
9	0,351–0,694	0,47	0	60 %	20 %	20 %	0
10	0,247–0,582	0,4	0	40 %	60 %	0	0
		0,48					

Between I and III groups							
11	0,043–0,428	0,2	40 %	40 %	20 %	0	0
12	0,093–0,321	0,2	40 %	60 %	0	0	0
13	0,096–0,307	0,19	40 %	60 %	0	0	0
14	0,163–0,662	0,4	20 %	20 %	40 %	20 %	0
15	0,118–0,513	0,32	20 %	60 %	20 %	0	0
		0,26					

I – radiologists with more than eight years in clinical practice; II – radiologists with three years to less eight years of experience; III – radiologists with less than three years of experience

Fleiss' kappa with dichotomized ASPECTS (≤6 and >6) among the three groups was 0.391. In group I, Fleiss's kappa was fair (0.366), in groups II and III it was moderate (0.452 and 0.462, respectively) [5].

It should be noted that minimal differences in inter-rater agreement were identified (Cohen's kappa and Fleiss's kappa), and their results were in the same range, demonstrating an insignificant level of agreement among themselves when assessed on the ASPECT score by radiologists with more than eight years of experience in urgent neuroradiology and moderate level of agreement between doctors with less than eight years of experience [5].

The coefficient of interrater agreement (Cohen's kappa) with dichotomized ASPECTS (≤ 7 and > 7) in group I ranged between 0.294–0.588 and to a greater extent (60%) was moderate, and also included the percentage (40%) of Cohen's kappa results, corresponding to fair agreement (which is higher than the results for dichotomized ASPECTS (≤ 6 and > 6). In group II, Cohen's kappa ranged between 0.27–0.518 and was more consistent with fair agreement. The result of group III (0.271–0.64) was the highest, amounting to 0.469 and corresponding to moderate agreement (Table 17) [5].

Table 17 – Statistical analysis of inter-rater agreement (Cohen's k and Fleiss's k) in assessed according to dichotomized ASPECTS (≤7 and >7) by RVC radiologists

ASPECTS 7				Sight	Fair	Moderate	Substantial	Almost perfect agreement
Years of experience	Fleiss's Kappa	min k – max k (Cohen's kappa)	μ (Cohen's kappa)	0–0,2	0,21–0,4	0,41–0,6	0,61–0,8	0,81–1
> 8 years	0,439	0,294–0,588	0,446	0	40 %	60 %	0	0
3- 8 years	0,384	0,27–0,518	0,392	0	50 %	50 %	0	0
<3 years	0,466	0,271–0,64	0,469	0	30 %	50 %	20 %	0
Overall coefficient of interrater agreement (Cohen's kappa)								
	0,376	0,007–0,8	0,39	12 %	40 %	46 %	2 %	0

Interrater Cohen's kappa ranged from 0.007 (slight agreement) to 0.8 (substantial agreement). Among all experts, agreement was either moderate or insignificant almost equally [5].

The coefficient of interrater agreement (Cohen's kappa) was higher among groups II and III and corresponded to a value of 0.45, which indicates moderate agreement. The lowest agreement rates were observed among groups I and III, Cohen's kappa was 0.31, which is fair agreement (this result did not differ significantly from Cohen's k between groups I and II - 0.34) (Table 18) [5].

Table 18 – Interrater agreement (Cohen's kappa) among groups of specialists to dichotomized ASPECTS (≤7 and >7)

ASPECTS 7			Sight	Fair	Moderate	Substantial	Almost perfect agreement
Radiologist	min k – max k (Cohen's kappa)	μ	0–0,2	0,21–0,4	0,41–0,6	0,61–0,8	0,81–1
Between I and II groups							
1	0,331–0,504	0,42	0	40 %	60 %	0	0
2	0,154–0,485	0,32	20 %	60 %	20 %	0	0
3	0,331–0,541	0,40	0	60 %	40 %	0	0
4	0,128–0,412	0,20	80 %	0	20 %	0	0
5	0,326–0,469	0,38	0	80 %	20 %	0	0
		0,34					
Between II and III groups							
6	0,451–0,6	0,68	0	0	100 %	0	0
7	0,217–0,8	0,43	0	20 %	80 %	0	0
8	0,313–0,595	0,48	0	20 %	80 %	0	0
9	0,131–0,595	0,32	20 %	40 %	40 %	0	0
10	0,124–0,504	0,37	0	40 %	60 %	0	0
		0,45					

Between I and III groups							
11	0,007–0,244	0,14	80 %	20 %	0	0	0
12	0,118–0,388	0,29	20 %	80 %	0	0	0
13	0,099–0,374	0,27	20 %	80 %	0	0	0
14	0,231–0,587	0,41	0	40 %	60 %	0	0
15	0,2–0,6	0,44	20 %	0	80 %	0	0
		0,31					

I – radiologists with more than eight years in clinical practice; II – radiologists with three years to less eight years of experience; III – radiologists with less than three years of experience

Fleiss' kappa with dichotomized ASPECTS (≤ 7 and > 7) among the three groups was 0.376. In groups I and III, Fleiss' kappa was moderate (0.439 and 0.466, respectively), and in group II it was fair (0.384). There were also no significant differences in the numerical indicators of inter-rater agreement (Cohen's k and Fleiss's k) with ASPECTS (≤ 7 and > 7), so in the future we analyzed inter-rater agreement only using Fleiss's kappa [5]. Cohen's kappa and Fleiss's kappa interrater agreement coefficients were higher with dichotomized ASPECTS (≤ 6 and > 6) [4].

The coefficient of inter-rater agreement (Cohen's kappa) with ASPECTS (≤ 6 and > 6) prevailed in groups II and III (0.46), with dichotomous division (≤ 7 and > 7) it was higher in group III of experts, which indicates a trend among specialists with less than three years of experience, ASPECT scores are overestimated [5].

The agreement coefficient (Cohen's k) was higher when comparing the results of groups II and III (both dichotomized ASPECTS [≤ 6 and > 6] and [≤ 7 and > 7]), while the agreement between groups I and III was the smallest for both dichotomized ASPECTS [5]. These results show that the consistency of expert assessments of emergency neuroradiology doctors does not depend on the specialist's experience, and in the conditions of this study, radiologists with extensive experience showed quite heterogeneous results, but definitely lower results than experts with the least length of service [5]. Thus, the influence of work experience in the RVC of emergency neuroradiology radiologists on both the diagnostic efficiency in detecting middle cerebral artery stroke and on the consistency with respect to most CT signs of ischemic stroke was revealed. However, the experience of specialists with expert qualifications does not correlate with the level of inter-rater agreement regarding the ASPECTS assessment, which reflects the low reproducibility of this scale.

CHAPTER 4. RESULTS OF TESTING AUTOMATED CT IMAGE ANALYSIS SYSTEMS AND THEIR EFFECT ON THE MEDICAL DECISION MAKING

At the fourth stage of the dissertation research, we selected for testing three software programs based on convolutional neural networks, positioning themselves as a software device for computer-aided detection, used by a radiologist to interpret the nature and distribution of pathological changes in brain tissue by assessing the ASPECT scale based on data CT images.

Below are given the testing selection criteria:

1. Availability of a test online access;
2. The description of the software program indicates the function of rating areas with early ischemic changes in the acute period (ASPECTS);
3. The declared accuracy in diagnosing middle cerebral artery stroke is more than 75% (according to the third stage of the dissertation research, as the minimum accuracy among the tested doctors).

According to the inclusion criteria, for software program were selected (two systems of domestic developers and one foreign system).

Since the research was aimed at the general assessment of diagnostic indicators of currently available systems and not at the assessment of a particular product, all programs in the research were disguised as A, B, C, D.

Using the method of analytical validation, artificial intelligence software was tested on a reference data set (imaging sampling №1), prepared in accordance with the scientific task, and also registered in accordance with the regulations [7, 15]. The CT image data set was similar to that of the third phase of the study, which assessed the diagnostic performance of radiologists with varying level of experience in detecting middle cerebral artery territory stroke.

During statistical analysis, all metrics were assessed in the range of 0–1: <0.6 - unsuitable; 0.61–0.8—needs improvement; >0.81—can be accepted for clinical validation [15]. The results of the analysis of data from sample 1 are presented in Table 19.

Table 19 – Comparative characteristics of automated CT image analysis systems

Indicator	Artificial intelligence software A	Artificial intelligence software B	Artificial intelligence software C
Number of TruePositive Results	44	30	22
Number of FalsePositive Results	27	5	9
Number of FalseNegative Results	6	20	18
Number of TrueNegative Results	23	45	35
Not defined	0	0	16
Sensitivity	0,88	0,60	0,55
Specificity	0,46	0,90	0,80
Likelihood Ratio of a Positive Test	1,63	6,00	2,69
Likelihood Ratio of a Negative Test	0,26	0,44	0,57
Accuracy	0,67	0,75	0,68

In the thesis work, all selected automatic analysis systems had an accuracy in the range from 0.67 to 0.75, which, according to clinical recommendations [15], indicates the need for their further improvement. The highest accuracy, a parameter that determines the number of correctly identified judgments, was demonstrated by program B (0.75). The specificity of program B (0.90) was also higher compared to the results of other programs (0.46; 0.80) [2].

Program A demonstrated a high rate of identifying true positive results (44), while the specificity of this program was quite low (0.46); the program identified 27 false positive results, which is a sign of overdiagnosis and overtraining. The accuracy of this program is lower than the accuracy of the results of other programs. At the same time, the program has demonstrated high sensitivity, which reduces the risk of underdiagnosis [2].

Program B had the highest specificity rates (0.90) with low sensitivity (0.60), which will allow, once it is refined and the required total parameters of diagnostic accuracy are achieved, to recommend this algorithm as a method of primary assessment with rapid triage of patients with suspected stroke [2].

A limitation in assessing the diagnostic effectiveness of one of the algorithms (C) was its inability to interpret 16 CT of the brain from the presented independent data set due to an unavoidable technical error when loading radiation images. In turn, the accuracy of program C was slightly higher than the results of program A; it demonstrated a large number of false negative results, which indicates its tendency to underdiagnosis [2].

Low sensitivity of programs is a parameter showing low efficiency in identifying patients with stroke, since some of the results will be classified as negative. Thus, algorithms B and C require further refinement (with further training) in order to improve the efficiency of the analysis [2].

The likelihood ratio of a positive result shows how many times more likely patients with stroke are likely to get a positive result than healthy patients. Program B achieved the highest results, which is confirmed by high specificity indicators. In terms of likelihood of a negative result, program C showed the highest results, which is a result of the low sensitivity of the program [2].

To assess the relationship between sensitivity and specificity, ROC curves were constructed to calculate the area under the curve (AUC) for all analyzed groups (Figure 17). Analysis of the indicators revealed that only the results obtained using program B give values (AUC = 0.75) close to the recommended parameters, according to clinical guidelines; the predictions of other programs are lower [2].

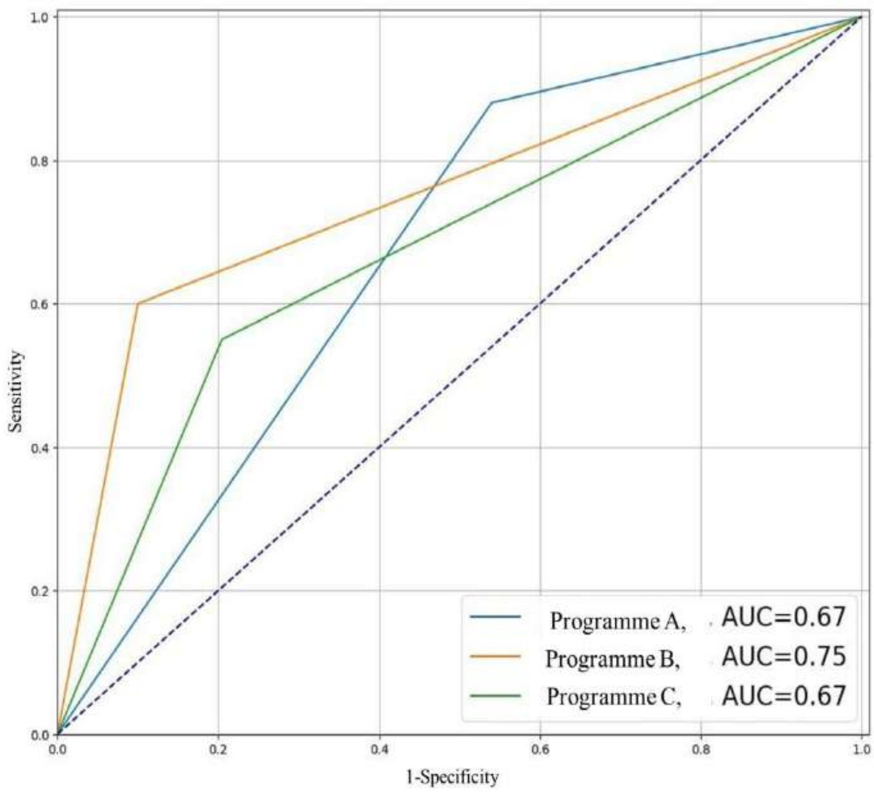

Figure 17 – ROC curves to assess the diagnostic effectiveness of all automatic analysis systems

According to the results of the study, none of the reviewed systems for automatic detection of stroke based on CT images reached the threshold accuracy values required for further clinical validation, which indicates the need for their further refinement [2].

The accuracy of all automatic analysis systems was significantly lower than the results of doctors with expert qualifications (0.94), and also lower than the results of the assessment of doctors with less than 3 years of experience (0.88) from the RVC tested in the third stage of the dissertation research (Table 20).

Table 20 – Comparative characteristics of indicators of diagnostic efficiency in joint testing of radiologists and automatic analysis systems

	A	B	C	D*	F*	E*
Number of TruePositive Results	44	30	22	45	42	29
Number of FalsePositive Results	27	5	9	2	4	4
Number of FalseNegative Results	6	20	18	5	8	21
Number of TrueNegative Results	23	45	35	48	46	46
Not defined	0	0	16	0	0	0
Sensitivity	0.88	0.60	0.55	0.90	0.84	0.57
Sensitivity, %	88.00	60.00	55.00	90.50	84.00	57.30
Specificity	0.46	0.90	0.80	0.97	0.93	0.93
Specificity, %	46.00	90.00	80.00	97.00	92.50	92.70
Likelihood Ratio of a Positive Test	1.63	6.00	2.69	30.17	11.20	7.82
Likelihood Ratio of a Negative Test	0.26	0.44	0.57	0.10	0.17	0.46
Accuracy	0.67	0.75	0.68	0.94	0.88	0.75
Accuracy, %	67.00	75.00	68.00	93.80	88.30	75.00

D - RVC radiologists with more than eight years in clinical practice, F - RVC radiologists with less than three years of experience, E - GH radiologists with less than three years of experience

One of the algorithms (B) had comparative accuracy with radiologists with less than 3 years in clinical practice and no experience in emergency neuroradiology (0.75). The highest likelihood ratio of a positive result was also found among doctors with expert qualifications (with 8 years of experience in the RVC), while this indicator was 5 times higher than the best result for the programs (6.00 versus 30.17). The likelihood values of a positive result for groups of doctors with less than three years of experience, regardless of experience in the RVC, are also higher (11.20 with experience, 7.82 without experience) than for programs (A-1.63, B-6.00, C-2.69) [2].

To assess the relationship between sensitivity and specificity, ROC curves were constructed to calculate the area under the curve (AUC) for all analyzed groups. Analysis of the indicators revealed that only the results obtained using program B give values (AUC = 0.75) close to those of doctors without experience in emergency medicine (AUC = 0.75); the accuracy of predictions of other programs is lower (Figure 18).

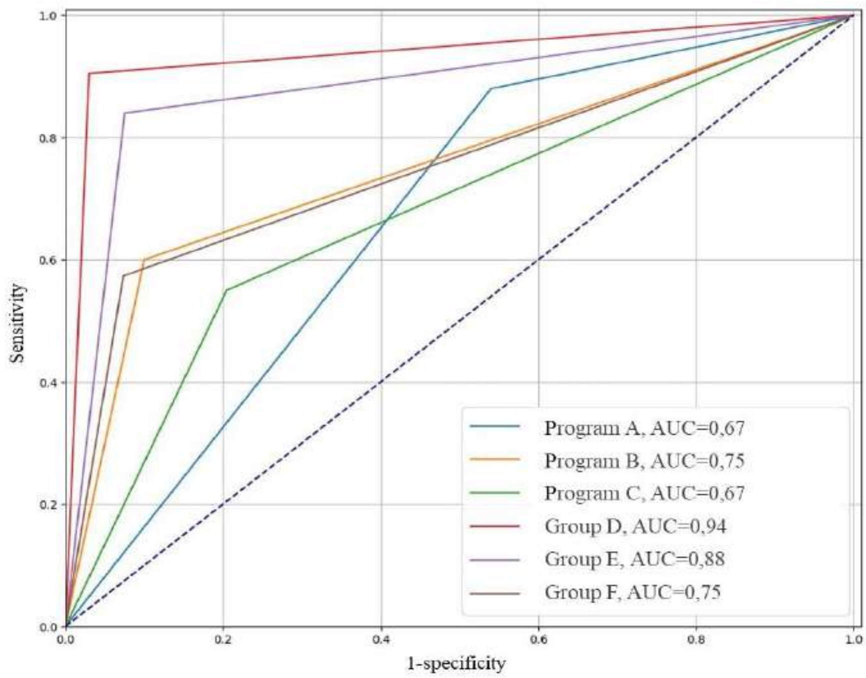

Figure 18 – ROC curves to assess the diagnostic efficiency in joint testing of radiologists (RVC radiologists with more than eight years of experience; RVC radiologists with less than three years of experience; GH radiologists with less than three years of experience) and automatic analysis systems

However, it should be noted that despite the high rate of false-positive responses, program A is practically not inferior to the results of doctors with expert qualifications regarding true positive and false negative responses, which allows us to judge its low level of underdiagnosis. For doctors with less than three years of experience and no experience in the RVC, compared with program A, false negative responses were 3.5 times more, and true positive responses were 1.5 times less. Regarding these indicators, doctors with more than three years of experience from the RVC were also slightly inferior to program A. This suggests that potentially, program A may have a positive effect in reducing missed pathology in young specialists with more than three years of experience, regardless of experience in assessing stroke [2].

Thus, at present, the results of systems for automatically detecting ischemic changes on CT images are not comparable with the average data of the results of radiologists with experience in assessing stroke in terms of diagnostic efficiency.

Most software products show high specificity rates and low sensitivity rates, which indicates infrequent cases of overdiagnosis and a large number of cases of underdiagnosis. For a more reliable understanding of the diagnostic capabilities of these software products, clinical trials should be continued using both the method of analytical validation on various samples and the method of clinical validation.

CHAPTER 5. IMPACT OF IMPLEMENTING AUTOMATED CT IMAGE ANALYSIS SYSTEMS ON THE MEDICAL DECISION MAKING

At the fifth stage of the thesis work, possible options for introducing automated CT image analysis system were studied as a method of supporting clinical decision-making in the diagnosis of middle cerebral artery stroke.

According to Obuchowski N.A. et al. [88], there are four options for using AI algorithms in CT imaging: (1) a first - reader mode, where the AI algorithm first evaluates the image and then the results of this assessment are reviewed by a human, (2) a second- reader mode, where the AI algorithm is applied after how a human accepts or rejects AI findings, (3) a triage mode, where the algorithm prioritized worklist for reader according to suspiciousness of ischemia on brain CT, (4) a pre-screening mode, where the prescreening AI writes report for negative brain CT scan. Reader gives interpretation for remaining cases. In order to study the implementation of the first two options for using automated CT image analysis systems, we conducted two options for joint testing of doctors with less than three years of experience and an artificial intelligence algorithm selected at the fourth stage of the thesis work. The subject of study was program A, since despite the insufficient level of accuracy, this system showed results that indicate a low level of underdiagnosis of ischemic changes (sensitivity 88%, true positive responses - 44 out of 100, false positive responses 27 out of 100). Analysis of non-contrast CT images by this program was performed with the function of automatic assessment of areas with ischemia changes in the acute period (i.e., assessment on the ASPECT score). Seven radiologists with less than 3 years of experience were also included in the study. Doctors were divided into two groups according to their experience in CT diagnostics of ischemic stroke: specialists from the RVC (N=4) and their colleagues from the GH (N=3).

In the first testing option, in the process of studying the implementation of the second reader mode, collections of images 1B and 2B were used. On the CT collections of images 1B, doctors were asked to assess the presence/absence of stroke, then the doctors were informed of the results of the evaluation of the artificial intelligence algorithm: presence (ASPECTS 9 and lower)/absence (ASPECTS 10) of stroke, as well as the side of the lesion. Doctors could either agree with artificial intelligence or not change their decision. On CT images of collections of images 2B, specialists determined the side of the stroke and the prevalence of acute ischemic changes according to ASPECTS. Next, the radiologists were informed of the results of the evaluation of the artificial intelligence algorithm: the ASPECT score indicating the affected areas, as well as the side of the lesion. Doctors could accept or reject system’s findings. The results of the first phase of testing are presented in Tables 21 and 22.

Table 21 – Indicators of diagnostic efficiency of radiologists with different experience in urgent neuroradiology and less than three years of experience in the first option of testing by collections of images 1B

			Number of TruePositive Results	Number of FalsePositive Results	Number of FalseNegative Results	Number of TrueNegative Results	Sensitivity, %	Specificity, %	Accuracy, %
Artificial intelligence software			44	27	6	23	88.00	46.00	67.00
RVC Radiologists	1	Initial	49	9	1	41	98.00	82.00	90.00
		First phase	50	9	0	41	100.00	82.00	91.00
	2	Initial	49	10	1	40	98.00	80.00	89.00
		First phase	49	12	1	38	98.00	76.00	87.00
	3	Initial	50	12	0	38	100.00	76.00	88.00
		First phase	50	13	0	37	100.00	74.00	87.00
	4	Initial	50	6	0	44	100.00	88.00	94.00
		First phase	50	7	0	43	100.00	86.00	93.00

GH Radiologists	5	Initial	47	8	3	42	94.00	84.00	89.00
		First phase	49	7	1	43	98.00	86.00	92.00
	6	Initial	48	31	2	19	96.00	38.00	67.00
		First phase	49	28	1	22	98.00	44.00	71.00
	7	Initial	36	17	14	33	72.00	66.00	69.00
		First phase	41	17	9	33	82.00	66.00	74.00

Table 22 – Average diagnostic performance indicators of radiologists with different levels of experience in urgent neuroradiology and less than three years of experience in the first phase of testing on collections of images 1B

		Number of TruePositive Results	Number of FalsePositive Results	Number of FalseNegative Results	Number of TrueNegative Results	Sensitivity, %	Specificity, %	Accuracy, %
Artificial intelligence software		44.00	27.00	6.00	23.00	88.00	46.00	67.00
RVC Radiologists	Initial	49.50	9.25	0.50	40.75	99.00	81.50	90.25
	First phase	49.75	10.25	0.25	39.75	99.50	79.50	89.50
GH Radiologists	Initial	43. (6)	18. (6)	6. (3)	31. (3)	67. (3)	62. (6)	75.00
	First phase	46. (3)	17. (3)	3. (6)	32. (6)	92. (6)	65. (3)	79.00

Despite the performance of individual radiologists, the overall values of diagnostic efficiency (accuracy, specificity) among RVC doctors in the first version of testing decreased to a greater extent due to specificity (from 81.5% to 79.5%) and an increase in the number of false positive answers (from 9. 25 to 10.25), contributing to an increase in overdiagnosis in this group of experts. However, in the group of specialists without experience in the RVC, there was an improvement in efficiency indicators (accuracy, sensitivity and specificity) with the first version of testing, mainly due to an increase in sensitivity (from 67. (3) % to 92. (6) %) and a reduction in false negatives answers (from 6. (3) to 3. (6)), that is, with a decrease in the level of underdiagnosis. These results indicate the positive impact of artificial intelligence on the diagnostic performance of doctors without experience in emergency neuroradiology when applying the second reader mode. The results of the first joint testing of radiologists and the automated CT image analysis program on a sampling of CT images 2B are presented in Table 23.

Table 23 – Results of the initial assessment and first joint testing of radiologists with different levels of experience in urgent neuroradiology and less than three years of experience with an automated CT image analysis system

<table>
<tr><td colspan="2" rowspan="3"></td><td colspan="4">Second-Reader Mode</td></tr>
<tr><td colspan="2">Initial assessment</td><td colspan="2">First joint testing</td></tr>
<tr><td>≤ 7 и> 7</td><td>≤ 6 и> 6</td><td>≤ 7 и> 7</td><td>≤ 6 и> 6</td></tr>
<tr><td rowspan="2">Interrater agreement</td><td>RVC Radiologists</td><td>0.593</td><td>0.541</td><td>0.598</td><td>0.633</td></tr>
<tr><td>GH Radiologists</td><td>0.234</td><td>0.260</td><td>0.365</td><td>0.304</td></tr>
<tr><td colspan="2">General interobserver agreement</td><td>0.425</td><td>0.409</td><td>0.530</td><td>0.464</td></tr>
<tr><td rowspan="3">Agreement with artificial intelligence software</td><td>RVC Radiologists</td><td>0.584</td><td>0.501</td><td>0.625</td><td>0.592</td></tr>
<tr><td>GH Radiologists</td><td>0.300</td><td>0.324</td><td>0.480</td><td>0.415</td></tr>
<tr><td>Interobserver agreement</td><td>0.440</td><td>0.411</td><td>0.556</td><td>0.480</td></tr>
</table>

At the initial assessment, both when dichotomized ASPECTS ≤7 and > 7, and ≤ 6 and > 6, the agreement of the RVS specialists within the group was moderate. When applying the second reader mode, no significant changes in Fleiss's kappa scores were detected, which indicates that there is no significant influence of the clinical decision support system on the inter-rater agreement of RVC doctors. A similar trend in agreement was observed among GH doctors, whose Fleiss kappa scores indicated fair agreement between specialists at the initial assessment and after the first testing on collections of images 2B. It is worth noting that agreement with the program of RVC specialists during the initial assessment and during the first version of testing with dichotomized ASPECTS ≤ 6 and > 6 was inferior to agreement with each other, however, the Fleiss kappa indicators for dichotomized ASPECTS ≤ 7 and > 7 were higher when applying the second reader mode. Also, during the initial assessment, RVC specialists gave a score of ≤ 7 in 50.5% of cases; during the first testing option, this figure decreased to 42.5%. Similarly, RVC doctors, before providing interpretation of the artificial intelligence algorithm, in 64.5% of cases assessed the prevalence of ischemic changes in the territory of the middle cerebral artery as ≤ 6 points according to ASPECTS, however, when applying the second reading model, this figure decreased to 57%. These results indicate that there were no significant changes in the agreement between specialists when implementing the second reader mode, but at the same time, their tendency to reduce the ASPECTS score after providing interpretation data of the algorithm.

We also assessed the agreement of radiologists during the initial assessment and the first testing on collections of images 2B regarding each of the ASPECTS areas. The results are presented in Table 24.

Table 24 – Indicators of interrater agreement among radiologists with varying level experience in urgent neuroradiology and less than three years of experience at the initial assessment of ASPECTS areas and after the first joint testing on a set of 2B CT images

	Initial assessment										Second joint testing (First-Reader Mode)									
	C	IC	L	I	M1	M2	M3	M4	M5	M6	C	IC	L	I	M1	M2	M3	M4	M5	M6
Within the group																				
RVC Radiologists	0.520	0.165	0.480	0.504	0.317	0.525	0.478	0.652	0.518	0.472	0.648	0.227	0.557	0.643	0.393	0.631	0.654	0.696	0.534	0.493
GH Radiologists	0.520	0.070	0.141	0.109	0.250	0.010	0.138	0.183	0.134	0.157	0.534	0.041	0.400	0.217	0.410	0.294	0.442	0.428	0.732	0.495
Between all groups	0.497	0.090	0.321	0.238	0.312	0.281	0.325	0.470	0.335	0.340	0.566	0.204	0.443	0.379	0.428	0.466	0.549	0.594	0.523	0.483
Agreement with the program																				
RVC radiologists + program	0.532	0.191	0.440	0.431	0.335	0.484	0.480	0.632	0.516	0.472	0.389	0.091	0.455	0.657	0.297	0.167	0.294	0.500	0.420	0.269
GH radiologists + program	0.458	0.013	0.265	0.158	0.269	0.057	0.192	0.225	0.187	0.604	0.566	0.175	0.519	0.342	0.491	0.412	0.480	0.483	0.673	0.567
Between all groups and the program	0.496	0.112	0.337	0.248	0.317	0.284	0.342	0.467	0.350	0.371	0.396	0.114	0.409	0.402	0.347	0.415	0.322	0.475	0.446	0.325

Among RVC radiologists, during the first testing on collections of images 2B, agreement regarding each of the ASPECTS areas did not change significantly, with the exception of zone M3 (from moderate to substantial agreement). A more dynamic picture was observed among GH radiologists: agreement increased most noticeably relative to areas M5 (from slight to substantial), L, M3, M4 (from slight to moderate agreement).

It is also worth noting the general increase in Fleiss's kappa values relative to the ASPECTS areas of the basal nuclear region and insula (compared to the “hemispheric” zones) after the first stage of testing on a collections of images 2B from KMC doctors, but in the range of moderate agreement. The overall agreement between GH doctors changed to a greater extent regarding the “hemispheric” areas (M1-M6) from slight to moderate.

In the second version of testing, carried out 2 months later, the influence of the first reader mode was assessed on collections of images 1C and 2C. Doctors were immediately provided with data from the interpretation of CT images by the algorithm, after which they could either agree with the automated CT image analysis system or offer their own assessment option.

The results of the second joint testing on sampling of 1C and 2C CT images of the brain by doctors with less than three years of experience and an artificial intelligence algorithm are presented in Tables 25 and 26.

Table 25 – Indicators of diagnostic efficiency of radiologists with different experience in urgent neuroradiology and less than three years of experience in the second option of testing by collections of images 1C

			Number of TruePositive Results	Number of FalsePositive Results	Number of FalseNegative Results	Number of TrueNegative Results	Sensitivity, %	Specificity, %	Accuracy, %
Artificial intelligence software			44	27	6	23	88.00	46.00	67.00
RVC Radiologists	1	Initial	49	9	1	41	98.00	82.00	90.00
		Second phase	50	4	0	46	100.00	92.00	96.00
	2	Initial	49	10	1	40	98.00	80.00	89.00
		Second phase	50	6	0	44	100.00	88.00	94.00
	3	Initial	50	12	0	38	100.00	76.00	88.00
		Second phase	50	5	0	45	100.00	90.00	95.00
	4	Initial	50	6	0	44	100.00	88.00	94.00
		Second phase	49	7	1	43	98.00	86.00	92.00
GH Radiologists	5	Initial	47	8	3	42	94.00	84.00	89.00
		Second phase	48	3	2	47	96.00	94.00	95.00
	6	Initial	48	31	2	19	96.00	38.00	67.00
		Second phase	48	16	2	34	96.00	68.00	82.00
	7	Initial	36	17	14	33	72.00	66.00	69.00
		Second phase	48	8	2	42	96.00	84.00	90.00

Table 26 – Average diagnostic performance indicators of radiologists with different levels of experience in urgent neuroradiology and less than three years of experience in the second phase of testing on collections of images 1C

		Number of TruePositive Results	Number of FalsePositive Results	Number of FalseNegative Results	Number of TrueNegative Results	Sensitivity, %	Specificity, %	Accuracy, %
Artificial intelligence software		44.00	27.00	6.00	23.00	88.00	46.00	67.00
RVC Radiologists	Initial	49.50	9.25	0.50	40.75	99.00	81.50	90.25
	Second phase	49.75	5.50	0.25	44.50	99.50	89.00	94.25
GH Radiologists	Initial	43. (6)	18. (6)	6. (3)	31. (3)	67. (3)	62. (6)	75.00
	Second phase	48.00	9.00	2.00	41.00	96.00	82.00	89.00

Contrary to the first testing on sampling 1B, with the second option (on sampling 1C), the average diagnostic performance indicators (accuracy, sensitivity and specificity) of RVC doctors increased to a greater extent due to accuracy (from 90.2% to 94.25%) and specificity (from 81.5% to 89%) with a reduction in the number of false positive (from 18. (6) to 9) and false negative (from 6. (3) to 2) responses, and a decrease in the level of under- and overdiagnosis. A similar trend was also observed among specialists without experience in the RVC. In this group of specialists, there was a pronounced positive dynamics of average efficiency values with an increase in sensitivity values (from 67. (3) % to 96%), specificity (from 62. (6) % to 82%) and accuracy (from 75% to 89). %), with a reduction in the number of false positives (from 18. (6) to 9) and false negatives (from 6. (3) to 2). The results of the second testing on the sampling 1C indicate an improvement in the interpretation of CT images of patients with suspected stroke, both among doctors with and without experience in the RVC, when using the first reader mode.

The results of the second testing on collections of images 2C are presented in Table 27.

Table 27 – Results of initial assessment and second joint testing of radiologists with different levels of experience in urgent neuroradiology and less than three years of experience with the automated CT image analysis program

		First- Reader Mode			
		Initial assessment		Second joint testing	
		≤ 7 и> 7	≤ 6 и> 6	≤ 7 и> 7	≤ 6 и> 6
Interrater agreement	RVC Radiologists	0.593	0.541	0.734	0.739
	GH Radiologists	0.234	0.260	0.573	0.600
General interobserver agreement		0.425	0.409	0.669	0.657
Agreement with artificial intelligence software	RVC Radiologists	0.584	0.501	0.734	0.743
	GH Radiologists	0.300	0.324	0.658	0.646
	Between all groups	0.440	0.411	0.687	0.673

During the initial assessment (carried out during the first version of testing), the agreement of the RVC specialists within the group was moderate both when dichotomized ASPECTS ≤ 7 and > 7, and ≤ 6 and > 6. When using the first reader mode, an increase in inter-rater agreement was found to be substantial. A similar trend in changes in agreement rates was observed among GH radiologists, whose Fleiss's kappa values indicated fair agreement between specialists during the initial assessment and an increase in agreement to moderate (almost substantial) during the second testing. It is worth noting that agreement with the program of RVC specialists during the initial assessment with dichotomized ASPECTS ≤ 6 and > 6 and ≤ 7 and > 7 was inferior to agreement with each other, however, the Fleiss's kappa indicators when dichotomized ASPECTS ≤ 6 and > 6 and ≤ 7 and > 7 became higher when using the first reader mode, which suggests that experts trusted the program more at this stage than when using the second reader mode. Also, during the initial assessment, RVC specialists gave a score of ≤ 7 in 50.5% of cases; during the second testing option on sample 2C, this figure decreased to 39% (compared to the first testing option - 42.5%). Doctors without experience in assessing stroke (working in GH) during the initial interpretation gave a score of ≤ 7 in 68% of cases; in the second version of testing on collections of images 2C, this percentage decreased to 53 (at the first stage to 55.5%). Similarly, RVC doctors, before providing an interpretation of the artificial intelligence algorithm, in 64.5% of cases assessed the prevalence of ischemic changes in the middle cerebral artery territory as ≤ 6 points according to ASPECTS; when using the first reader mode, this figure decreased to 52.5% (to 57% when applying the second reading model). Their colleagues without experience in urgent neuroradiology scored ≤ 6 points on ASPECTS during the initial assessment in 78% of cases, and in 63% during the second testing (in 65% of cases during the first). These results suggest that as interrater agreement increases, examiners tend to decrease ASPECTS scores more markedly when using the first reader mode.

We also assessed the agreement of radiologists during the initial assessment and the second joint testing on collections of images 2C regarding each of the ASPECTS areas. The results are presented in Table 28.

Table 28 – Indicators of interrater agreement among radiologists with varying level experience in urgent neuroradiology and less than three years of experience at the initial assessment of ASPECTS areas and after the second joint testing on a set of 2C CT images

	Initial assessment										Second joint testing (First-Reader Mode)									
	C	IC	L	I	M1	M2	M3	M4	M5	M6	C	IC	L	I	M1	M2	M3	M4	M5	M6
Within the group																				
RVC Radiologists	0.520	0.165	0.480	0.504	0.317	0.525	0.478	0.652	0.518	0.472	0.796	0.557	0.567	0.528	0.571	0.668	0.564	0.619	0.640	0.735
GH Radiologists	0.520	0.070	0.141	0.109	0.250	0.010	0.138	0.183	0.134	0.157	0.640	0.747	0.457	0.365	0.547	0.606	0.478	0.672	0.622	0.525
Between all groups	0.497	0.090	0.321	0.238	0.312	0.281	0.325	0.470	0.335	0.340	0.731	0.587	0.551	0.449	0.550	0.646	0.506	0.650	0.653	0.592
Agreement with the program																				
RVC radiologists + program	0.532	0.191	0.440	0.431	0.335	0.484	0.480	0.632	0.516	0.472	0.815	0.614	0.605	0.535	0.586	0.683	0.592	0.639	0.664	0.674
GH radiologists + program	0.458	0.013	0.265	0.158	0.269	0.057	0.192	0.225	0.187	0.604	0.719	0.756	0.577	0.454	0.578	0.693	0.535	0.707	0.672	0.579
Between all groups and the program	0.496	0.112	0.337	0.248	0.317	0.284	0.342	0.467	0.350	0.371	0.754	0.623	0.583	0.476	0.565	0.669	0.533	0.662	0.667	0.595

Among RVC doctors, during the second testing on collections of images 2C, agreement regarding each of the ASPECTS areas changed significantly in relation to areas C, M2, M5, M6 (from moderate to substantial agreement) and IC (from slight to moderate agreement). A similar picture was observed among GH specialists: agreement significantly increased relative to areas IC, M2, M4, M5, M6 (from slight to substantial agreement), and L (from slight to moderate agreement).

There was a uniform overall increase in Fleiss's kappa values relative to ASPECTS areas (both basal nuclear and insular regions, as well as hemispheric regions) after the second testing on the collections of images 2C across all radiologists, with moderate to substantial agreement.

Based on the results of the fifth stage of the thesis work, it can be concluded that the second reader mode has a negative impact on the performance of doctors with less than three years of experience and experience in diagnosing stroke on CT images. Also, the second reader mode is less effective in improving the performance of doctors without experience in the RVC and with less than three years of experience. At the same time, the use of a first reader mode when introducing an artificial intelligence algorithm into the diagnostic process of doctors with less than three years of experience has a positive effect in increasing the diagnostic efficiency of detecting middle cerebral artery stroke and their inter-expert agreement when assessed according to ASPECTS. It is necessary to especially highlight the absence of a summative decrease in diagnostic efficiency in relation to an increase in overdiagnosis rates with a complementary assessment by doctors and a program, with an identical positive effect in the form of a decrease in overdiagnosis rates among radiologists.

FINAL STATEMENT

To date, neuroimaging occupies a key role in the diagnosis of ischemic stroke. Taking into account a number of reasons (time of the examination, absence of absolute contraindications, availability), CT is the leading method of neuroimaging in the diagnosis of ischemic stroke, which should first solve this main task – the exclusion of ischemic brain injury and/or intracerebral hemorrhage. Detecting signs of thrombosis of the main cerebral arteries and ischemia in the early stages is a difficult diagnostic task, especially for radiologists with little experience. With the annual introduction of young specialists in hospitals, their adaptation period, and the general shortage of radiologists for direct interpretation of X-ray examinations (including CT) in the federal subjects of the Russian Federation, it is likely that variability in the interpretation of radiation images by medical doctors will naturally increase, leading to a deterioration in the quality of medical care.

Considering that the middle cerebral artery territory is most commonly affected in stroke [53], the ASPECTS was developed in 2000 to provide a unified approach to stroke diagnosis in this area. Many researchers consider this 10-point system to be a reliable diagnostic method. However, the use of this scale has a number of limitations, including high variability in the interpretation of the presence of ischemic changes by different experts, which may influence the further course of the patient's treatment [3, 75]. To partially solve the problems of subjectivity in assessment with the ASPECTS, it is proposed to introduce an artificial intelligence algorithm as a system to support medical decision making. To date, the use of such systems has been based on analysis of noncontrast CT images, CT angiography, and CT perfusion. Their application aims to automatically determine the score on the ASPECTS, quantify the stroke core, penumbra, and collateral blood flow status, and localize arterial occlusions [113]. The use of such algorithms foresees more effective detection of ischemic changes, a reduction in the number of cases of under- and overdiagnosis, and a reduction in variability between experts in the assessment of CT images of patients requiring emergency medical care [3, 49, 113].

A multistage research was performed to investigate the impact of automated CT image analysis systems on the process of medical decision making in the diagnosis of ischemic middle cerebral artery stroke using computed tomography.

After studying the domestic and foreign publications on this problem in the first stage, a database based on the results of CT examinations of 150 patients with the clinical picture of middle cerebral artery stroke was created and registered in the second stage [19]. Later, two collections of CT images were created based on this database to test medical doctors and automated analysis systems. The first collection included 50 CT examinations of patients with confirmed ischemic stroke and 50 CT examinations without pathological changes after noncontrast CT, CT angiography, and CT perfusion. The second collection included only CT scans of patients with ischemic stroke, which was also detected by methods of radiation diagnosis.

In the third stage, we tested 21 radiologists with different experience and expertise in emergency neuro-radiology as the first part of the study and fifteen RVC radiologists – as the second part of the study. The first collection of CT images was used to determine the diagnostic effectiveness of the specialists and their agreement in detecting early ischemic changes, and the second collection was used to evaluate agreement in dichotomous classification of points on the ASPECTS (the volume of early ischemic changes in the middle cerebral artery territory). The influence of professional experience of specialists in emergency neuroradiology of regional vascular centers on diagnostic effectiveness and agreement for most CT signs of ischemic stroke was demonstrated. The accuracy of specialists with expert qualification (experience in the assessment of ischemic stroke and clinical practice of more than eight years) was 93.8%, while their colleagues from regional vascular centers three years to less eight years of experience had an accuracy of 91.3% and colleagues with experience of up to three years – 88.3%. At the same time, years of clinical practice did not affect the diagnostic efficiency of medical doctors who did not have relevant experience in stroke assessment (i.e., who worked in hospitals that were not affiliated with regional vascular centers): specialists with more than eight years of clinical practice achieved 82.3% accuracy, specialists with three years to less eight years of experience – 85.0%, specialists with less than three years of clinical practice – 75%. It also appeared that specialists tended to overdiagnoses patients when evaluating CT scans, regardless of their years of clinical practice or experience in emergency pathology. This trend was more evident among medical doctors from general hospitals with less than three years of experience. The most specific and reproducible CT evidence of ischemic stroke was the sign of hyperdense MCA – all experts within their groups showed the highest response (from 0.095 to 0.75). The most ambiguous sign of ischemic stroke was sulcal effacement when interpreted by radiologists (-0.020 (random response) to 0.689). Signs of decreased brain matter density and impaired cortex differentiation had moderate and substantial reliability among regional vascular center specialists (from 0.493 to 0.671). In addition, we defined inter-rater reliability between specialists of regional vascular centers with different experience in determining the extent of ischemic changes on the ASPECTS. There was high estimated variability with low inter-rater reliability among specialists, regardless of dichotomous classification – 6 and 7 (indications for thrombectomy and thrombolysis) – and their experience (mean 0.391 (for ASPECTS 6) and 0.376 (for ASPECTS 7) – low reliability). Considering the moderate and substantial reliability between specialists of regional vascular centers regarding the signs of ischemic stroke and the high variability in the assessment by ASPECTS, this scale showed low reproducibility values in our research, confirming the subjectivity of the assessment by ASPECTS [4, 5].

It is worth noting that in foreign publications, studies to determine inter-rater reliability play an important role, as standardized assessment is the key to successful diagnosis of ischemic stroke in the acute stage [4]. The work of Farzin B. et al [52] investigated the degree of inter-rater and intra-rater reliability between 15 experts with different specializations (neurologists, radiologists, endovascular surgeons, and neuroradiologists) and professional experience (mostly more than 10 years) when performing assessments by ASPECTS. Even with dichotomous classification (0-5/5-10 points), intra-rater reliability did not reach a significant level (0.561 kappa statistic), meaning that at least 5 of 15 experts reached a different conclusion in 15% of the cases. Intra-rater reliability (experts' agreement with their own assessment during the first or second review) ranged from 0.599 to 0.943 [4, 52]. Also, Farzin B. et al [52] analyzed articles on the variability of reliability in assessment with the ASPECTS, published between 2000 and 2015. The methods of the analyzed studies differed from each other in several characteristics, including whether clinical information about the patient was provided to the experts, the time allotted for the assessment, access to all areas of the CT examination, and the ability of the experts to set their own window parameters. The results of this review of the current state of ASPECTS reproducibility reflect a high degree of variability in inter-rater agreement. For example, in the studies described, the coefficients of inter-rater reliability (Fleiss' kappa) ranged from 0.26 to 0.97 for ASPECTS without dichotomous classification and from 0.16 to 0.93 for ASPECTS with dichotomous classification, ([≤ 6 and > 6] and [≤ 7 and > 7]) [4, 52]. Despite this diversity, substantial inter-rater agreement with dichotomous classification (independent of [≤ 6 and > 6] or [≤ 7 and > 7]) was reported in the publication related to ASPECTS [72], although some studies reported moderate reliability [75].

In this thesis we also obtained quite different results, but on average they did not reach significant inter-rater agreement. We also found an inverse relationship between the degree of inter-rater reliability and the years of work of a specialist (although this dependence is directly proportional for foreign publications). This is probably due to the fact that there are obvious weaknesses in the standardization of the scale despite multiple reviews: There is no unified approach to describing the features of early ischemic changes, the ranges of ASPECTS do not have exact boundaries, and when the scale was introduced into neuroradiological practice in the Russian Federation, a unified and structural approach to its evaluation was not adopted due to the existing limitations of its application [4].

In the fourth stage, an analytical validation of three programs (A, B, C) for the analysis of computed tomograms of the brain was performed to investigate the possibility of using automated systems for the detection of ischemic middle cerebral artery stroke. First, diagnostic strength was assessed using one collection of images. According to the clinical recommendations (80%) for testing software based on artificial intelligence technologies [15], none of the systems involved can be approved for clinical validation with accuracy values in the range of 67% to 75%. Software products B and C had high specificity (0.9 and 0.8, respectively) and low sensitivity (0.6 and 0.55, respectively), indicating rare cases of overdiagnosis and frequent cases of underdiagnosis. To obtain a more reliable understanding of the diagnostic capabilities of these software products, clinical studies should be continued using both the analytical validation method with different collections and the clinical validation method. However, program A showed results that allow us to assess its low underdiagnosis of pathological changes (sensitivity 88 and true positives – 44 out of 100, false positives – 27 out of 100). Specialists with less than three years of experience and no professional experience in regional vascular centers had 3.5 times more false-negative and 1.5 times fewer true-positive results compared with program A. With respect to these indicators, specialists with up to three years of experience in regional vascular centers were also slightly inferior to program A. Considering the potential positive impact on young professionals with up to three years of experience, regardless of stroke assessment experience, in terms of reducing cases of missing pathology, this artificial intelligence algorithm was selected for further investigation and tested with two collections of CT images [2].

It should be noted that the accuracy index (0.80) in this case is controversial according to the recommendations that allow the program to be approved for clinical validation. Specialists of regional vascular centers, regardless of the years of their clinical practice, achieved higher indicators of diagnostic strength than the level of the indicated parameter in the clinical recommendations [2, 15].

In the study by P.V. Gavrilov et al [8], which dealt with the detection of round formations in the lungs on X-rays, the accuracy of specialists without experience was 73.0% and with experience 77.0%, depending on the length of service of the radiologists and their experience in thoracic radiology. Indicators of sensitivity and specificity in detecting round lung formations were also slightly higher (81.0%; 75.0%) than for specialists without experience (75.0%; 72.0%) but did not differ significantly. With such results from experienced specialists, the indicator that formally clears the tested program for further clinical validation appears to be important.

However, according to the results of the tests for specialists with different experience and expertise in the diagnosis of ischemic stroke, this indicator is relevant only for specialists without experience in emergency medicine and with less than three years of clinical practice.

It is also worth considering the presence of a pathology that is difficult to diagnose. For example, in the thesis work of A. A. Meldo [16], which dealt with the development and implementation of an artificial intelligence system in the X-ray diagnosis of pulmonary nodules and masses, the CT collections submitted to specialists for evaluation were divided into difficult-to-interpret cases (typical lung cancer, atypical lung cancer, and no lung cancer) and cases with predominantly typical visualization patterns (typical lung cancer, no lung cancer). The accuracy of five specialists with different experience (more than five years and less than five years) in cancer diagnosis ranged from 51.0% to 71.1% when interpreting the first collection and when evaluating the second collection – from 82.0% to 96.0%. Given such a pronounced range of accuracy values in the study, it is necessary to differentiate the indicator that formally allows the tested program for further clinical validation in the group of this pathology.

It is probably necessary to consider the current efficiency indicators of specialists in the diagnosis of various pathological processes to determine the criteria for approval of algorithms for clinical validation. Otherwise, the introduction of automated analysis systems may lead to negative dynamics in the accuracy of detection of pathological changes due to the interference of specialists without experience or reduce confidence in the program due to obviously wrong results received [2].

In the fifth stage of the research, possible options for the introduction of automated analysis systems of computed tomography images into the clinical practice of a radiologist were investigated by modeling two different situations of interaction between a specialist and program A. Two models of introduction of an artificial intelligence algorithm as a medical decision support system were applied: the first (parallel) reader mode and a second reader mode. The first (parallel) reader mode assumed that the CT images were first analyzed by an AI algorithm and then a specialist reviewed the diagnostic results of the automated analysis system. In the second reader mode, a radiologist reviewed the examinations, further analysis was performed by the AI algorithm, and final interpretation of the data was performed by a radiologist, with possible inclusion of additional findings detected by the AI. Tests were performed with collections 1 and 2 to determine the impact of the program on the diagnostic strength of specialists with up to three years of experience in urgent neuroradiology or their reliability in terms of their assessment with ASPECTS. Seven specialists with less than three years of experience from the second stage of the research participated in the test. They were divided into two groups: Emergency medicine specialists (regional vascular centers) and specialists without experience in regional vascular centers. Using the second reader mode, overall diagnostic strength scores (accuracy, specificity) decreased to a greater extent among RVC specialists due to specificity (from 81.5% to 79.5%) and an increase in the number of false-positive results (from 9.25 to 10.25), which contributed to an increase in overdiagnosis in this expert group despite their individual results. However, in the group of specialists who had no experience with the regional vascular centers, there was an improvement in performance indicators (accuracy, sensitivity, and specificity), mainly because of an increase in sensitivity (from 67, (3) % to 92, (6)) % and a decrease in false-negative results (from 6, (3) to 3, (6)), i.e., with a decrease in underdiagnoses. These results suggest a positive effect of artificial intelligence on the diagnostic strength of specialists without experience when the second reader mode is used.

No significant changes in Fleiss' kappa indicators were found when the second reader mode was used (moderate response), regardless of the dichotomous classification of ASPECTS scores (6 and 7), suggesting that the medical decision support system does not affect inter-rater reliability in assessing the extent of ischemic changes according to ASPECTS. Reliability of specialists without professional experience in regional vascular centers changed with increasing values of Fleiss' kappa, but also in the range of fair reliability. The second test with the first reader mode was performed after two months. The average indicators of diagnostic strength (accuracy, sensitivity, and specificity) increased more for accuracy (from 90.2% to 94.25%) and specificity (from 81.5% to 89%) among RVC specialists, with decreases in the number of false-positive (from 18, (6) to 9) and false-negative (from 6, (3) to 2) results and in the degree of under- and overdiagnosis. A similar trend was observed among specialists without professional experience in regional vascular centers. In this group of specialists, there was a pronounced positive dynamic of average efficiency scores with an increase in sensitivity (from 67, (3) % to 96%), specificity (from 62, (6) % to 82%), and accuracy (from 75% to 89%), with a decrease in the number of false-positive (from 18, (6) to 9) and false-negative (from 6, (3) to 2) results. The results of using the first reader mode showed improvement in the interpretation of CT images of patients with suspected of MCA stroke during assessment with the artificial intelligence algorithm both with and without specialists experienced in stroke assessment.

Similarly, use of the parallel reader mode among RVC specialists showed an increase in inter-rater reliability from moderate to substantial when evaluating with ASPECTS with dichotomous classification of scores ≤ 7 and > 7 and ≤ 6 and > 6. A similar trend of reliability changes was observed among specialists from general hospitals, whose Fleiss' kappa indicators showed fair inter-rater agreement at the first assessment and an increase in fair to moderate (almost significant) at the second test.

The results of the fifth stage of this thesis showed a pronounced positive effect of using the parallel reader mode with an increase in diagnostic strength of specialists with up to three years of experience, with an increase in inter-rater reliability when using ASPECTS. The introduction of an automated ischemic stroke detection system into the diagnostic process of less experienced RVC specialists (with less than three years of experience) allows them to improve their ability to interpret CT images and to approach the diagnostic strength of specialists with expert qualifications (with more than eight years of experience) due to indicators of sensitivity and accuracy.

It is worth noting that all specialists involved in the research who used both the parallel and second reader modes tended to lower the ASPECTS score after a joint assessment with the automated stroke detection program. In the initial analysis, RVC specialists reported a score of ≤ 7 in 50.5% of cases. This indicator decreased to 42.5% in the first test and to 39% in the second test. Specialists with no experience in stroke assessment (non-RVC) reported a score of ≤ 7 in 68% of cases; this percentage decreased to 55.5 in the first test and 53 in the second test. Similarly, before the evaluation of the artificial intelligence algorithm, RVC specialists estimated the extent of ischemic changes in the middle cerebral artery territory according to ASPECTS to be ≤ 6 in 64.5% of cases; in the second test, this indicator decreased to 57%, and in the parallel analysis – to 52.5%. Their colleagues without experience in urgent neuroradiology reported ≤ 6 points according to ASPECTS at the first examination in 78% of cases, in 65% of cases – at the first test, and in 63% of cases – in the second test. These results suggest that as inter-rater reliability increases, specialists tend to lower the ASPECTS score more significantly when using a parallel reader mode.

According to foreign publications, researchers have found positive results when specialists interact with automated CT image analysis systems. For example, Brinjikji W. et al. [51] tested the software e-ASPECTS together with twelve radiologists and four neurologists. The study included native CT examinations of 60 patients with confirmed middle cerebral artery stroke. First, the specialists evaluated the non-contrast CT images of the brain and scored them according to ASPECTS. Two months later, the specialists evaluated the image data again, but this time together with the software. The authors found that inter-rater reliability improved significantly with the use of e-ASPECTS. The study did not examine the characteristics of using modes to implement medical decision support systems, so we cannot confirm the effectiveness of using the parallel reader mode.

In the thesis work of U.A. Smolnikova [20], which deals with the possibilities of automated analysis systems of digital X-rays in the diagnosis of pulmonary nodules and masses, despite the good results (in some cases better than those of specialists) of analytical validation of automated analysis systems with joint interpretation (parallel reader Mode, according to N.A. Obuchowski N. A. et al [88]) of X-rays by a specialist and automated analytical systems led to a summation of human errors and those of the programs, which resulted in a deterioration of diagnostic strength (a decrease in sensitivity from 83% to 56.7%, specificity from 99% to 93.9%). Using the second reader mode with an initial assessment of X-rays by a specialist followed by the use of an automated analysis system and repeated decision making, a decrease in sensitivity to 66.7% and an increase in specificity to 95% (by 10.6%) were observed. It is worth noting that in this study, the use of the second reader mode was more appropriate because it improved the sensitivity of professional judgments, which is different from the results of the research of this thesis. When choosing an automated analysis system, it is probably worthwhile not only to take a personalized approach, but also to evaluate the characteristics of the interaction between specialists and an artificial intelligence depending on their area of expertise.

Despite the positive results of the use of automated analysis systems, further research is needed to develop their ideal interaction with specialists and improve the indicators of diagnostic strength. However, it is already possible to assess the prospects for introducing artificial intelligence algorithms into a specialist’s clinical practice.

CONCLUSIONS

1. The thesis work proves the direct dependence of the degree of diagnostic strength in the detection of ischemic changes in the middle cerebral artery area by computed tomography on the years of practice of radiologists and their experience in the regional vascular center was established.
2. There was substantial inter-rater reliability of regional vascular center specialists in detecting hyperdense middle cerebral artery sign (0.684) and sign of loss of gray- white differentiation (0.629), as well as moderate agreement regarding the sign of sulcal effacement (0.525) and a hypoattenuation of the cortex (0.529).
3. Despite substantial reliability for the individual CT signs of ischemic stroke (ranging from 0.529 to 0.684) and high diagnostic strength in detecting the middle cerebral artery stroke (accuracy ranging from 88.3% to 93.8%), we revealed low reproducibility of the ASPECTS scale (0.391 for the dichotomous classification of ASPECTS scores with a cutoff of 6 and 0.376 for the dichotomous classification of ASPECTS scores with a cutoff of 7) by specialists in regional vascular centers, regardless of the duration of treatment.
4. The diagnostic strength of currently existing automated CT image analysis systems (accuracy between 67.0% and 75.0%) in detecting ischemic middle cerebral artery stroke is lower than the results obtained by specialists in regional vascular centers, regardless of how long they have been working (accuracy between 88.3% and 93.8%).
5. The optimal option for implementing automated image analysis systems with a low degree of underdiagnosis and a high risk of overdiagnosis is to use a mode in which the CT examinations are first evaluated by software, followed by an assessment by a specialist, and the decision is made taking into account the results of the interpretation of the artificial intelligence algorithm. This approach allows an increase in diagnostic strength for specialists with less than three years

of experience (accuracy of 90.25% to 94.25% for RVC specialists and accuracy of 75% to 89% for general hospital specialists) and a decrease in variability in the assessment of ischemic changes according to ASPECTS (from 0.409 to 0.657 for dichotomous classification according to ASPECTS with a cutoff of 6 and from 0.425 to 0.669 for dichotomous classification according to ASPECTS with a cutoff of 7).

PRACTICAL RECOMMENDATIONS

1. The use of automated image analysis systems makes it possible to increase the diagnostic strength of radiologists with up to three years of experience in detecting middle cerebral artery ischemic stroke by computed tomography.
2. The use of automated CT image analysis systems makes it possible to increase the reproducibility of the ASPECTS among radiologists with up to three years of experience.
3. Currently existing programs for automated assessment of ischemic changes differ in their diagnostic strength and do not achieve the values required for further clinical validation.
4. It is recommended to test the implemented algorithm on an independent, verified collection of CT images.
5. In testing automated CT image analysis systems, the evaluation of parameters of high sensitivity is of paramount importance, which has an impact on reducing the risk of underdiagnosis of pathological changes.
6. When introducing automated analysis systems into the diagnostic process, it is advisable to choose a model in which the CT examinations are first evaluated by software, followed by an assessment by a specialist and final decision-making taking into account the results of the interpretation by the artificial intelligence algorithm.

PROSPECTS FOR FURTHER DEVELOPMENT OF THE TOPIC

Timely recognition and treatment of acute ischemic stroke is critical for reducing morbidity and mortality in the Russian population. The use of AI in the diagnosis of middle cerebral artery ischemic stroke offers numerous opportunities for more accurate and subsequently faster interpretation of radiation images, which may ultimately improve the quality of medical care and functional outcomes of the disease.

One of the most promising areas for the use of artificial intelligence systems is the creation of platforms with software products for computed tomographic detection of ischemic stroke, acute brachycephalic and cerebral artery occlusion, and automated assessment of CT perfusion. Further research is also needed to detect ischemic strokes in the vertebral-basilar region using artificial intelligence and to build prognostic models for disease development. The most important of these models are predicting the outcome of thrombolytic therapy and the outcomes of other treatments.

It is not yet clear whether this integration will prove successful in clinical practice. How will this affect the overall view of diagnostic and therapeutic interventions? How will this affect the legal agenda and the legal responsibilities of specialists and health care providers in the event of disputes, even though there are official guidelines and recommendations for the use of automated analysis systems?

Further research is needed to confirm the effectiveness of artificial intelligence methods in reducing variability in the interpretation of radiation images and to ensure wider application under various practical conditions. To obtain a full understanding of the diagnostic capabilities of these artificial intelligence products, clinical studies should be continued using both the analytical validation method with different collections and the clinical validation method.

LIST OF ABBREVIATIONS

AI – artificial intelligence

AIS – acute ischemic stroke;

ASPECTS (Alberta stroke program early CT score) – a 10-point quantitative topographic CT scan score used for middle cerebral artery (MCA) stroke patients;

CT – computed tomography;

CVA – cerebrovascular accidents;

DICOM (Digital Imaging and Communications in Medicine) – medical standard for creation, storage, transfer and visualization of digital medical images and documents of examined patients;

DWI – diffusion-weighted images;

EIC – early ischemic changes;

GH – general hospitals;

HAS – hyperdense middle cerebral artery;

ITT – intravenous thrombolytic therapy;

JPEG (Joint Photographic Experts Group) – one of the formats used for storing photos

MCA – middle cerebral artery;

MRI – magnetic resonance imaging;

mRS – Modified Rankin scale;

NIHSS – National Institutes of Health Stroke Scale;

PACS (Picture Archiving and Communication System) — is a clinical data management system consisting of several open-source medical imaging technologies that can be used to store, access, transfer and manage medical images and digital reports in DICOM format;

PVC – primary vascular compartments;

RF – Russian Federation;

RVC – regional vascular center.

GLOSSARY OF TERMS

A neural network is a mathematical model, as well as its software or hardware implementation, which is built on the principle of organization and functioning of biological neural networks - networks of nerve cells of a living organism, from the point of view of machine learning, a neural network is a special case of pattern recognition methods, discriminant analysis.

Analytical validation is an assessment of the correctness of processing input data by software providing for the creation of reliable output data; evaluated using reference labeled data sets.

Artificial intelligence is a set of technological solutions that allows you to simulate human cognitive functions (including self-learning and the search for solutions without a predetermined algorithm) and obtain, when performing specific tasks, the results that are comparable with the results of human intellectual activity.

Automated CT Image Analysis System – a system that allows you to produce detection, tracking and classification of objects, in particular image processing in medicine, it contributes to obtaining information from image data for staging medical diagnosis for patients.

Clinical approbation - evaluation of effectiveness through use within the standard production process, consists of two components: clinical correlation (assessment the presence of a significant clinical relationship between the results and the target clinical condition of the software), and clinical validation (confirmation of achievement of the intended goal in the target population in the context of clinical work through the use of accurate and reliable output).

Deep learning is a set of machine learning methods (with a teacher, with partial teacher-assisted, unsupervised, reinforced) based on representational learning, rather than specialized algorithms for specific tasks.

Diagnostic efficiency is a parameter that characterizes the capabilities of a given test systems at the same time correctly identify positive results as positive, and negative results as negative.

Hypodiagnosis - an erroneous medical conclusion about the absence of a disease in the examined person or its complications that are present or more pronounced than indicated in conclusion.

Machine learning is a class of artificial intelligence methods, a characteristic feature of which is not a direct solution of the problem but learning through the application of solutions of many similar tasks. To construct such methods, the data of the mathematical statistics are used, as well as numerical methods, mathematical 109 analysis, optimization methods, probability theory, graph theory, various techniques for working with data in digital form.

Overdiagnosis - an erroneous medical conclusion about the presence of a disease in the patient or its complications, which are absent or less pronounced than indicated in the conclusion.

REFERENCES

1. Analysis of epidemiological indicators of recurrent strokes in the regions of the Russian Federation (based on the results of the territorial-population register 2009–2014) / L. V. Stakhovskaya, O. A. Klochikhina, M. D. Bogatyreva, S. A. Chugunova. – 2016. – No. 18 (9). – pp. 8–11 (In Russian);
2. Andropova P.L. Diagnostic efficiency of separate artificial intelligence – based automated CT image analysis systems in detecting middle cerebral artery stroke / Andropova P.L., Gavrilov P.V., Kolesnikova P.A. [et al.] // The Siberian Journal of Clinical and Experimental Medicine. – 2023. – № 3(39). – pp. 194-200. (In Russian).
3. Andropova P.L. ASPECTS: interobserver agreement between radiologist / Andropova P.L., Gavrilov P.V., Savintseva Zh.I. // Diagnostic radiology and radiotherapy. – 2022. – № 1(13). – C. 7-13 (In Russian).
4. Andropova P.L. Interexpert Agreement between Neuroradiologists in the Diagnosis of Middle cerebral Artery Stroke by Computed Tomography / Andropova P.L., Gavrilov P.V., Kazantseva I.P. [et al.] // Medical Visualization. – 2023. – URL: https://doi.org/10.24835/1607-0763-1315 (In Russian).
5. Andropova P.L. Interexpert Agreement Between Emergency Neuroradiologists with Different Levels of Experience in the Rating of ASPECTS / Andropova P.L., Gavrilov P.V., Kazantseva I.P. [et al.] // Radiology - practice. – 2022. – № 5(95). – C. 10-25 (In Russian).
6. Andropova P.L. Application of artificial intelligence systems in neuroradiology of acute ischemic stroke / Andropova P.L., Gavrilov P.V., Savintseva Zh.I. [et al.] // Diagnostic radiology and radiotherapy. – 2021. – № 2(12). – C. 30-35 (In Russian).
7. Vasiliev, A. Yu. Analysis of data from radiation research methods based on the principles of evidence-based medicine / Vasiliev A. Yu., Maly A. Yu., Serov N. S. - Moscow: GEOTAR-Media, 2008. – 32 p. (In Russian).

8. Gavrilov P.V. Identifying pulmonary nodules and masses on digital X-rays: role of the radiologist’s experience / Gavrilov P.V., Ushkov A.D., Smolnikova U.A. // Medical Alliance. - 2019. - No.2. - p.51–56 (In Russian).
9. Roadmap for the development of “end-to-end” digital technology “Neurotechnologies and artificial intelligence”. URL: https://digital.gov.ru/ru/documents/6658. Retrieved June 17, 2020 (In Russian).
10. Stroke: Clinical Guide / Hennerici M.G., Boguslawski J., Sacco R.L. Per. from English – Generally ed. corr. member RAMS Skvortsova V.I. 2nd ed. M.: MEDpressinform, 2008. – 224 p. (In Russian).
11. Stroke: a guide for doctors / Ed. Stakhovskaya L.V., Kotova S.V. — Moscow: LLC “Medical Information Agency”, 2013. — 400 p. (In Russian).
12. Ischemic stroke and transient ischemic attack in adults: Clinical Guidelines / Akzhigitov R.G., Alekyan B.G., Alferova V.V. [et al.]. – M.: Ministry of Health of the Russian Federation, 2021. – 181 p. (In Russian).
13. Ischemic stroke and transient ischemic attack in adults: Clinical Guidelines / All-Russian Society of Neurologists [et al.]. – M.: Ministry of Health of the Russian Federation, as amended. and additional in ed. from 2022. – 214 p. (In Russian).
14. Ischemic stroke: malignant middle cerebral artery infarction: Clinical Guidelines / V.V. Krylov, O.N. Dreval, R.S. Dzhindzhikhadze [and others]. – M.: Ministry of Health of the Russian Federation, 2015. – 17 p. (In Russian).
15. Clinical trials of software based on intelligent technologies (radiology) / Morozov S. P., Vladzimirsky A. V., Klyashtorny V. G. [et al.] // Best practices in radiological and instrumental diagnostics. Issue 23. - 2019 – 34p. (In Russian).
16. Meldo A. A. Development and implementation of an artificial intelligence system in radiation diagnostics of focal formations in the lungs: dis. ...Dr. med. sciences; 3.1.25 / A.A. Meldo. - St. Petersburg, 2022. - 235. URL: https://disser.spbu.ru/files/2021/disser_meldo.pdf (08/29/2023) (In Russian).
17. National standard of the Russian Federation GOST R 59921.0-2022 "Artificial intelligence systems in clinical medicine. Basic principles" (In Russian).

18. National standard of the Russian Federation GOST R 59921.5-2022 “Artificial intelligence systems in clinical medicine. Part 5. Requirements for the structure and order of using a data set for training and testing algorithms" (In Russian).
19. Results of computed tomography examinations of patients with a clinical presentation of middle cerebral artery stroke / Andropova P.L., Gavrilov P.V., Savintseva Zh.I. [et al.] // Database Registration Certificate RU 2022620850, 18.04.2022. Request № 2022620723 from 07.04.2022 (In Russian).
20. Smolnikova U.A. Identifying pulmonary nodules or masses on chest radiography: capabilities of radiologists and automated analysis programs / Smolnikova U.A., Gavrilov P.V., Ushkov A.D. // Materials of the Congress of the Russian Society of Roentgenologists and Radiologists. – 2019. – P. 192-193. (In Russian).
21. Decree of the President of the Russian Federation dated October 10, 2019 No. 490 “On the development of artificial intelligence in the Russian Federation” (together with the “National Strategy for the Development of Artificial Intelligence for the period until 2030”). URL: http://www.consultant.ru/document/cons_doc_LAW_335184/. Access date: 03/15/2022 (In Russian).
22. Feigin, V. Stroke: Clinical Guide / V. Feigin, D. Wiebers, R. Brown. – M.: Binom, 2005. – 307 p. (In Russian).
23. Epidemiology of stroke in the Russian Federation: results of territory's population registry (2009-2010) / Stakhovskaia, L. V., Klochikhina, O. A., Bogatyreva, M. D., & Kovalenko, V. V. // Zhurnal nevrologii i psikhiatrii imeni S.S. Korsakova. – 2013. –№5. – C. 4-10 (In Russian).
24. Advanced [18F] FDG and [11C] flumazenil PET analysis for individual outcome prediction after temporal lobe epilepsy surgery for hippocampal sclerosis / J. Yankam Njiwa, K. R. Gray, N. Costes [et al.] // NeuroImage: Clinical. – 2015. – Vol. 7. – P. 122-131.
25. Advances in Auto-Segmentation / C. E. Cardenas, J. Yang, B. M. Anderson [et al.] // Seminars in Radiation Oncology. – 2019. – Vol. 29. – № 3. – P. 185-197.

26.Agreement and Variability in the Interpretation of Early CT Changes in Stroke Patients Qualifying for Intravenous rtPA Therapy / J. C. Grotta, D. Chiu, M. Lu [et al.] // Stroke. – 1999. – Vol. 30. – № 8. – P. 1528-1533.
27.Alberta Stroke Program Early Computed Tomographic Scoring Performance in a Series of Patients Undergoing Computed Tomography and MRI: Reader Agreement, Modality Agreement, and Outcome Prediction / R. A. McTaggart, T. G. Jovin, M. G. Lansberg [et al.] // Stroke. – 2015. – Vol. 46. – № 2. – P. 407-412.
28.Alberta Stroke Program Early CT Scoring of CT Perfusion in Early Stroke Visualization and Assessment / R. I. Aviv, J. Mandelcorn, S. Chakraborty [et al.] // American Journal of Neuroradiology. – 2007. – Vol. 28. – № 10. – P. 1975-1980.
29.Andropova P. A. Computer-Aided Diagnosis System in the Diagnosis of Multiple Sclerosis / P. Andropova, D. Cheremisin, A. Meldo. – Text: electronic // Proceedings of International Scientific Conference on Telecommunications, Computing and Control: Smart Innovation, Systems and Technologies / eds. N. Voinov [et al.]. – Singapore: Springer Singapore, 2021. – Vol. 220. – P. 277-284. – URL: https://link.springer.com/10.1007/978-981-33-6632-9_24 (date accessed: 22.08.2023).
30.Application of Machine Learning to Automated CT Analysis of Cerebral Edema in Large Cohorts of Ischemic Stroke Patients / R. Dhar, Y. Chen, H. An, J.-M. Lee // Frontiers in Neurology. – 2018. – Vol. 9. – P. 687.
31.Artificial Intelligence and Acute Stroke Imaging / J. E. Soun, D. S. Chow, M. Nagamine [et al.] // AJNR. American journal of neuroradiology. – 2021. – Vol. 42. – № 1. – P. 2-11.
32.Artificial Intelligence and Machine Learning in Radiology: Opportunities, Challenges, Pitfalls, and Criteria for Success / J. H. Thrall, X. Li, Q. Li [et al.] // Journal of the American College of Radiology. – 2018. – Vol. 15. – № 3. – P. 504-508.
33.Artificial intelligence-driven ASPECTS for the detection of early stroke changes in non-contrast CT: a systematic review and meta-analysis / A. Adamou, E. T.

Beltsios, A. Bania [et al.] // Journal of NeuroInterventional Surgery. – 2022.– P. jnis-2022-019447.
34. Assistance from Automated CT ASPECTS Software Improves Reader Performance / P. R. Delio, M. L. Wong, J. P. Tsai [et al.] // Journal of Stroke and Cerebrovascular Diseases. – 2021. – Vol. 30. – № 7. – P. 105829.
35. Automated CT quantification of cerebral edema following hemispheric infarction: Application of a machine-learning algorithm to evaluate CSF shifts on serial head CTs / Y. Chen, R. Dhar, L. Heitsch [et al.] // NeuroImage: Clinical. – 2016. – Vol. 12.– P. 673-680.
36. Butcher K. Acute Stroke Imaging Part I: Fundamentals / K. Butcher, D. Emery // Canadian Journal of Neurological Sciences / Journal Canadien des Sciences Neurologiques. – 2010. – Vol. 37. – № 1. – P. 4-16.
37. Can structural MRI aid in clinical classification? A machine learning study in two independent samples of patients with schizophrenia, bipolar disorder and healthy subjects / H. G. Schnack, M. Nieuwenhuis, N. E. M. Van Haren [et al.] // NeuroImage. – 2014. – Vol. 84. – P. 299-306.
38. Chen L. Fully automatic acute ischemic lesion segmentation in DWI using convolutional neural networks / L. Chen, P. Bentley, D. Rueckert // NeuroImage: Clinical. – 2017. – Vol. 15. – P. 633-643.
39. Classifying Acute Ischemic Stroke Onset Time using Deep Imaging Features / K. C. Ho, W. Speier, S. El-Saden, C. W. Arnold // AMIA ... Annual Symposium proceedings. AMIA Symposium. – 2017. – Vol. 2017. – P. 892-901.
40. Comparison of the performance between Frontier ASPECTS software and different levels of radiologists on assessing CT examinations of acute ischaemic stroke patients / L. Li, Y. Chen, Y. Bao [et al.] // Clinical Radiology. – 2020. – Vol. 75. – № 5. – P. 358-365.
41. Computer-assisted identification and quantification of multiple sclerosis lesions in MR imaging volumes in the brain / J. R. Mitchell, S. J. Karlik, D. H. Lee, A.

Fenster // Journal of Magnetic Resonance Imaging. – 1994. – Vol. 4. – № 2. – P. 197-208.

42. Convolutional neural networks: an overview and application in radiology / R. Yamashita, M. Nishio, R. K. G. Do, K. Togashi // Insights into Imaging. – 2018. – Vol. 9. – № 4. – P. 611-629.
43. Crowdsourcing image analysis for plant phenomics to generate ground truth data for machine learning / N. Zhou, Z. D. Siegel, S. Zarecor [et al.] // PLOS Computational Biology. – 2018. – Vol. 14. – № 7. – P. e1006337.
44. Cruz J. A. Applications of machine learning in cancer prediction and prognosis / J. A. Cruz, D. S. Wishart // Cancer Informatics. – 2007. – Vol. 2. – P. 59-77.
45. CT interpretation by ASPECTS in hyperacute ischemic stroke predicting functional outcomes / W. Phuttharak, K. Sawanyawisuth, B. Sangpetngam, S. Tiamkao // Japanese Journal of Radiology. – 2013. – Vol. 31. – № 10. – P. 701-705.
46. CT Perfusion ASPECTS in the Evaluation of Acute Ischemic Stroke: Thrombolytic Therapy Perspective / N. Sillanpaa, J. T. Saarinen, H. Rusanen [et al.] // Cerebrovascular Diseases Extra. – 2011. – Vol. 1. – № 1. – P. 6-16.
47. CT visualization of intracranial arterial thromboembolism. / G. Gács, A. J. Fox, H. J. Barnett, F. Vinuela // Stroke. – 1983. – Vol. 14. – № 5. – P. 756-762.
48. Decoding post-stroke motor function from structural brain imaging / J. M. Rondina, M. Filippone, M. Girolami, N. S. Ward // NeuroImage: Clinical. – 2016. – Vol. 12. – P. 372-380.
49. Dermatologist-level classification of skin cancer with deep neural networks / A. Esteva, B. Kuprel, R. A. Novoa [et al.] // Nature. – 2017. – Vol. 542. – № 7639. – P. 115-118.
50. Differential Prognosis of Isolated Sulcal effacement and Hypoattenuation on CT in Acute Stroke / K. S. Butcher, S. B. Lee, M. W. Parsons [et al.] // Stroke. – 2007. – Vol. 38. – № 3. – P. 941-947.

51.e-ASPECTS software improves interobserver agreement and accuracy of interpretation of aspects score / W. Brinjikji, M. Abbasi, C. Arnold [et al.] // Interventional Neuroradiology. – 2021. – Vol. 27. – № 6. – P. 781-787.
52.Early CT changes in patients admitted for thrombectomy: Intrarater and interrater agreement / B. Farzin, R. Fahed, F. Guilbert [et al.] // Neurology. – 2016. – Vol. 87. – № 3. – P. 249-256.
53.Early Identification of Potentially Salvageable Tissue with MRI-Based Predictive Algorithms after Experimental Ischemic Stroke / M. J. Bouts, I. A. Tiebosch, A. Van Der Toorn [et al.] // Journal of Cerebral Blood Flow & Metabolism. – 2013. – Vol. 33. – № 7. – P. 1075-1082.
54.Early prediction of irreversible brain damage after ischemic stroke at CT / R. von Kummer, H. Bourquain, S. Bastianello [et al.] // Radiology. – 2001. – Vol. 219. – № 1. – P. 95-100.
55.Effect of baseline Alberta Stroke Program Early CT Score on safety and efficacy of intra-arterial treatment: a subgroup analysis of a randomised phase 3 trial (MR CLEAN) / A. J. Yoo, O. A. Berkhemer, P. S. S. Fransen [et al.] // The Lancet. Neurology. – 2016. – Vol. 15. – № 7. – P. 685-694.
56.Effects of Alteplase for Acute Stroke on the Distribution of Functional Outcomes: A Pooled Analysis of 9 Trials / K. R. Lees, J. Emberson, L. Blackwell [et al.] // Stroke. – 2016. – Vol. 47. – № 9. – P. 2373-2379.
57.Endovascular Treatment of Middle Cerebral Artery M2 Occlusion Strokes: Clinical and Procedural Predictors of Outcomes / M. Mokin, C. T. Primiani, Z. Ren [et al.] // Neurosurgery. – 2017. – Vol. 81. – № 5. – P. 795-802.
58.Extent of hypoattenuation on CT angiography source images predicts functional outcome in patients with basilar artery occlusion / V. Puetz, P. N. Sylaja, S. B. Coutts [et al.] // Stroke. – 2008. – Vol. 39. – № 9. – P. 2485-2490.
59.Feng G. C. Mistakes and How to Avoid Mistakes in Using Intercoder Reliability Indices / G. C. Feng // Methodology. – 2015. – Vol. 11. – № 1. – P. 13-22.

60. Fiebach J. B. Moderne Kernspintechniken beim Schlaganfall / J. B. Fiebach, P. D. Schellinger // Der Radiologe. – 2003. – Vol. 43. – № 3. – P. 251-264.
61. Gillies R. J. Radiomics: Images Are More than Pictures, They Are Data / R. J. Gillies, P. E. Kinahan, H. Hricak // Radiology. – 2016. – Vol. 278. – Radiomics. – № 2. – P. 563-577.
62. Global, regional, and national burden of stroke, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016 / C. O. Johnson, M. Nguyen, G. A. Roth [et al.] // The Lancet Neurology. – 2019. – Vol. 18. – № 5. – P. 439-458.
63. Guidelines for the Early Management of Adults With Ischemic Stroke: A Guideline From the American Heart Association/ American Stroke Association Stroke Council, Clinical Cardiology Council, Cardiovascular Radiology and Intervention Council, and the Atherosclerotic Peripheral Vascular Disease and Quality of Care Outcomes in Research Interdisciplinary Working Groups: The American Academy of Neurology affirms the value of this guideline as an educational tool for neurologists / H. P. Adams, G. Del Zoppo, M. J. Alberts [et al.] // Stroke. – 2007. – Vol. 38. – № 5. – P. 1655-1711.
64. Hallgren K. A. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial / K. A. Hallgren // Tutorials in Quantitative Methods for Psychology. – 2012. – Vol. 8. – № 1. – P. 23-34.
65. Ho M.-L. Cerebral Edema / M.-L. Ho, R. Rojas, R. L. Eisenberg // American Journal of Roentgenology. – 2012. – Vol. 199. – № 3. – P. W258-W273.
66. Hsu K.-C. Backward somersault as a cause of childhood stroke: a case report of isolated middle cerebral artery dissection in an adolescent boy / K.-C. Hsu, H.-W. Kao, S.-J. Chen // The American Journal of Emergency Medicine. – 2008. – Vol. 26. – № 4. – P. 519.e3-519.e5.
67. Huang S. Quantitative prediction of acute ischemic tissue fate using support vector machine / S. Huang, Q. Shen, T. Q. Duong // Brain Research. – 2011. – Vol. 1405. – P. 77-84.

68. Hyperdense Sylvian Fissure MCA ‘Dot’ Sign: A CT Marker of Acute Ischemia / P. A. Barber, A. M. Demchuk, M. E. Hudon [et al.] // Stroke. – 2001. – Vol. 32. – № 1. – P. 84-88.
69. Hyperacute stroke: experience essential when reading unenhanced CT scans / J. H. W. Pexman, M. D. Hill, A. M. Buchan [et al.] // AJNR. American journal of neuroradiology. – 2004. – Vol. 25. – № 3. – P. 516; author reply 516-518.
70. Identifying spatial imaging biomarkers of glioblastoma multiforme for survival group prediction: GBM Survival Group Prediction / M. Zhou, B. Chaudhury, L. O. Hall [et al.] // Journal of Magnetic Resonance Imaging. – 2017. – Vol. 46. – № 1. – P. 115-123.
71. Impact of Window Setting Optimization on Accuracy of Computed Tomography and Computed Tomography Angiography Source Image-based Alberta Stroke Program Early Computed Tomography Score / E. M. Arsava, J. T. Saarinen, A. Unal [et al.] // Journal of Stroke and Cerebrovascular Diseases. – 2014. – Vol. 23. – № 1. – P. 12-16.
72. Interobserver Variation of ASPECTS in Real Time / S. B. Coutts, A. M. Demchuk, P. A. Barber [et al.]. – Text: electronic // Stroke. – 2004. – Vol. 35. – № 5. – URL: https://www.ahajournals.org/doi/10.1161/01.STR.0000127082.19473.45 (date accessed: 22.08.2023).
73. Ischemic Brain Tissue Water Content: CT Monitoring during Middle Cerebral Artery Occlusion and Reperfusion in Rats / I. Dzialowski, E. Klotz, S. Goericke [et al.] // Radiology. – 2007. – Vol. 243. – № 3. – P. 720-726.
74. Jha B. Pearls & Oy-sters: Hyperdense or pseudohyperdense MCA sign: A Damocles sword? / B. Jha, M. Kothari // Neurology. – 2009. – Vol. 72. – № 23. – P. e116-e117.
75. Kobkitsuksakul C. Interobserver agreement between senior radiology resident, neuroradiology fellow, and experienced neuroradiologist in the rating of Alberta Stroke Program Early Computed Tomography Score (ASPECTS) / C. Kobkitsuksakul, O. Tritanon, V. Suraratdecha // Diagnostic and Interventional

Radiology. – 2018. – URL: http://cms.galenos.com.tr/Uploads/Article_55149/Diagn%20Interv%20Radiol-24-104-En.pdf (date accessed: 22.08.2023).

76. Koo C. K. What Constitutes a True Hyperdense Middle Cerebral Artery Sign? / C. K. Koo, E. Teasdale, K. W. Muir // Cerebrovascular Diseases. – 2000. – Vol. 10. – № 6. – P. 419-423.
77. Kumar R. Receiver operating characteristic (ROC) curve for medical researchers / R. Kumar, A. Indrayan // Indian Pediatrics. – 2011. – Vol. 48. – № 4. – P. 277-287.
78. Long-Term Prediction of Functional Outcome After Stroke Using the Alberta Stroke Program Early Computed Tomography Score in the Subacute Stage / L. D. Alexander, J. A. Pettersen, J. J. Hopyan [et al.] // Journal of Stroke and Cerebrovascular Diseases. – 2012. – Vol. 21. – № 8. – P. 737-744.
79. Mainali S. Detection of Early Ischemic Changes in Noncontrast CT Head Improved with ‘Stroke Windows’ / S. Mainali, M. Wahba, L. Elijovich // ISRN Neuroscience. – 2014. – Vol. 2014. – P. 1-4.
80. Matrix metalloproteinase-9 and urokinase plasminogen activator mediate interleukin-1-induced neurotoxicity / P. Thornton, E. Pinteaux, S. M. Allan, N. J. Rothwell // Molecular and Cellular Neuroscience. – 2008. – Vol. 37. – № 1. – P. 135-142.
81. Minds treating brains: understanding the interpretation of non-contrast CT ASPECTS in acute ischemic stroke / A. T. Wilson, S. Dey, J. W. Evans [et al.] // Expert Review of Cardiovascular Therapy. – 2018. – Vol. 16. – № 2. – P. 143-153.
82. Multi-center prediction of hemorrhagic transformation in acute ischemic stroke using permeability imaging features / F. Scalzo, J. R. Alger, X. Hu [et al.] // Magnetic Resonance Imaging. – 2013. – Vol. 31. – № 6. – P. 961-969.
83. Multiclass Support Vector Machine-Based Lesion Mapping Predicts Functional Outcome in Ischemic Stroke Patients / N. D. Forkert, T. Verleger, B. Cheng [et al.] // PLOS ONE. – 2015. – Vol. 10. – № 6. – P. e0129569.

84. Multiple Sclerosis Identification by 14-Layer Convolutional Neural Network with Batch Normalization, Dropout, and Stochastic Pooling / S.-H. Wang, C. Tang, J. Sun [et al.] // Frontiers in Neuroscience. – 2018. – Vol. 12. – P. 818.
85. Navarro-Orozco D. Neuroanatomy, Middle Cerebral Artery / D. Navarro-Orozco, J. C. Sánchez-Manso. – Text: electronic // StatPearls. – Treasure Island (FL): StatPearls Publishing, 2023. – URL: http://www.ncbi.nlm.nih.gov/books/NBK526002/ (date accessed: 22.08.2023).
86. Neutrophil Cerebrovascular Transmigration Triggers Rapid Neurotoxicity through Release of Proteases Associated with Decondensed DNA / C. Allen, P. Thornton, A. Denes [et al.] // The Journal of Immunology. – 2012. – Vol. 189. – № 1. – P. 381-392.
87. Nogles T. E. Middle Cerebral Artery Stroke / T. E. Nogles, M. A. Galuska. – Text: electronic // StatPearls. – Treasure Island (FL): StatPearls Publishing, 2023. – URL: http://www.ncbi.nlm.nih.gov/books/NBK556132/ (date accessed: 22.08.2023).
88. Obuchowski N. A. Statistical considerations for testing an AI algorithm used for prescreening lung CT images / N. A. Obuchowski, J. A. Bullen // Contemporary Clinical Trials Communications. – 2019. – Vol. 16. – P. 100434.
89. Performance of e-ASPECTS software in comparison to that of stroke physicians on assessing CT scans of acute ischemic stroke patients / C. Herweh, P. A. Ringleb, G. Rauch [et al.] // International Journal of Stroke. – 2016. – Vol. 11. – № 4. – P. 438-445.
90. Prakkamakul S. ASPECTS CT in Acute Ischemia: Review of Current Data / S. Prakkamakul, A. J. Yoo // Topics in Magnetic Resonance Imaging. – 2017. – Vol. 26. – № 3. – P. 103-112.
91. Predicting Stroke Outcome Using Clinical- versus Imaging-based Scoring System / J. H. Baek, K. Kim, Y.-B. Lee [et al.] // Journal of Stroke and Cerebrovascular Diseases. – 2015. – Vol. 24. – № 3. – P. 642-648.

92. Prediction of Hemorrhagic Transformation Severity in Acute Stroke From Source Perfusion MRI / Y. Yu, D. Guo, M. Lou [et al.] // IEEE Transactions on Biomedical Engineering. – 2018. – Vol. 65. – № 9. – P. 2058-2065.
93. Prediction of stroke thrombolysis outcome using CT brain machine learning / P. Bentley, J. Ganesalingam, A. L. Carlton Jones [et al.] // NeuroImage: Clinical. – 2014. – Vol. 4. – P. 635-640.
94. Prediction of Tissue Outcome and Assessment of Treatment Effect in Acute Ischemic Stroke Using Deep Learning / A. Nielsen, M. B. Hansen, A. Tietze, K. Mouridsen // Stroke. – 2018. – Vol. 49. – № 6. – P. 1394-1401.
95. Recommendations on telestroke in Europe / A. M. Alasheev, G. J. Hubert, G. C. Santo [et al.] // Zhurnal nevrologii i psikhiatrii im. S.S. Korsakova. – 2020. – Vol. 120. – № 3. – P. 33.
96. Reliability and accuracy of individual Alberta Stroke Program Early CT Score regions using a medical and a smartphone reading system in a telestroke network / A. J. Salazar, N. Useche, M. Granja [et al.] // Journal of Telemedicine and Telecare. – 2021. – Vol. 27. – № 7. – P. 436-443.
97. Reliability of Visual Assessment of Non-Contrast CT, CT Angiography Source Images and CT Perfusion in Patients with Suspected Ischemic Stroke / T. Van Seeters, G. J. Biessels, J. M. Niesten [et al.] // PLoS ONE. – 2013. – Vol. 8. – № 10. – P. e75615.
98. Role of imaging in early diagnosis of acute ischemic stroke: a literature review / M. A. Akbarzadeh, S. Sanaie, M. Kuchaki Rafsanjani, M.-S. Hosseini // The Egyptian Journal of Neurology, Psychiatry and Neurosurgery. – 2021. – Vol. 57. – № 1. – P. 175.
99. Sakai K. Machine learning studies on major brain diseases: 5-year trends of 2014–2018 / K. Sakai, K. Yamada // Japanese Journal of Radiology. – 2019. – Vol. 37. – № 1. – P. 34-72.
100. Samuel's Checkers Player / E. Martin, S. Kaski, F. Zheng [et al.]. – Text: electronic // Encyclopedia of Machine Learning / eds. C. Sammut, G. I. Webb. –

Boston, MA: Springer US, 2011. – P. 881-881. – URL: http://link.springer.com/10.1007/978-0-387-30164-8_740 (date accessed: 22.08.2023).

101. Schröder J. A Critical Review of Alberta Stroke Program Early CT Score for Evaluation of Acute Stroke Imaging / J. Schröder, G. Thomalla. // Frontiers in Neurology. – 2017. – Vol. 7. – URL: http://journal.frontiersin.org/article/10.3389/fneur.2016.00245/full (date accessed: 22.08.2023).

102. Shen D. Deep Learning in Medical Image Analysis / D. Shen, G. Wu, H.-I. Suk // Annual Review of Biomedical Engineering. – 2017. – Vol. 19. – № 1. – P. 221-248.

103. Streamlining door to recanalization processes in endovascular stroke therapy / A. Aghaebrahim, C. Streib, S. Rangaraju [et al.] // Journal of NeuroInterventional Surgery. – 2017. – Vol. 9. – № 4. – P. 340-345.

104. Stroke care: initial data from a county-based bypass protocol for patients with acute stroke / S. F. Zaidi, J. Shawver, A. Espinosa Morales [et al.] // Journal of Neurointerventional Surgery. – 2017. – Vol. 9. – № 7. – P. 631-635.

105. Sulcal effacement with Preserved Gray–White Junction: A Sign of Reversible Ischemia / D. C. Haussen, A. Lima, M. Frankel [et al.] // Stroke. – 2015. – Vol. 46. – № 6. – P. 1704-1706.

106. The Alberta Stroke Program Early CT Score in clinical practice: what have we learned? / V. Puetz, I. Dzialowski, M. D. Hill, A. M. Demchuk // International Journal of Stroke: Official Journal of the International Stroke Society. – 2009. – Vol. 4. – № 5. – P. 354-364.

107. The ASPECTS template is weighted in favor of the striatocapsular region / T. G. Phan, G. A. Donnan, M. Koga [et al.] // NeuroImage. – 2006. – Vol. 31. – № 2. – P. 477-481.

108. The validity and reliability of signs of early infarction on CT in acute ischaemic stroke / D. W. J. Dippel, M. Du Ry Van Beest Holle, F. Van Kooten, P. J. Koudstaal // Neuroradiology. – 2000. – Vol. 42. – № 9. – P. 629-633.
109. Tilt-Corrected Region Boundaries May Enhance the Alberta Stroke Program Early Computed Tomography Score for Less Experienced Raters / C. J. Culbertson, S. Christensen, M. Mlynash [et al.] // Journal of Stroke and Cerebrovascular Diseases. – 2020. – Vol. 29. – № 7. – P. 104820.
110. Time to Treatment with Endovascular Thrombectomy and Outcomes from Ischemic Stroke: A Meta-analysis / J. L. Saver, M. Goyal, A. Van Der Lugt [et al.] // JAMA. – 2016. – Vol. 316. – № 12. – P. 1279.
111. Use of the Alberta Stroke Program Early CT Score (ASPECTS) for assessing CT scans in patients with acute stroke / J. H. Pexman, P. A. Barber, M. D. Hill [et al.] // AJNR. American journal of neuroradiology. – 2001. – Vol. 22. – № 8. – P. 1534-1542.
112. Using the baseline CT scan to select acute stroke patients for IV-IA therapy / M. D. Hill, A. M. Demchuk, T. A. Tomsick [et al.] // AJNR. American journal of neuroradiology. – 2006. – Vol. 27. – № 8. – P. 1612-1616.
113. Validation of automated CT Alberta Stroke Program Early CT Score (ASPECTS) software for detection of early ischemic changes on non-contrast brain CT scans / L. Wolff, O. A. Berkhemer, A. C. G. M. Van Es [et al.] // Neuroradiology. – 2021. – Vol. 63. – № 4. – P. 491-498.
114. Validation of Computed Tomographic Middle Cerebral Artery 'Dot' Sign: An Angiographic Correlation Study / M. C. Leary, C. S. Kidwell, J. P. Villablanca [et al.] // Stroke. – 2003. – Vol. 34. – № 11. – P. 2636-2640.
115. Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy / P. A. Barber, A. M. Demchuk, J. Zhang, A. M. Buchan // The Lancet. – 2000. – Vol. 355. – № 9216. – P. 1670-1674.

116. Wardlaw J. M. Early Signs of Brain Infarction at CT: Observer Reliability and Outcome after Thrombolytic Treatment—Systematic Review / J. M. Wardlaw, O. Mielke // Radiology. – 2005. – Vol. 235. – № 2. – P. 444-453.

117. Wey H.-Y. A review of current imaging methods used in stroke research / H.-Y. Wey, V. R. Desai, T. Q. Duong // Neurological Research. – 2013. – Vol. 35. – № 10. – P. 1092-1102.

118. Why are acute ischemic stroke patients not receiving IV tPA? Results from a national registry / S. R. Messé, P. Khatri, M. J. Reeves [et al.] // Neurology. – 2016. – Vol. 87. – № 15. – P. 1565-1574.