

Создание коллекции МСКТ-изображений и клинических данных при острых нарушениях мозгового кровообращения

Ф.А. Шарифуллин^{1,2*}, Д.Д. Долотова³, Т.Г. Бармина¹, С.С. Петриков¹, Л.С. Коков^{1,2},
Г.Р. Рамазанов¹, Е.Р. Благосклонова³, И.В. Архипов³, И.М. Скоробогач¹, Н.Н. Черемушкин¹,
В.В. Донитова⁴, Б.А. Кобринский⁴, А.В. Гаврилов^{3,5}

Отделение лучевой диагностики

¹ ГБУЗ «Научно-исследовательский институт скорой помощи им. Н.В. Склифосовского ДЗМ»

Российская Федерация, 129090, Москва, Б. Сухаревская пл., д. 3

² Институт профессионального образования ФГАОУ ВО Первый Московский государственный медицинский университет им. И.М. Сеченова Минздрава РФ (Сеченовский Университет)

Российская Федерация, 119991, Москва, ул. Трубецкая, д. 8, стр. 2

³ ООО «Гаммамед-Софт»

Российская Федерация, 127473, Москва, 3-й Самотёчный пер., д. 11

⁴ Федеральный исследовательский центр «Информатика и управление» Российской академии наук

Российская Федерация, 119333, Москва, ул. Вавилова, д. 44, корп. 2

⁵ Научно-исследовательский институт ядерной физики им. Д.В. Скобельцына Московского государственного университета им. М.В. Ломоносова

Российская Федерация, 119991, Москва, Ленинские горы, д. 1, стр. 58

* Контактная информация: Шарифуллин Фаат Абдул-Каюмович, доктор медицинских наук, главный научный сотрудник отделения лучевой диагностики ГБУЗ «НИИ СП им. Н.В. Склифосовского ДЗМ». Email: drfaat@narod.ru

АКТУАЛЬНОСТЬ

Применение методов нейровизуализации является неотъемлемой частью процесса оказания помощи больным с острыми нарушениями мозгового кровообращения (ОНМК), при этом золотым стандартом обследования данной категории больных является компьютерная томография (КТ). Значительно расширить возможности анализа КТ-изображений возможно с помощью современных методов машинного обучения, в том числе на основе применения принципов радиомики. Однако, так как использование этих методов требует наличия больших массивов *DICOM* (*Digital Imaging and Communications in Medicine*)-изображений, их внедрение в клиническую практику ограничено проблемой набора репрезентативных выборок. Кроме того, в настоящее время в открытом доступе практически не представлены коллекции, содержащие КТ-изображения больных с ОНМК, которые были бы пригодны для машинного обучения.

ЦЕЛЬ

В связи с вышесказанным, целью данной работы являлось создание коллекции *DICOM*-изображений нативной КТ и КТ-ангиографии у пациентов с различными типами ОНМК.

МАТЕРИАЛ И МЕТОДЫ

Основой для создания коллекции стали истории болезни пациентов, госпитализированных в региональный сосудистый центр НИИ СП им. Н.В. Склифосовского. Для формирования коллекции использовалась разработанная нами ранее специализированная платформа, позволяющая вводить клинические данные о случаях ОНМК, прикреплять к каждому случаю *DICOM*-изображения проведенных исследований, а также оконтуривать и тегировать (размечать) 3D-области интереса. Для тегирования был разработан словарь, элементы которого описывают тип патологического образования, локализацию и бассейн кровоснабжения.

РЕЗУЛЬТАТЫ

В ходе работы была сформирована коллекция клинических случаев и изображений, включающая анонимизированную информацию о 220 пациентах, из них 130 — с ишемическим инсультом, 40 — с геморрагическим инсультом, а также 50 человек без цереброваскулярной патологии. Клинические данные включали сведения о типе ОНМК, наличии сопутствующих заболеваний и осложнений, длительности госпитализации, способе лечения и исходе. Всего для пациентов были введены результаты 370 исследований нативной КТ и 102 исследования КТ-ангиографии. На каждой серии изображений врачом-экспертом были оконтурены и протегированы области интереса, соответствующие прямым и косвенным признакам ОНМК.

ВЫВОД

Сформированная коллекция изображений позволит в последующем применить различные методы анализа данных и машинного обучения в решении важнейших практических задач, в том числе диагностики типа ОНМК, оценки объема поражения, прогноза степени неврологического дефицита.

Ключевые слова:

набор данных, инсульт, компьютерная томография, *DICOM*-изображения, радиомика, машинное обучение

Ссылка для цитирования

Шарифуллин Ф.А., Долотова Д.Д. Бармина Т.Г., Петриков С.С., Коков Л.С., Рамазанов Г.Р. и др. Создание коллекции МСКТ-изображений и клинических данных при острых нарушениях мозгового кровообращения. *Журнал им. Н.В. Склифосовского неотложная медицинская помощь*. 2020;9(2):231–237. <https://doi.org/10.23934/2223-9022-2020-9-2-231-237>

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов

Благодарность, финансирование

Работа выполнена при финансовой поддержке РФФИ в рамках научного проекта № 18-29-26007мк

| | |
|------|---|
| ГИ | — геморрагический инсульт |
| ИИ | — ишемический инсульт |
| КТ | — компьютерная томография |
| МРТ | — магнитно-резонансная томография |
| МСКТ | — многосрезовая компьютерная томография |
| ОНМК | — острое нарушение мозгового кровообращения |
| СППР | — система поддержки принятия решений |

| | |
|---------|---|
| ASPECTS | — <i>Alberta Stroke Program Early CT Score</i> , шкала оценки ранних КТ-признаков ишемии головного мозга |
| DICOM | — <i>Digital Imaging and Communications in Medicine</i> — медицинский отраслевой стандарт создания, хранения, передачи и визуализации цифровых медицинских изображений и документов обследованных пациентов |
| NIHSS | — шкала определения тяжести ишемического инсульта Национального института здоровья США (NIH) |

ВВЕДЕНИЕ

Острые нарушения мозгового кровообращения (ОНМК) являются одной из лидирующих причин заболеваемости, смертности и инвалидизации как в России, так и во всем мире. Ежегодно в России регистрируется более 450 000 случаев инсульта, при этом смертность в остром периоде ОНМК достигает 35%, увеличиваясь на 12–15% к концу первого года; а в течение 5 лет после инсульта умирают 44% пациентов. Постинсультная инвалидизация занимает первое место среди всех причин инвалидности и составляет 3,2 на 10 000 населения [1–5].

Одним из перспективных направлений оптимизации диагностики ОНМК является внедрение систем поддержки принятия решений (СППР), в том числе на основе применения методов машинного обучения, на этапе интерпретации радиологических изображений. Актуальность этого направления обусловлена рядом факторов. Несмотря на высокое насыщение медицинских учреждений аппаратами КТ по всей территории России, отмечается дефицит кадров. Кроме того, в крупных медицинских центрах, где круглосуточно обследуют большое число пациентов, имеет место фактор усталости и ослабления концентрации внимания. В таких ситуациях наличие СППР могло бы минимизировать влияние подобных факторов на качество оказания медицинской помощи.

Известно, что для построения любой СППР необходимо наличие репрезентативной выборки. В большинстве научных проектов авторам приходится сначала осуществить сбор данных, что требует значительных временных и организационных ресурсов. В некоторых случаях разработчики пытаются использовать изображения, накопленные в медицинском учреждении за предыдущие годы, однако работа с ними требует тщательной проверки каждого клинического случая на предмет его соответствия критериям включения и учета того фактора, что протоколы ведения больных и схемы лечения могли в течение этого времени меняться [6–7]. Иногда исследователи используют коллекции, опубликованные в открытом доступе, однако они чаще всего содержат малое количество наблюдений [8]. Кроме того, сопутствующая клиническая информация, важность которой нельзя недооценивать, может быть не представлена вообще или представлена весьма кратким списком переменных. В настоящее время в открытом доступе представлено весьма ограниченное

количество датасетов, содержащих *DICOM*-изображения КТ-исследований пациентов с ОНМК.

В связи с вышесказанным, **целью данной работы** являлось создание коллекции МСКТ-изображений и клинических данных пациентов с ОНМК.

МАТЕРИАЛ И МЕТОДЫ

Набор данных в коллекцию производился из историй болезни пациентов, госпитализированных в региональный сосудистый центр НИИ СП им. Склифосовского в период с 2016 по 2019 год. Критериями включения пациентов являлись:

- возраст старше 18 лет;
- основной диагноз: ОНМК по ишемическому (ИИ) или геморрагическому (ГИ) типу;
- наличие *DICOM*-изображений и протокола КТ, выполненной сразу после поступления пациента в учреждение; г) для пациентов с ишемическим инсультом (ИИ) — его подтверждение по данным магнитно-резонансной томографии (МРТ).

Исключались из исследования истории болезни пациентов, отвечающих следующим критериям:

- наличие тяжелой сопутствующей патологии (пневмония, онкология, травма и др.);
- в случае смертельного исхода отсутствие информации о причине смерти;
- наличие только прореженной серии *DICOM*-изображений в *PACS*-архиве медицинского учреждения.

Для группы сравнения критериями включения являлись отсутствие клинических и радиологических признаков ОНМК, травматических и других поражений головного мозга, а также наличие *DICOM*-изображений, полученных в результате проведения нативной КТ.

При создании коллекции использовалась разработанная нами ранее платформа для сбора и анализа данных и изображений на базе аппаратно-программного комплекса «Гамма-Мультивокс» [9]. Функциональность системы включает следующие основные блоки: блок работы с клиническими данными, блок работы с изображениями, блок оценки статистических характеристик и блок экспорта. Блок работы с клиническими данными включал экранные формы для ввода данных о каждом клиническом случае и ввода формализованных заключений КТ-исследований (рис. 1). О каждом пациенте была внесена следующая информация:

а) основной диагноз, длительность госпитализации и исход обращения;
 б) данные об имеющихся у пациента факторах риска (гипертоническая болезнь, атеросклероз, прием антикоагулянтов, ОНМК в анамнезе, отягощенный семейный анамнез, сахарный диабет, почечная патология, ожирение);
 в) данные о проведенном лечении (название операции и ее результат);

г) балльная оценка тяжести состояния при поступлении и выписке из стационара по следующим шкалам: модифицированная шкала Рэнкина (определение степени функциональной независимости после инсульта), шкала *NIHSS* (для оценки тяжести неврологической симптоматики) и шкала комы Глазго.

Также в платформу были внесены результаты нативной КТ и КТ-ангиографии, выполненных на томографе *Toshiba Aquilion PRIME* с толщиной среза 0,5 мм. Формализованные заключения содержали информацию о наличии прямых и косвенных признаков ОНМК, балльную оценку объема поражения по *ASPECTS* для ИИ, характеристики ликворных пространств, описание экстра- и интракраниальных артерий, тип Виллизиева круга.

Блок работы с изображениями включал инструменты фильтрации, полуавтоматического оконтуривания областей интереса и тегирования. Созданный справочник тегов предусматривал присвоение каждой области интереса тегов, описывающих тип патологического образования, его локализацию и бассейн кровоснабжения (рис. 2).

Все данные, включая *DICOM*-файлы, были анонимизированы. Хранение данных было реализовано в базе данных *Microsoft SQL*, файлы *DICOM*-изображений хранятся отдельно в файловом сетевом хранилище.

Сводная информация по распределению количественных признаков (возраст, баллы по *NIHSS*) представлена в виде значений медианы, первого и третьего квартиля (Me [Q1; Q3]), так как распределение признаков было отличным от нормального.

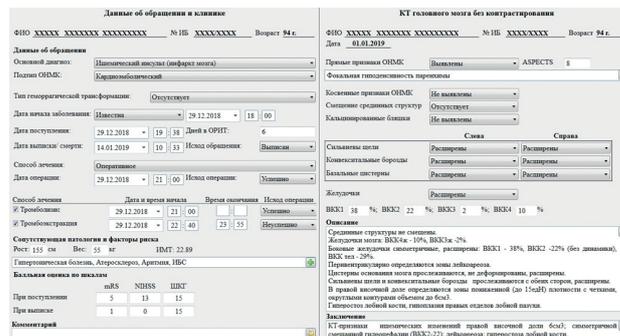


Рис. 1. Вид экранных форм для внесения клинической информации и протокола КТ-исследования
 Fig. 1. Type of screen forms for entering clinical information and protocol of CT examination

Таблица 1
 Распределение пациентов по возрасту и полу
 Table 1
 The distribution of patients by age and gender

| Группа | Вся выборка | Мужчины | Женщины |
|-------------------------|----------------------|---------------------|---------------------|
| Норма | n=50 34 [26; 51] | n=21 34 [30; 49] | n=29 34 [23; 54] |
| Ишемический инсульт | n=130 69 [57; 80] | n=68 65 [53; 77] | n=62 77 [65; 83] |
| Геморрагический инсульт | n=40 58 [49; 70] | n=28 57 [47; 70] | n=12 67 [51; 70] |

РЕЗУЛЬТАТЫ

В ходе работы была создана коллекция, включающая в себя информацию о 220 пациентах, из них 130 — с ИИ, 40 — с ГИ, а также 50 человек без цереброваскулярной патологии. Сводная информация о половозрастной структуре каждой из групп представлена в табл. 1.

Средний балл по шкале *NIHSS* при поступлении в группах с ИИ и ГИ составил 8 [3; 15] и 12 [5; 33] соответственно. Летальность в группе пациентов с ИИ

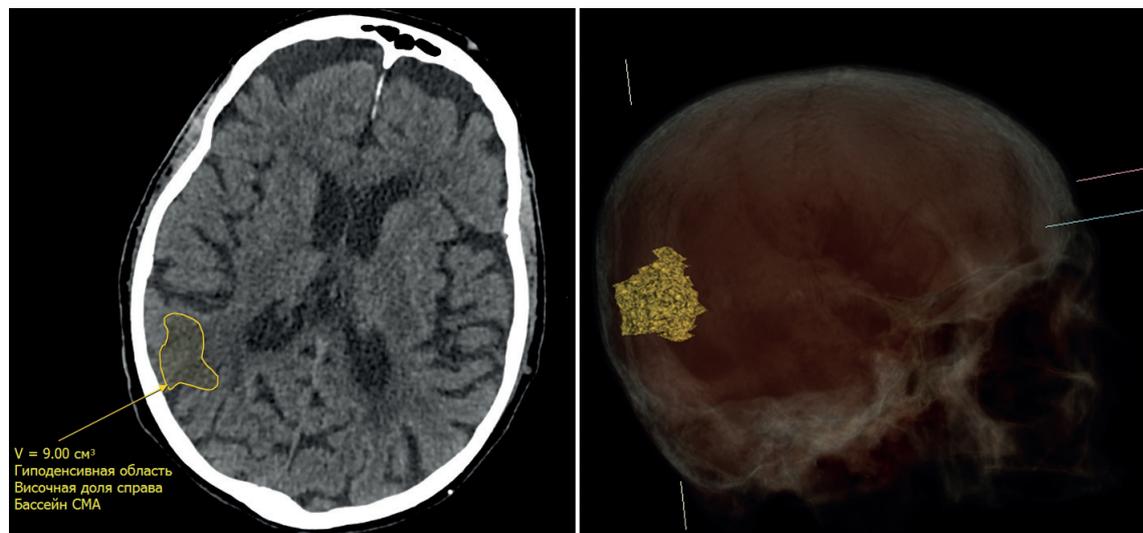


Рис. 2 Разметка и тегирование патологических областей на *DICOM*-изображениях с формированием трехмерной области интереса
 Fig. 2. Marking and tagging areas of interest on *DICOM*-images with 3D region of interest

составила 18,6%, в группе с ГИ — 50%. Распределение пациентов с ИИ по различным патогенетическим подтипам по шкале инсультов *TOAST* являлось следующим: атеротромбоэмболический — 30, кардиоэмболический — 39, лакунарный — 12, инсульт другой установленной этиологии — 4, инсульт другой неустановленной этиологии — 48.

Среди пациентов с ИИ в 18 наблюдениях был проведен тромболитизис, в 7 — тромбэкстракция, в 5 — тромбоаспирация, в 2 случаях была выполнена реваскуляризация экстракраниальных артерий. У 17 пациентов с ИИ наблюдалась геморрагическая трансформация (13%).

Всего было проведено 330 нативных КТ- исследований: у 88 пациентов из группы ИИ и у 22 пациентов из группы ГИ исследование было проведено дважды. На всех сериях была выполнена разметка выявленных прямых и косвенных признаков ОНМК (гиподенсивные области с ишемией, гиперденсивная средняя мозговая артерия, внутримозговые и внутривентрикулярные кровоизлияния). Области с кистозно-глиозными изменениями, соответствующие проявлениям «старых» ОНМК, также были оконтурены. У пациентов с ИИ в 78,5% случаев была проведена КТ-ангиография ($n=102$).

ОБСУЖДЕНИЕ

Внедрение в клиническую практику системы поддержки принятия решений, которая значительно ускорит и увеличит эффективность оказания медицинской помощи при ОНМК, является важной задачей. Автоматический анализ данных нейровизуализации позволит в кратчайшие сроки проводить раннюю дифференциальную диагностику, прогнозировать возможный исход заболевания и предоставить рекомендации наиболее эффективного способа лечения индивидуально для каждого пациента. Наличие репрезентативной выборки, состоящей из большого количества структурированных и достоверных данных, является основой для реализации различных методов анализа медицинских изображений, в том числе с применением машинного обучения. Эффективность и точность моделей напрямую зависит также от качества исходных данных (обучающей выборки) и требует их тщательной предварительной обработки.

При проведении исследований возникает необходимость поиска случаев, соответствующих выбранным критериям, в локальных хранилищах одного медицинского учреждения с последующим трудоемким процессом разметки изображений экспертами вручную. Именно по этой причине используемые при обучении систем выборки чаще всего содержат менее 100 диагностических серий [10].

Несмотря на актуальность автоматизации процесса диагностики ОНМК, наличие соответствующих коллекций изображений в открытом доступе в мире немногочисленно, а в России подобные проекты вовсе отсутствуют, что неизбежно приводит к потере точности работы моделей, полученных на зарубежных популяционных данных. Наиболее распространенной целью организации общедоступных датасетов в основном является содействие коллективам в создании и совершенствовании алгоритмов автоматической сегментации объема поражения, и зачастую коллекции представлены данными диагностики ГИ [6, 11] или изображениями с уже развившимися крупными

зонами ишемии (не включают случаи в острейшем периоде), и практически не содержат клинической информации о пациенте. Некоторые датасеты не включают разметку очагов поражения, или размечены с помощью алгоритмов автоматической сегментации без участия экспертов, что делает их использование в качестве обучающей выборки невозможным без предварительной обработки. Несомненно, в наибольшей степени актуальны системы экстренной и ранней диагностики пациентов с ИИ, однако на данный момент сложно найти данные для их разработки. Более того, такие крупные проекты, посвященные ишемическим поражениям, как *ENIGMA Stroke Recovery* [12], *ATLAS* [13], а также конкурс *ISLES* [8], проводимый в 2016–2018 гг., сосредоточены исключительно на сборе данных магнитно-резонансной томографии и некоторых типов функциональных исследований головного мозга. Изображения указанных модальностей, несомненно, обладают большой информативностью для принятия клинических решений, однако не являются широко распространенными методами диагностики ОНМК. Еще одним распространенным недостатком общедоступных наборов изображений является их размещение в сжатом виде или предварительное конвертирование в форматы с потерей качества и метаданных, что значительно ограничивает исследователей в выборе методов анализа.

В данной статье описан подход к формированию коллекции МСКТ-изображений и клинических данных при острых нарушениях мозгового кровообращения, который позволит значительно расширить круг задач, решаемых СППР на основе анализа медицинских изображений. Специфика собранной клинической информации, наличие данных нейровизуализации для одного пациента в динамике, разметка прямых и косвенных признаков поражений, их четкая идентификация по типу и локализации позволят выявить статистически значимую взаимосвязь между биомаркерами изображений, клиническими данными и эффектами лечения. Размещение коллекции в открытом доступе будет способствовать привлечению специалистов различного профиля к участию в создании крупномасштабного систематизированного отечественного банка данных, позволит находить и использовать выборку с необходимым набором параметров при обучении и тестировании алгоритмов сегментации и анализа изображений при такой социально-значимой патологии как ОНМК.

ЗАКЛЮЧЕНИЕ

В результате данной работы была создана коллекция изображений мультиспиральной компьютерной томографии и клинических данных пациентов с острым нарушением мозгового кровообращения. В состав коллекции вошли анонимизированные данные о 130 пациентах с ишемическим инсультом, 40 пациентах с геморрагическим инсультом и 50 пациентах без цереброваскулярной патологии. Были собраны *DICOM*-изображения 330 исследований нативной компьютерной томографии и 102 исследования компьютерной томографической ангиографии. У 110 пациентов компьютерное томографическое исследование было выполнено в динамике. Все выявленные на изображениях прямые и косвенные признаки острого нарушения мозгового кровообращения были оконтурены экспертами, каждой трехмерной области интереса были

присвоены теги, описывающие тип патологического образования, его локализацию и бассейн кровоснабжения. Собранные данные могут быть использованы при построении систем поддержки принятия врачебных решений, в том числе на основе методов машинного обучения и анализа биомаркеров изображений, в решении таких важнейших практических задач, как

дифференциальная диагностика типов острого нарушения мозгового кровообращения, автоматическое определение объема области поражения, оценка рисков геморрагической трансформации, прогноз исхода клинического случая и степени неврологического дефицита. В последующем планируется публикация собранной коллекции в открытом доступе.

ЛИТЕРАТУРА

1. Корниенко В.Н., Пронина И.Н. (ред.) *Диагностическая нейрорадиология*. Москва: Издательство: "Т. М. Андреева"; 2006.
2. Сулина З.А., Пирадов М.А. (ред.) *Инсульт: диагностика, лечение, профилактика*. Москва: МЕДпресс-информ; 2008.
3. Пирадов М.А., Крылов В.В., Белкин А.А., Петриков С.С. Инсульты. В кн.: Б.Р. Гельфанда, И.Б. Заболотских (ред.) *Интенсивная терапия. Национальное руководство: краткое издание*. 2-е изд., перераб. и доп. Москва: ГЭОТАР-Медиа; 2017. Гл. 2. с. 288–309.
4. Сулина З.А., Варакин Ю.Я. Клинико-эпидемиологические исследования – перспективное направление изучения цереброваскулярной патологии (сообщение первое). *Анналы клинической и экспериментальной неврологии*. 2009; 3(3): 4–11.
5. Верещагин Н.В., Пирадов М.А., Сулина З.А. Принципы диагностики и лечения больных в остром периоде инсульта. *Атмосфера. Нервные болезни*. 2002;(1):8–14.
6. Chilamkurthy S, Ghosh R, Tanamala S, Biviji M, Campeau NG, Venugopal VK, et al. Development and Validation of Deep Learning Algorithms for Detection of Critical Findings in Head CT Scans. arXiv:1803.05854 [cs. CV]. 2018. URL: <https://arxiv.org/abs/1803.05854> [Дата обращения 21.05.2020].
7. Hajimani E, Ruano MG, Ruano AE. An intelligent support system for automatic detection of cerebral vascular accidents from brain CT images. *Comput Methods Programs Biomed*. 2017;146:109–123. PMID: 28688480 <https://doi.org/10.1016/j.cmpb.2017.05.005>
8. ISLES: Ischemic Stroke Lesion Segmentation Challenge. URL: <http://www.isles-challenge.org/> [Дата обращения 21.05.2020].
9. Dolotova D, Donitova V, Arhipov I, Sharifullin F, Zagriazkina T, Kobrinskii B, et al. A Platform for Collection and Analysis of Image Data on Stroke. *Stud Health Technol Inform*. 2019;262:312–315. PMID: 31349330 <https://doi.org/10.3233/SHT1190081>
10. Mikhail P, Le MGD, Mair G. Computational Image Analysis of Nonenhanced Computed Tomography for Acute Ischaemic Stroke: A Systematic Review. *J Stroke Cerebrovasc Dis*. 2020;29(5):104715. doi:10.1016/j.jstrokecerebrovasdis.2020.104715
11. Hssayeni M. Computed Tomography Images for Intracranial Hemorrhage Detection and Segmentation (version 1.3.1). *PhysioNet*. 2020. <https://doi.org/10.13026/4nae-zg36>.
12. *Enigma stroke recovery*. URL: <http://enigma.ini.usc.edu/ongoing/enigma-stroke-recovery/> [Дата обращения 21.05.2020].
13. Liew S, Anglin J, Banks N, et al. A large, open source dataset of stroke anatomical brain images and manual lesion segmentations. *Sci Data*. 2018;5:180011. <https://doi.org/10.1038/sdata.2018.11>

REFERENCES

1. Kornienko VN, Pronina IN. (eds.) *Diagnosticheskaya neyroradiologiya*. Moscow: Izdatel'stvo:T. M. Andreeva Publ.; 2006. (In Russ.)
2. Suslina ZA, Piradov MA. (eds.) *Insul't: diagnostika, lechenie, profilaktika*. Moscow: MEDpress-inform Publ.; 2008. (In Russ.)
3. Piradov MA, Krylov VV, Belkin AA, Petrikov SS. Insul'ty. In: Gelf'fanda BR, Zabolotskikh IB (eds.) *Intensivnaya terapiya*. 2nd ed., rev. and exp. Moscow: GEOTAR-Media Publ.; 2017. Ch. 2. pp. 288–309. (In Russ.)
4. Suslina ZA, Varakin YuYa. Clinico-epidemiological studies a perspective direction for investigation of cerebrovascular disorders. *Annals of Clinical and Experimental Neurology*. 2009;3(3):4–11. (In Russ.)
5. Vereshchagin NV, Piradov MA, Suslina ZA. Printsipy diagnostiki i lecheniya bol'nykh v ostrom periode insul'ta. *Atmosfera. Nervnye bolezni*. 2002;(1):8–14. (In Russ.)
6. Chilamkurthy S, Ghosh R, Tanamala S, Biviji M, Campeau NG, Venugopal VK, et al. Development and Validation of Deep Learning Algorithms for Detection of Critical Findings in Head CT Scans. arXiv:1803.05854 [cs. CV]. 2018. Available at: <https://arxiv.org/abs/1803.05854> [Accessed 21 May 2020].
7. Hajimani E, Ruano MG, Ruano AE. An intelligent support system for automatic detection of cerebral vascular accidents from brain CT images. *Comput Methods Programs Biomed*. 2017;146:109–123. PMID: 28688480 <https://doi.org/10.1016/j.cmpb.2017.05.005>
8. ISLES: Ischemic Stroke Lesion Segmentation Challenge. Available at: <http://www.isles-challenge.org/> [Accessed 21 May 2020].
9. Dolotova D, Donitova V, Arhipov I, Sharifullin F, Zagriazkina T, Kobrinskii B, et al. A Platform for Collection and Analysis of Image Data on Stroke. *Stud Health Technol Inform*. 2019;262:312–315. PMID: 31349330 <https://doi.org/10.3233/SHT1190081>
10. Mikhail P, Le MGD, Mair G. Computational Image Analysis of Nonenhanced Computed Tomography for Acute Ischaemic Stroke: A Systematic Review. *J Stroke Cerebrovasc Dis*. 2020;29(5):104715. doi:10.1016/j.jstrokecerebrovasdis.2020.104715
11. Hssayeni M. Computed Tomography Images for Intracranial Hemorrhage Detection and Segmentation (version 1.3.1). *PhysioNet*. 2020. <https://doi.org/10.13026/4nae-zg36>.
12. *Enigma stroke recovery*. Available at: <http://enigma.ini.usc.edu/ongoing/enigma-stroke-recovery/> [Accessed 21 May 2020].
13. Liew S, Anglin J, Banks N, et al. A large, open source dataset of stroke anatomical brain images and manual lesion segmentations. *Sci Data*. 2018;5:180011. <https://doi.org/10.1038/sdata.2018.11>

ИНФОРМАЦИЯ ОБ АВТОРАХ

- Шарифуллин Фаат Абдул-Каюмович** доктор медицинских наук, главный научный сотрудник отделения лучевой диагностики ГБУЗ «НИИ СП им. Н.В. Склифосовского ДЗМ»; <https://orcid.org/0000-0001-7483-7899>, drfaat@narod.ru; 10%: организация сбора данных, написание текста рукописи
- Долотова Дарья Дмитриевна** кандидат медицинских наук, ведущий научный сотрудник ООО «Гаммамед-Софт»; <https://orcid.org/dariadolotova@gmail.com>; 10%: разработка дизайна исследования, анализ собранных данных, написание текста рукописи
- Бармина Татьяна Геннадьевна** кандидат медицинских наук, старший научный сотрудник отделения лучевой диагностики ГБУЗ «НИИ СП им. Н.В. Склифосовского ДЗМ»; <https://orcid.org/0000-0003-2418-680X>, barminat@inbox.ru; 10%: анализ КТ-изображений
- Петриков Сергей Сергеевич** член-корреспондент РАН, доктор медицинских наук, директор ГБУЗ «НИИ СП им. Н.В. Склифосовского ДЗМ»; <https://orcid.org/0000-0003-3292-8789>, petrikovss@sklif.mos.ru; 8%: окончательное утверждение публикации рукописи

- Коков Леонид Сергеевич** член-корреспондент РАН, профессор, доктор медицинских наук, руководитель отделения лучевой диагностики ГБУЗ «НИИ СП им. Н.В. Склифосовского ДЗМ»; <https://orcid.org/0000-0002-3167-3692>, lskokov@mail.ru;
8%: редактирование текста рукописи
- Рамазанов Ганипа Рамазанович** заведующий научным отделением неотложной неврологии и восстановительного лечения ГБУЗ «НИИ СП им. Н.В. Склифосовского ДЗМ»; <https://orcid.org/0000-0001-6824-4114>, ramazanovgr@sklif.mos.ru;
8%: организация сбора данных
- Благосклонова Евгения Романовна** научный сотрудник ООО «Гаммамед-Софт»; <https://orcid.org/0000-0002-4678060X>, evromsp@gmail.com;
8%: разработка программного обеспечения для формирования коллекции, сбор данных, анализ собранных данных
- Архипов Иван Владимирович** ведущий программист ООО «Гаммамед-Софт»; <https://orcid.org/0000-0003-4278-2285>, arkhivania@gmail.com;
8%: разработка программного обеспечения для формирования коллекции и работы с изображениями
- Скоробогач Иван Михайлович** врач ординатор ГБУЗ «НИИ СП им. Н.В. Склифосовского ДЗМ»; <https://orcid.org/0000-0002-5428-6687>, dr.skoroboga@gmail.com;
6%: сбор данных
- Черемушкин Николай Николаевич** врач ординатор ГБУЗ «НИИ СП им. Н.В. Склифосовского ДЗМ»; <https://orcid.org/0000-0003-0208-3033>, tiforge.invest@gmail.com;
6%: сбор данных
- Донитова Виктория Владимировна** научный сотрудник ФИЦ «Информатика и управление» РАН; <https://orcid.org/0000-0002-7838-584X>, vdonitova@gmail.com;
6%: анализ собранных данных
- Кобринский Борис Аркадьевич** доктор медицинских наук, профессор, заведующий отделом систем поддержки принятия клинических решений Института современных информационных технологий в медицине ФИЦ «Информатика и управление» РАН; <https://orcid.org/0000-0002-3459-8851>, kba_05@mail.ru;
6%: разработка программного обеспечения для формирования коллекции, сбор данных, анализ собранных данных
- Гаврилов Андрей Васильевич** кандидат технических наук, заведующий лабораторией медицинских компьютерных систем «НИИ ЯФ им. Д.В. Скобелыцына МГУ им. М.В. Ломоносова»; <https://orcid.org/0000-0002-7838-584X>, agavrilov49@gmail.com;
6%: разработка концепции и дизайна исследования

Received on 19.11.2019

Accepted on 24.03.2020

Поступила в редакцию 19.11.2019

Принята к печати 24.03.2020

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

F.A. Sharifullin^{1,2*}, D.D. Dolotova³, T.G. Barmina¹, S.S. Petrikov¹, L.S. Kokov^{1,2}, G.R. Ramazanov¹, Y.R. Blagosklonova³, I.V. Arkhipov³, I.M. Skorobogach¹, N.N. Cheremushkin¹, V.V. Donitova⁴, B.A. Kobrinski⁴, A.V. Gavrilov^{3,5}

Department of Diagnostic Radiology

¹ N.V. Sklifosovsky Research Institute for Emergency Medicine

3 B. Suharevskaya Sq., Moscow 129090, Russian Federation

² I.M. Sechenov First Moscow State Medical University of the Ministry of Health of the Russian Federation (Sechenov University)

8 b. 2 Trubetskaya St., Moscow 119991, Russian Federation

³ Gammamed-Soft, LLC

11 3th Samotechny Pereulok, Moscow 127473, Russian Federation

⁴ Federal Research Center "Informatics and Management" of the Russian Academy of Sciences

44 b. 2 Vavilova St., Moscow 119333, Russian Federation

⁵ D.V. Skobel'syn Research Institute of Nuclear Physics, M.V. Lomonosov Moscow State University

1 b. 58 Leninskiye Gory, Moscow 119991, Russian Federation

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

BACKGROUND The use of neuroimaging methods is an integral part of the process of assisting patients with acute cerebrovascular events (ACVE), and computed tomography (CT) is the «gold standard» for examining this category of patients. The capabilities of the analysis of CT images may be significantly expanded with modern methods of machine learning including the application of the principles of radiomics. However, since the use of these methods requires large arrays of DICOM (Digital Imaging and Communications in Medicine)-images, their implementation into clinical practice is limited by the lack of representative sample sets. In addition, at present, collections (datasets) of CT images of stroke patients, that are suitable for machine learning, are practically not available in the public domain.

AIM OF STUDY Regarding the aforesaid, the aim of this work was to create a DICOM images dataset of native CT and CT-angiography of patients with different types of stroke.

MATERIAL AND METHODS The collection was based on the medical cases of patients hospitalized in the Regional Vascular Center of the N.V. Sklifosovsky Research Institute for Emergency Medicine. We used a previously developed specialized platform to enter clinical data on the stroke cases, to attach CT DICOM-images to each case, to contour 3D areas of interest, and to tag (label) them. A dictionary was developed for tagging, where elements describe the type of lesion, location, and vascular territory.

RESULTS A dataset of clinical cases and images was formed in the course of the work. It included anonymous information about 220 patients, 130 of them with ischemic stroke, 40 with hemorrhagic stroke, and 50 patients without cerebrovascular disorders. Clinical data included information about type of stroke, presence of concomitant diseases and complications, length of hospital stay, methods of treatment, and outcome. The results of 370 studies of native CT and 102 studies of CT-angiography were entered for all patients. The areas of interest corresponding to direct and indirect signs of stroke were contoured and tagged by radiologists on each series of images.

CONCLUSION The resulting collection of images will enable the use of various methods of data analysis and machine learning in solving the most important practical problems including diagnosis of the stroke type, assessment of lesion volume, and prediction of the degree of neurological deficit.

Keywords: dataset, stroke, computed tomography, DICOM-images, radiomics, machine learning

For citation Sharifullin FA, Dolotova DD, Barmina TG, Petrikov SS, Kokov LS, Ramazanov GR, et al. Creation of a Dataset of MSCT-Images and Clinical Data for Acute Cerebrovascular Events. *Russian Sklifosovsky Journal of Emergency Medical Care*. 2020;9(2):231–237. <https://doi.org/10.23934/2223-9022-2020-9-2-231-237> (in Russ.)

Conflict of interest Authors declare lack of the conflicts of interests

Acknowledgments, sponsorship This work was financially supported by the Russian Federal Property Fund in the framework of the scientific project No. 18-29-26007mk

Affiliations

| | |
|--------------------------|--|
| Faat A. Sharifullin | Dr. Med. Sci., Chief Researcher, Department of Diagnostic Radiology, N.V. Sklifosovsky Research Institute for Emergency Medicine; https://orcid.org/0000-0001-7483-7899 , drfaat@narod.ru ; 10%, organizing data collection, writing a manuscript |
| Darya D. Dolotova | Cand. Med. Sci., Leading Research Assistant, Gammamed-Soft LLC; https://orcid.org/ , dariadolotova@gmail.com ; 10%, research design development, analysis of collected data, manuscript text writing |
| Tatyana G. Barmina | Cand. Med. Sci., Senior Researcher, Department of Diagnostic Radiology, N.V. Sklifosovsky Research Institute for Emergency Medicine; https://orcid.org/0000-0003-2418-680X , barminat@inbox.ru ; 10%, analysis of CT images |
| Sergey S. Petrikov | Corresponding Member of the Russian Academy of Sciences, Dr. Med. Sci., Director of N.V. Sklifosovsky Research Institute for Emergency Medicine; https://orcid.org/0000-0003-3292-8789 , petrikovss@sklif.mos.ru ; 8%, final approval of manuscript publication |
| Leonid S. Kokov | Corresponding Member of the Russian Academy of Sciences, Professor, Dr. Med. Sci., Head of Department of Diagnostic Radiology, N.V. Sklifosovsky Research Institute for Emergency Medicine; https://orcid.org/0000-0002-3167-3692 , lskokov@mail.ru ; 8%, manuscript text editing |
| Ganipa R. Ramazanov | Head of the Scientific Department of Emergency Neurology and Reconstructive Treatment of N.V. Sklifosovsky Research Institute for Emergency Medicine; https://orcid.org/0000-0001-6824-4114 , ramazanovgr@sklif.mos.ru ; 8%, data collection organization |
| Evgenia R. Blagosklonova | Researcher, Gammamed-Soft LLC; https://orcid.org/0000-0002-4678060H , evromsp@gmail.com ; 8%, software development for the collection, data collection, analysis of data collected |
| Ivan V. Arkhipov | Leading Programmer, Gammamed-Soft LLC; https://orcid.org/0000-0003-4278-2285 , arkhivania@gmail.com ; 8%, software development for the collection and working with images |
| Ivan M. Skorobogach | Resident of the N.V. Sklifosovsky Research Institute for Emergency Medicine; https://orcid.org/0000-0002-5428-6687 , dr.skoroboga@gmail.com ; 6%, data collection |
| Nikolay N. Cheremushkin | Resident of the N.V. Sklifosovsky Research Institute for Emergency Medicine of the Moscow Health Department; https://orcid.org/0000-0003-0208-3033 , tiforge.invest@gmail.com ; 6%, data collection |
| Victoria V. Donitova | Researcher, FRC Computer Science and Control of RAS; https://orcid.org/0000-0002-7838-584H , vdonitova@gmail.com ; 6%, analysis of collected data priority |
| Boris A. Kobriniski | Dr. Med. Sci., Professor, Head of the Department of Support Systems for Clinical Decisions, Institute of Modern Information Technologies in Medicine, FRC Computer Science and Control of RAS; https://orcid.org/0000-0002-3459-8851 , kba_05@mail.ru ; 6%, software development for the collection, data collection, analysis of collected data |
| Andrey V. Gavrilov | Cand. Tec. Sci., Head of the Laboratory of Medical Computer Systems, D.V. Skobeltsyn Research Institute of Nuclear Physics of M.V. Lomonosov Moscow State University; https://orcid.org/0000-0002-7838-584X , agavrilov49@gmail.com ; 6%, research concept and design development |