

Совместный анализ рентгенологических протоколов и компьютерных томограмм для автоматического уточнения патологических состояний головного мозга

Ю.Д. Агафонова¹, А.В. Гайдель^{1,2}, П.М. Зельтер³, А.В. Капишников³,
А.В. Кузнецов^{1,4,5}, Е. Н. Суровцев³, А.В. Никоноров^{1,2}

¹ Самарский национальный исследовательский университет имени академика С.П. Королёва, 443086, Россия, г. Самара, Московское шоссе, д. 34;

² ИСОИ РАН – филиал ФНИЦ «Кристаллография и фотоника» РАН, 443001, Россия, г. Самара, ул. Молодогвардейская, д. 151;

³ ФГБОУ ВО СамГМУ Минздрава России, 443099, Россия, г. Самара, ул. Чапаевская, д. 89;

⁴ Sber AI, 121170, Moscow, Russia, Kutuzovsky prospekt, 32 building 2;

⁵ Artificial Intelligence Research Institute (AIRI), 105064, Moscow, Russia, Nizhniy Susalnyy pereulok, 5

Аннотация

Рассматривается задача валидации радиологических медицинских протоколов и изображений компьютерной томографии для автоматизированного анализа состояния головного мозга. Предлагается два метода решения задачи: метод на основе мультимодальной модели ruCLIP и метод, основанный на совместном использовании двух отдельных классификаторов – для текстового отчета и для изображения КТ головного мозга. Обсуждаются способы оценки полученных результатов. Предложенные подходы позволяют верно классифицировать на 15 возможных диагнозов 99,6 % радиологических отчетов из контрольной выборки.

Ключевые слова: глубокое обучение, компьютерная томография, автоматизация диагностики, распознавание образов, обработка естественного языка.

Цитирование: Агафонова, Ю.Д. Совместный анализ рентгенологических протоколов и компьютерных томограмм для автоматического уточнения патологических состояний головного мозга / Ю.Д. Агафонова, А.В. Гайдель, П.М. Зельтер, А.В. Капишников, А.В. Кузнецов, Е.Н. Суровцев, А.В. Никоноров // Компьютерная оптика. – 2023. – Т. 47, № 1. – С. 152-159. – DOI: 10.18287/2412-6179-CO-1201.

Citation: Agafonova YD, Gaidel AV, Zelter PM, Kapishnikov AV, Kuznetsov AV, Surovtsev EN, Nikonorov AV. Joint analysis of radiological reports and CT images for automatic validation of pathological brain conditions. Computer Optics 2023; 47(1): 152-159. DOI: 10.18287/2412-6179-CO-1201.

Введение

В настоящее время всё активнее развиваются методы автоматизированной медицинской диагностики заболеваний, поскольку они позволяют повысить скорость и эффективность работы медицинских учреждений по всему миру [1]. Для разработки систем поддержки принятия решений при медицинской диагностике требуются большие объемы данных, которые зачастую вручную проходят предварительную обработку и разметку. Автоматизация этого процесса могла бы существенно расширить возможности разработчиков подобных информационных систем.

Между тем медицинские лаборатории по всему миру хранят огромные объемы данных, включая радиологические изображения различной природы, в специальных базах данных, таких как PACS [2]. Эти данные хранятся в изначальном виде, без какой бы то ни было предварительной обработки и разметки, но зачастую изображения в таких базах сопровождаются радиологическими отчетами, написанными профессиональными радиологами при первичном анализе исследования.

Автоматическое сопоставление изображений в таких базах с радиологическими отчетами могло бы позволить уточнять имеющиеся диагнозы в режиме реального времени: если система обнаруживает несовпадение изображения и текста к нему, то она сигнализирует об этом с целью получения мнения ещё одного медицинского специалиста. Это могло бы значительно повысить эффективность диагностики, не требуя при этом значительных массивов размеченных данных для обучения.

Существует множество программ и систем, обеспечивающих классификацию медицинских данных по изображениям либо на основе медицинских протоколов.

Так, в [3] упоминается программная система для автоматизированного мониторинга данных в произвольном формате в медицинской информационной системе. Данная система использует инструмент обработки медицинского языка и правила, полученные на основе статистического анализа базы данных, для обработки протоколов о рентгенографии грудной клетки (СХР) в произвольном тексте и выявления протоколов, описывающих новые или расширяющиеся новообразования, с целью мониторинга состояния

пациента. Однако данный метод опирается только на классификацию самих протоколов, не анализируя соответствующие им изображения.

Существуют также программы, которые классифицируют в основном изображения. В работе [4] упоминается классификация магнитно-резонансных томограмм головного мозга с помощью методов глубокого обучения. В качестве методов глубокого обучения используются методы на основе свёрточной нейронной сети (CNN), глубокой нейронной сети (DNN) и такие архитектуры нейронных сетей, как LeNet, AlexNet, ResNet. Особенность данного метода заключается в том, что при классификации учитываются дополнительные параметры, такие как возраст и пол. В данной работе изображения разбиваются всего на два класса: норма и патология, то есть дифференциальный диагноз не ставится.

В источнике [5] поднимается еще одна важная проблема – выбор конкретных двумерных срезов из всего множества изображений, которые появляются в результате проведения МРТ-диагностики. В данной работе также поднимается вопрос по выбору двумерных срезов из трехмерных. По исследованию можно сделать вывод о том, что авторы количественно оценили эффект утечки данных, вызванный разделением данных 3D-МРТ на основе уровня 2D-срезов, с использованием трех 2D-моделей CNN для классификации пациентов с болезнью Альцгеймера и болезнью Паркинсона. Однако данный метод решает только предварительный отбор используемых снимков для решения, при этом метод достаточно объемный, так как содержит множество операций для предобработки изображений, что негативно сказывается на скорости получения ответа от системы.

Стоит отметить, что в медицинских наблюдениях важно оценивать не только состояние пациентов на данный момент времени, но и отслеживать динамику изменения их состояния и проблемных областей. Однако пациентам зачастую не проводятся контрольные исследования или они проводятся нерегулярно, отчасти из-за отсутствия инструментов для лечения и времени, необходимого для обновления результатов исследования. Авторы [6] предлагают на основе обработки естественного языка (NLP) извлекать информацию из клинических документов с произвольным текстом. Идея подхода заключается в обнаружении отчёта, в котором описывается впервые выявленная опухоль, однако дифференциальная диагностика не производится.

Нельзя не отметить, что достоверная классификация изображений и медицинских протоколов важна не только для постановки диагноза, но и для поиска похожих клинических случаев, и для получения релевантных документов. Клинические медицинские записи содержат большое количество информации, которая обычно пишется в свободной текстовой форме и без лингвистического стандарта. Существуют алго-

ритмы, позволяющие восстанавливать знания из текстовой информации в медицинских документах [7]. То есть данные алгоритмы помогают искать похожие клинические случаи на основе различных медицинских протоколов от разных врачей, несмотря на индивидуальный стиль написания протоколов и различные сокращения. Однако авторы данных алгоритмов не уточняют точную архитектуру использованных нейронных сетей.

В работе [8] совместный анализ радиологических отчётов и изображений использовался для выявления патологических изменений лёгких. Было показано, что использование информации о локализации патологии из радиологических отчётов позволяет повысить эффективность обнаружения эмфиземы.

Целью работы является повышение эффективности диагностики патологических изменений головного мозга за счет уточнения диагнозов с помощью автоматического сопоставления рентгенологических протоколов и компьютерных томограмм.

1. Совместная классификация текстов и изображений

В качестве базового подхода к совместной классификации текстов и изображений можно предложить классифицировать отдельно изображения и отдельно текстовые описания диагнозов, с последующей оценкой соответствия между результатом предсказания класса для изображений и для текстов.

Для решения этой задачи был использован набор данных из 978 изображений компьютерной томографии головного мозга и 978 соответствующих им текстовых радиологических отчётов. Изображения были получены в ходе клинической практики сотрудниками ФГБОУ ВО СамГМУ Минздрава России. На рис. 1 представлены примеры изображений из выборки.

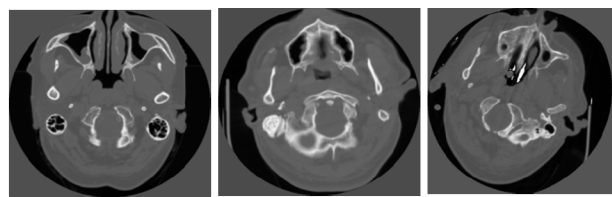


Рис. 1. Пример изображений из выборки

Изображения представляли собой избранные срезы компьютерной томографии мозга с медианным разрешением 512 на 512 отсчётов и 256 каналами яркости. Тексты отчётов были написаны в свободной форме на русском языке и содержали различное количество слов.

По описаниям изображения были разделены на классы. Классом является диагноз, который был указан пациенту в заключении рентгенолога. В том случае, если диагнозов было несколько, то брался первый диагноз из заключения. Таким образом, все изображения были поделены на 15 классов, они приведены в табл. 1.

Табл. 1. Количество изображений каждого класса

Наименование класса	Количество изображений
Патологии не выявлено	179
Ишемический инсульт	165
Дисциркуляторная энцефалопатия	160
Отек головного мозга	91
Геморрагический инсульт	82
Субарахноидальное кровоизлияние	54
Черепно-мозговая травма	50
Положительная динамика	44
Ликворная киста	42
Лакунарный инсульт	28
Атрофия головного мозга	25
Гидроцефалия	16
Эпидуральная гематома	15
Аневризма	14
Объемное образование головного мозга	13

Всё множество имеющихся пациентов Ω было разбито на обучающую выборку U и контрольную выборку \tilde{U} в соотношении примерно 3:1. Для каждого пациента $\omega \in \Omega$ в выборке содержалось изображение компьютерной томографии головного мозга этого пациента $x_\omega(n, m): [0; N-1] \times [0; M-1] \cap \mathbf{Z}^2 \rightarrow [0; Q-1] \cap \mathbf{Z}$, текст радиологического отчёта $y_\omega(t): [0; |y_\omega| - 1] \cap \mathbf{Z} \rightarrow \mathbf{W}$ и диагноз $z_\omega \in [0; L-1] \cap \mathbf{Z}$, где N – высота изображения, M – ширина изображения, $Q = 256$ – количество уровней яркости, $L = 15$ – количество различных диагнозов, $|y_\omega|$ – количество слов в отчёте для пациента ω , \mathbf{Z} – множество целых чисел, \mathbf{W} – множество возможных слов в языке. Под словами понимаются последовательности подряд идущих букв, отделённых друг от друга любыми другими символами.

Классификатор изображений – это функция $\Phi(x): \{x_\omega | \omega \in \Omega\} \rightarrow [0; L-1] \cap \mathbf{Z}$, которая ставит в соответствие изображению компьютерной томографии некоторый номер диагноза. Классификатор текстов – это аналогичная функция $\Psi(y): \{y_\omega | \omega \in \Omega\} \rightarrow [0; L-1] \cap \mathbf{Z}$, которая ставит в соответствие тексту радиологического отчёта некоторый номер диагноза. И ту и другую функцию предлагается реализовать в виде нейронной сети с использованием подходов глубокого обучения. Разумеется, для обучения нейронных сетей следует использовать только обучающую выборку U , а для проверки эффективности её работы – контрольную выборку \tilde{U} .

Достоверностью классификации будем называть отношение количества правильно классифицированных объектов из контрольной выборки к её общему объёму:

$$J_a(\Phi) = \frac{1}{|\tilde{U}|} \left| \left\{ \Phi(x_\omega) = z_\omega | \omega \in \tilde{U} \right\} \right|, \quad (1)$$

$$J_a(\Psi) = \frac{1}{|\tilde{U}|} \left| \left\{ \Psi(y_\omega) = z_\omega | \omega \in \tilde{U} \right\} \right|. \quad (2)$$

Здесь и далее под обозначением $|A|$ для конечного множества A имеется в виду количество элементов в этом множестве.

Нейронные сети на выходе имеют не номер класса, а значение дискриминантной функции для каждого класса, то есть они отображают объект распознавания в множество $[0; 1]^L$. Если на выходе получен вектор $F(\omega) = u \in [0; 1]^L$, то сам предсказанный класс может быть определён как номер элемента этого вектора с наибольшим значением, то есть

$$c = \operatorname{argmax}_{l \in [1; L] \cap \mathbf{Z}} u_l.$$

В качестве альтернативного показателя качества работы нейронной сети можно использовать среднюю точность ранжирования классов (LRAP). Если занумеровать элементы из контрольной выборки $\omega_i \in \tilde{U}$, то ответ нейронной сети можно представить в виде матрицы

$$\tilde{Y} = (F(\omega_i))_{i \in [1; |\tilde{U}|] \cap \mathbf{Z}} \in [0; 1]^{|\tilde{U}| \times L},$$

тогда идеальным ответом была бы матрица

$$Y = (z_{\omega_i} = j)_{i \in [1; |\tilde{U}|] \cap \mathbf{Z}} \in [0; 1]^{|\tilde{U}| \times L},$$

представляющая собой унитарный код для номеров классов объектов распознавания.

В этом случае LRAP оценивается по формуле:

$$J_{\text{LRAP}}(Y, \tilde{Y}) = \frac{1}{|\tilde{U}|} \sum_{i=1}^{|\tilde{U}|} \left(\frac{1}{\sum_{k=1}^L Y(i, k)} \times \sum_{j: i_j=1} \frac{|\{k \in [1; L] \cap \mathbf{Z} | Y(i, k) = 1 \wedge \tilde{Y}(i, k) \geq \tilde{Y}(i, j)\}|}{|\{k \in [1; L] \cap \mathbf{Z} | \tilde{Y}(i, k) \geq \tilde{Y}(i, j)\}|} \right). \quad (3)$$

Этот показатель качества основан на среднеобратном ранге [9].

В качестве функции потерь при обучении этой нейронной сети использовалась категориальная кросс-энтропия [10]:

$$J_H(Y, \tilde{Y}) = -\frac{1}{|\tilde{U}|} \sum_{i=1}^{|\tilde{U}|} \sum_{j=1}^L Y(i, j) \log \tilde{Y}(i, j).$$

Для классификации изображений была построена простейшая свёрточная нейронная сеть, содержащая пять свёрточных слоёв. Для всех пяти свёрточных слоёв была использована функция активации ReLU. Общая архитектура этой нейронной сети представлена на рис. 2 и напоминает упрощённую AlexNet [11]. Число эпох для обучения было подобрано экспериментально.

Для классификации текстов была использована рекуррентная нейронная сеть на основе LSTM [12]. Архитектура отдельного блока LSTM хорошо извест-

на и представлена на рис. 3. В настоящей работе на основе такого блока была построена простейшая нейронная сеть для классификации текстов, архитектура которой представлена на рис. 4. Следует ещё раз заметить, что набор классов для обеих нейронных сетей одинаковый и на выходе они выдают одинаковые по смыслу вектора дискриминантных значений.

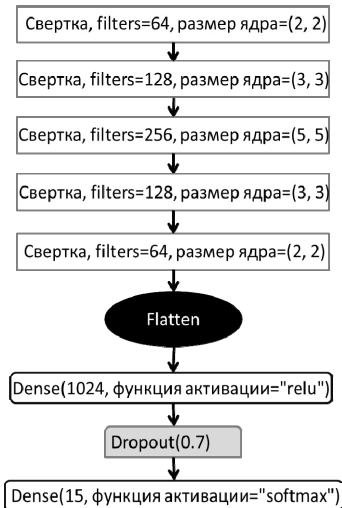


Рис. 2. Архитектура нейронной сети для классификации изображений

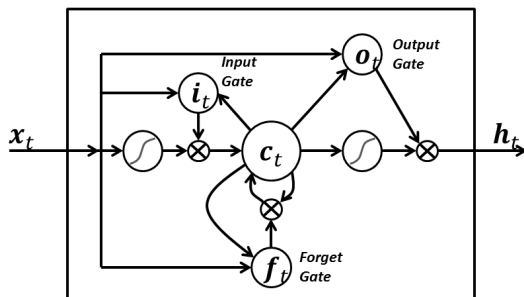


Рис. 3. Блок LSTM

Как известно, LSTM позволяет обучаться долгосрочным зависимостям, что позволяет использовать её для последовательностей токенов, таких как тексты. Кроме слоя LSTM, в архитектуре классификатора также можно заметить слой Embedding, который в качестве входных параметров принимал максимальную длину слова и максимальную длину всего описания в целом. Затем шел слой LSTM с числом нейронов 64. Далее было использовано два слоя субвекторизации, функция активации была выбрана ReLU и один полносвязный слой.

Предварительная обработка текста заключалась в замене слов на их номера в словаре, так что $W \subseteq Z$. После этого последовательность $y_{\omega}(t)$ разбивалась на n-граммы по 150 слов, так что на вход нейронной сети подавалась последовательность векторов

$$Y_{\omega}(k) = (y_{\omega}(k+j))_{j \in [1;n] \cap Z}.$$

Для обучения этой нейронной сети использовалась функция потерь MSE:

$$J_{MSE}(Y, \tilde{Y}) = \frac{1}{|\tilde{U}|} \sum_{i=1}^{|\tilde{U}|} \frac{1}{L} \sum_{j=1}^L (Y(i, j) - \tilde{Y}(i, j))^2. \quad (4)$$

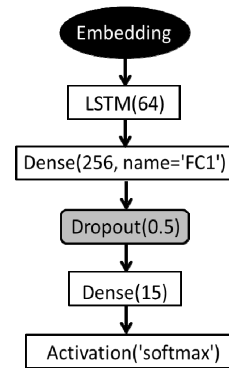


Рис. 4. Архитектура нейронной сети для классификации текстов

2. Мультимодальная модель ruCLIP

Нейросетевая модель ruCLIP (Russian Contrastive Language – Image Pre-training) – это мультимодальная модель русского языка, способная переводить изображения и тексты в единое векторное пространство. Она обучена для русского языка на открытых данных, собранных из Рунета. Всего для обучения использовалось около 240 млн уникальных пар «изображение – описание на естественном языке». Модель была представлена сотрудниками компании SberDevices, входящей в группу компаний Сбер [13]. Модель ruCLIP, в свою очередь, с незначительными изменениями основана на модели CLIP, разработанной сотрудниками компании OpenAI [14].

На рис. 5 схематично представлена работа ruCLIP. Видно, что модель состоит из двух частей (нейронных сетей). Image Encoder – это часть для кодирования изображений и перевода их в общее векторное пространство. В качестве архитектуры в оригинальной работе берутся ResNet разных размеров и Visual Transformer – тоже разных размеров. В ruCLIP Base в качестве image encoder используется ViT-B/16. Text Encoder – часть для кодирования текстов и перевода их в общее векторное пространство. В качестве архитектуры используется текстовый Transformer.

Таким образом, модель ruCLIP можно рассматривать как функцию $F_{CLIP}(x, y)$, переводящую изображение $x_{\omega}(n, m)$ и текст $y_{\omega}(t)$ в некоторый вещественный показатель их близости из множества $[0;1]$. Если же имеется множество занумерованных пациентов $\omega_i \in \Omega$, для каждого из которых существуют томограммы этого пациента $x_{\omega_i}(n, m)$ и соответствующий ему текст радиологического отчёта $y_{\omega}(t)$, то на выходе можно получить целую матрицу

$$Y_{CLIP}(i, j) = (F_{CLIP}(x_{\omega_i}, y_{\omega_j}))_{i, j \in [1;|\Omega|] \cap Z} \in [0;1]^{|\Omega| \times |Z|}.$$

Показателем эффективности работы модели в этом случае можно считать, например, среднее зна-

чение на диагонали этой матрицы. Чем больше среднее значение на диагонали, тем лучше.

$$J_{\text{CLIP}}(Y_{\text{CLIP}}) = \frac{1}{|\Omega|} \text{tr} Y_{\text{CLIP}}. \tag{5}$$

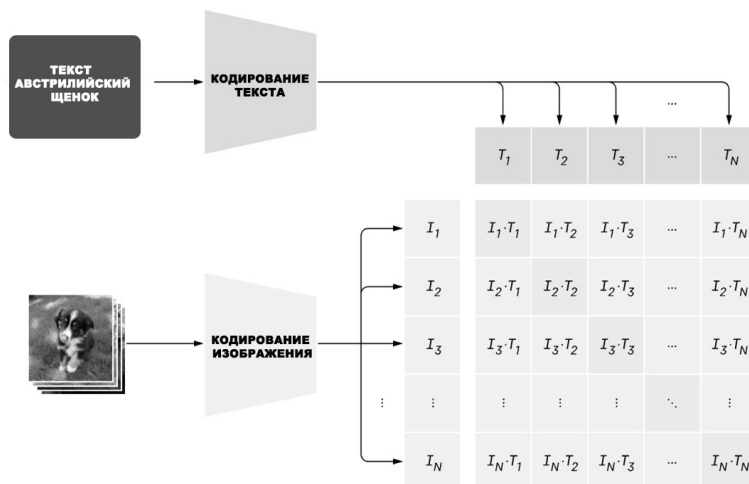


Рис. 5. Схема работы ruCLIP

В данной работе использовано две версии ruCLIP: оригинальная, выложенная её авторами, и дополнительно обученная на обучающей выборке U .

3. Результаты вычислительных экспериментов

В ходе экспериментов предложенные предсказательные модели были обучены на обучающей выборке, после чего эффективность их работы оценивалась на контрольной выборке. Для совместной классификации текстов и изображений классификаторы обучались по-разному.

Выборка для классификатора изображений состояла из 978 изображений, 75% из которых было использовано для обучения классификатора и 25% для проверки его работы. Обучение изначально длилось 1000 эпох. На рис. 6 представлено значение достоверности классификации (1) для обучения классификатора изображений в течение 1000 эпох.

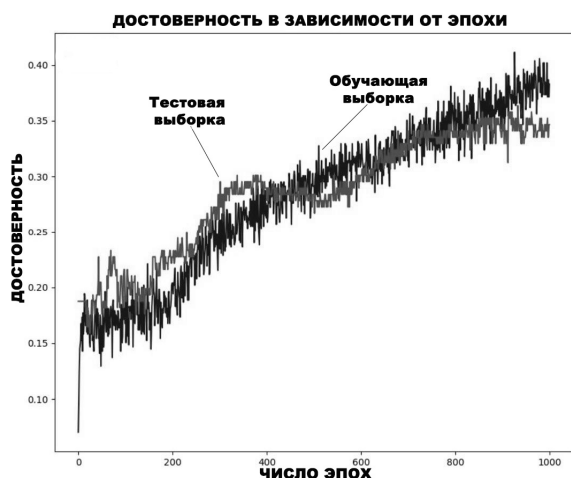


Рис. 6. Изменение достоверности классификации изображений для 1 000 эпох обучения

Значение достоверности составило 0,346, а LRAP – 0,533. Было принято решение увеличить число эпох, так как 15 классов – это достаточно много, и предсказательная модель может улучшиться в результате дальнейшего обучения.

На рис. 7 приведено значение достоверности классификации (1) во время обучения классификатора изображений, которое уже длилось 10000 эпох. Параметры нейронной сети остались без изменений.

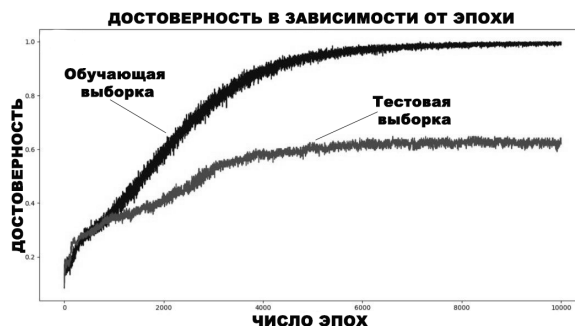


Рис. 7. Изменение достоверности классификации изображений для 10 000 эпох обучения

Для такого числа эпох достоверность классификации (1) составила 0,692, LRAP (3) – 0,795, что является неплохим результатом для классификации на 15 классов.

Классификатор текстов обучался на 978 медицинских отчетах, которые были разбиты на те же 15 классов, что и соответствующие им изображения. Классификатор обучался в течение 50 эпох. На рис. 8 видно, как функция потерь (4) убывает с каждой эпохой, а на рис. 9 – как с каждой эпохой растёт достоверность классификации (2). Значение достоверности классификации на тестовой выборке составило 0,996, что говорит о том, что классификация текстов радиологических отчётов представляет собой объективно несложную задачу.

Таким образом, мы практически безошибочно определяем класс, к которому относится текстовое описание медицинских изображений, и около 70% изображений также классифицируем правильно.



Рис. 8. Изменение функции потерь при обучении классификатора радиологических отчётов

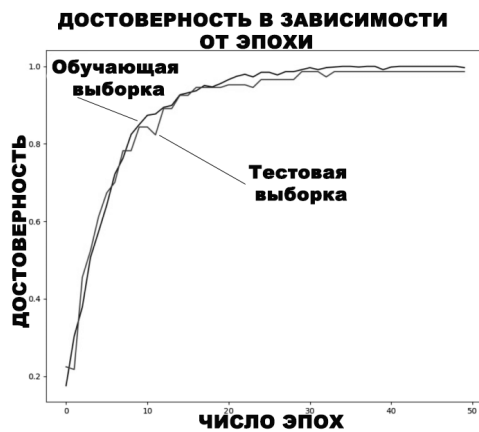


Рис. 9. Изменение достоверности при обучении классификатора радиологических отчётов

Перейдем ко второму подходу, основанному на ruCLIP . Изначально была взята базовая, предобученная мультимодальная модель, которая была обучена на большом количестве изображений, но при этом не была обучена на медицинских данных.

В отличие от первого метода, ruCLIP выдаёт результат схожести, не оценивая отдельно, насколько текст похож на тот или иной класс и насколько изображение похоже на тот или иной класс. Для демонстрации работы ruCLIP была построена матрица $Y_{\text{CLIP}}(i, j)$, представленная на рис. 10. В табл. 2 можно найти расшифровку текстов, которые обозначены буквами «а», «б», «в», «г» и «д». Как можно заметить по диагонали матрицы, модель с большим трудом может установить соответствия между изображениями компьютерной томографии головного мозга и их радиологическими отчётами. Среднее значение параметра схожести (5) для исходной версии ruCLIP составило 0,294, при максимальном значении 1,0 и минимальном значении 0,0.

Для обучения модели ruCLIP выборка была разбита на обучающую и контрольную. В обучающую выборку вошло 75% от общей выборки и оставшиеся

25% попали в контрольную выборку. Все изображения были приведены к единому размеру 224×224 отсчёта. Максимальная длина текста составила 76 слов. Обучение длилось 71 эпоху и продолжалось в течение около 12 часов.

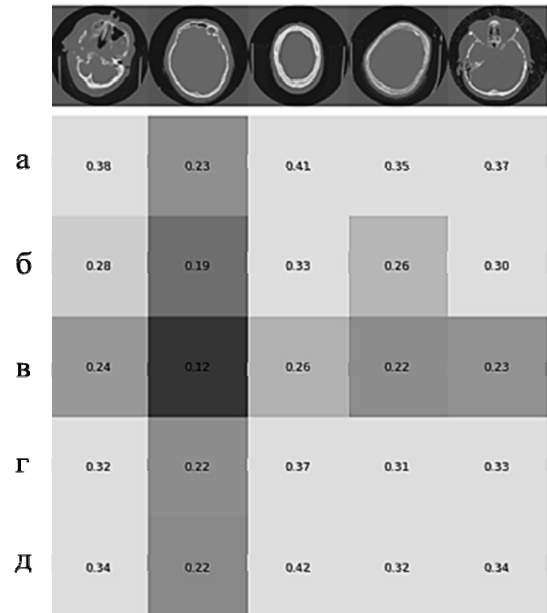


Рис. 10. Пример матрицы схожести для некоторых изображений КТ головного мозга и радиологических отчётов

Табл. 2. Примеры радиологических отчётов

Обозначение	Текст
а	ОНМК по геморрагическому типу левой гемисферы головного мозга с прорывом в желудочки
б	Патологии не выявлено
в	Патологии не выявлено
г	ОНМК по ишемическому типу в бассейне ПВСА, отек правой гемисферы головного мозга
д	Умеренный отек правой гемисферы головного мозга

После обучения было также подсчитано среднее значение схожести (5) для контрольной выборки, и оно составило 0,95.

Заключение

В ходе работы были предложены и исследованы два возможных подхода к совместному анализу томограмм и рентгенологических протоколов с целью автоматического уточнения патологических состояний головного мозга. Использование свёрточной нейронной сети для классификации изображений позволяет правильно классифицировать 69% изображений из контрольной выборки, а использование рекуррентной нейронной сети для классификации текстов позволяет правильно классифицировать 99,6% радиологических

отчётов. Совместное использование этих нейронных сетей может позволить автоматически проверять соответствие диагноза на изображении и в радиологическом отчёте. Также с этой целью можно использовать мультимодальную модель ruCLIP, которая после дополнительного обучения также способна предсказывать меру соответствия между заданным изображением и текстом. Пороговое значение подбирается экспериментально в зависимости от набора и типа данных.

Комплексное исследование эффективности на клинических данных показало, что использование двух простейших нейронных сетей позволяет добиться сопоставимых результатов по сравнению с моделью ruCLIP при том, что не требует больших объёмов памяти для хранения предсказательных моделей, больших объёмов выборки и времени на обучение. Модель ruCLIP же требует значительных вычислительных ресурсов, но при этом не показывает значительного прогресса при обучении на выборке из порядка тысячи медицинских изображений, что объясняется размерами самой модели и объёмами данных, которые были использованы для её обучения изначально. Эксперименты показывают, что для решения прикладной задачи сопоставления радиологических отчётов и изображений компьютерной томографии головного мозга достаточно использовать более простые нейронные сети, нежели ruCLIP.

Полученные в данной работе результаты могут быть в дальнейшем использованы при разработке автоматизированной системы уточнения поставленных диагнозов по данным, хранящимся в базах лечебных учреждений. Это может существенно повысить эффективность работы подобных медицинских учреждений, поскольку позволит в автоматическом режиме обнаруживать спорные случаи, требующие дополнительного внимания большего количества врачей-радиологов. Конечно, разработка подобной системы – это отдельная задача, которой авторы планируют заняться в будущем, а оценивание эффективности подобной системы потребует масштабного клинического исследования на большом количестве пациентов. В дальнейшем планируется провести совместный анализ для исследований других анатомических областей и модальностей, например, МРТ.

Благодарности

Работа выполнена при поддержке РФФИ (грант 19-29-01235МК).

References

[1] Yanase J, Triantaphyllou E. A systematic survey of computer-aided diagnosis in medicine: Past and present devel-

- opments. *Expert Syst Appl* 2019; 138: 112821. DOI: 10.1016/j.eswa.2019.112821.
- [2] Choplin RH, Boehme JM, Maynard CD, Picture archiving and communication systems: an overview. *Radiographics* 1992; 12(1): 127-129. DOI: 10.1148/radiographics.12.1.1734458.
- [3] Zingmond D, Lenert L. Monitoring free-text data using medical language processing. *Comput Biomed Res* 1993; 26(5): 467-481. DOI: 10.1006/cbmr.1993.1033.
- [4] Wahlang I, Maji AK, Saha G, Chakrabarti P, Jasinski M, Leonowicz Z, Jasinska E. Brain magnetic resonance imaging classification using deep learning architectures with gender and age. *Sensors* 2022; 22: 1766.
- [5] Yagis E, Atnafu SW, de Herrera AG, et al. Effect of data leakage in brain MRI classification using 2D convolutional neural networks. *Sci Rep* 2021; 11: 22544. DOI: 10.1038/s41598-021-01681-w.
- [6] Bala W, Steinkamp J, Feeney T, Gupta A, Sharma A, Kantrowitz J, Cordella N, Moses J, Draken FT. A web application for adrenal incidentaloma identification, tracking, and management using machine learning. *Appl Clin Inform* 2020; 11(4): 606-616. DOI: 10.1055/s-0040-1715892.
- [7] Dantas R, Bertoldi M, Wangenheim F. An approach for retrieval and knowledge communication using medical documents. *Proc 23rd Int Conf on Software Engineering and Knowledge Engineering (SEKE)* 2011: 169-174.
- [8] Sludnova A, Shutko V, Gaidel A, Zelter P, Kapishnikov A, Nikonorov A. Identification of pathological changes in the lungs using an analysis of radiological reports and tomographic images. *Computer Optics* 2021; 45(2): 261-266. DOI: 10.18287/2412-6179-CO-793.
- [9] Wu Y, Mukunoki M, Funatomi T, Minoh M, Lao S. Optimizing mean reciprocal rank for person re-identification. *8th IEEE Int Conf on Advanced Video and Signal Based Surveillance (AVSS)* 2011: 408-413. DOI: 10.1109/AVSS.2011.6027363.
- [10] Chen CH, Lin PH, Hsieh JG, Cheng SL, Jeng JH. Robust multi-class classification using linearly scored categorical cross-entropy. *3rd IEEE Int Conf on Knowledge Innovation and Invention (ICKII)* 2020: 200-203. DOI: 10.1109/ICKII50300.2020.9318835.
- [11] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM* 2017; 60(6): 84-90. DOI: 10.1145/3065386.
- [12] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997; 9(8): 1735-1780. DOI: 10.1162/neco.1997.9.8.1735.
- [13] Shonenkov A, Kuznetsov A, Dimitrov D, Shavrina T, Chesakov D, Maltseva A, Fenogenova A, Pavlov I, Emelyanov A, Markov S, Bakshandaeva D, Shybaeva V, Chertok A. RuCLIP – new models and experiments: a technical report. *arXiv Preprint*. 2022. Source: <https://arxiv.org/abs/2202.10784>.
- [14] Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, Krueger G. Learning transferable visual models from natural language supervision. *Int Conf on Machine Learning* 2021; 139: 8748-8763.

Сведения об авторах

Агафонова Юлия Дмитриевна, 2000 года рождения, в 2022 году окончила Самарский национальный исследовательский университет и получила квалификацию «Бакалавр». Область научных интересов: компьютерное зрение, обработка изображений, анализ изображений, цифровая фотография. E-mail: uka132@mail.ru.

Гайдель Андрей Викторович, 1989 года рождения, кандидат технических наук, в 2012 году окончил Самарский государственный аэрокосмический университет имени академика С.П. Королёва по направлению «Прикладная математика и информатика», в 2015 году защитил кандидатскую диссертацию. Работает доцентом кафедры технической кибернетики Самарского университета и научным сотрудником в Институте систем обработки изображений РАН – филиале ФНИЦ «Кристаллография и фотоника» РАН. Область научных интересов: компьютерная обработка изображений, теория распознавания образов, интеллектуальный анализ данных, теория алгоритмов. E-mail: andrey.gaidel@gmail.com.

Зельтер Павел Михайлович, 1986 года рождения, в 2010 году с отличием окончил Самарский государственный медицинский университет, в 2012 году – ординатуру по специальности «Рентгенология», в 2016 г. защитил кандидатскую диссертацию, работает доцентом кафедры лучевой диагностики и лучевой терапии СамГМУ. Область научных интересов: компьютерная томография, трехмерное моделирование томографических изображений, компьютерный анализ медицинских изображений. E-mail: pzelter@mail.ru.

Капишников Александр Викторович, 1955 года рождения. В 1972 году окончил лечебный факультет Куйбышевского медицинского института. Докторскую диссертацию защитил в 2004 году. Заведует кафедрой лучевой диагностики и лучевой терапии с курсом медицинской информатики СамГМУ Минздрава России. Область научных интересов: лучевая диагностика и ядерная медицина в пульмонологии, остеологии, нефрологии и онкологии, информационные технологии медицинской радиологии, САDe и САDx-системы для компьютерной поддержки диагностики.

Кузнецов Андрей Владимирович, родился в 1987 году. В 2010 году окончил СГАУ. В 2010 поступил в аспирантуру СГАУ, в 2013 г. защитил диссертацию на соискание степени кандидата технических наук. В настоящее время работает старшим научным сотрудником НИЛ-97 Самарского университета. Область научных интересов включает компьютерное зрение, машинное обучение, искусственные нейронные сети, обнаружение искажений и подделок изображений и т.п. Имеет более 40 публикаций, в том числе более 20 научных статей и 1 монографию.

E-mail: kuznetsoff.andrey@gmail.com. Страница в интернете: <http://nil97.ssau.ru/employee/detail.php?ID=35>

Суровцев Евгений Николаевич, 1987 года рождения, учился в Самарском государственном медицинском университете. Работает ассистентом кафедры лучевой диагностики и лучевой терапии. ФГБОУ ВО СамГМУ Минздрава России и заведующим отделением магнитно-резонансной томографии ООО «ЛДЦ МИБС – Тольятти». Область научных интересов: лучевая диагностика, магнитно-резонансная томография, онкология, обработка изображений. E-mail: evgenisurovcev@mail.ru.

Сведения об авторе **Никоноров Артем Владимирович** см. стр 101 этого номера.

ГРНТИ: 28.23.15

Поступила в редакцию 1 августа 2022 г. Окончательный вариант – 10 октября 2022 г.

Joint analysis of radiological reports and CT images for automatic validation of pathological brain conditions

Y.D. Agafonova¹, A.V. Gaidel^{1,2}, P.M. Zelter³, A.V. Kapishnikov³, A.V. Kuznetsov^{1,4,5}, E.N. Surovtsev³,
A.V. Nikonorov^{1,2}

¹ Samara National Research University, 443086, Samara, Russia, Moskovskoye Shosse 34;

² IPSI RAS – Branch of the FSRC “Crystallography and Photonics” RAS,
443001, Samara, Russia, Molodogvardeyskaya 151;

³ FSBEI HE SamSMU MOH Russia, 443099, Samara, Russia, Chapayevskaya 89;

⁴ Sber AI, 121170, Moscow, Russia, Kutuzovsky prospekt, 32 building 2;

⁵ Artificial Intelligence Research Institute (AIRI), 105064, Moscow, Russia, Nizhniy Susalnyy pereulok, 5

Abstract

We consider a problem of validation of radiological medical reports and computed tomography images for an automated analysis of brain structures. Two methods for solving the problem are proposed: a method based on the ruCLIP multimodal model, and a method based on the joint use of two separate classifiers – for a text report and for a brain CT image. We discuss methods evaluation and the obtained results. The proposed approaches make it possible to correctly **classify** 99.6% of radiological reports from a test sampling into 15 possible diagnoses.

Keywords: deep learning, computed tomography, computer-aided diagnosis, pattern recognition, natural language processing.

Citation: Agafonova YD, Gaidel AV, Zelter PM, Kapishnikov AV, Kuznetsov AV, Surovtsev EN, Nikonorov AV. Joint analysis of radiological reports and CT images for automatic validation of pathological brain conditions. *Computer Optics* 2023; 47(1): 152-159. DOI: 10.18287/2412-6179-CO-1201.

Acknowledgements: This work was supported by the Russian Science Foundation (Project No. 19-29-01235).

Authors' information

Julia Dmitrievna Agafonova (b. 2000) in 2022 she graduated from Samara National Research University with a bachelor's degree. Research interests: computer vision, image processing, image analysis, digital photography. E-mail: uka132@mail.ru.

Andrey Viktorovich Gaidel (b. 1989). Graduated from Samara State Aerospace University in 2012, majoring in Applied Mathematics and Informatics. He received his Candidate of Science degree in Physics and Math in 2015 from the SSAU. Currently he is a teaching assistant of the Technical Cybernetics sub-department and Engineer at SSAU's laboratory SRL-35 of Samara State Aerospace University, also working as an intern researcher of the Image Processing Systems Institute of the Russian Academy of Sciences – Branch of the FSRC “Crystallography and Photonics” RAS, Samara, Russia. His research interests currently focus on computer image processing, pattern recognition, data mining and theory of computation. E-mail: andrey.gaidel@gmail.com.

Pavel Mikhailovich Zelter (b. 1986) graduated with honors from Samara State Medical University in 2010, residency in Radiology in 2012. Assistant lecturer of Radiology department with course of Medical Informatics of Samara State Medical University. Research interests are currently focused on computerized tomography, COPD, digital analysis of medical images. E-mail: pzelter@mail.ru.

Aleksandr Viktorovich Kapishnikov (b.1955). In 1972 he graduated from Medical faculty of Kuibyshev Medical Institute. Doctor of Science in Medicine (2004). Head of Diagnostic Radiology and Medical Informatics department of Samara State Medical University. Research interests: diagnostic radiology and nuclear medicine in pulmonology, osteology, nephrology and oncology, information technology in medical imaging, computer-based diagnostic support system (CADe and CADx).

Andrey Vladimirovich Kuznetsov (b. 1987) graduated (2010) from SSAU. He studied as a post-graduate student at SSAU from 2010 and received his PhD in Technical Sciences in 2013. Nowadays he is a senior researcher at the Remote Sensing Data Analysis laboratory at Samara National Research. His research interests are currently focused on

computer vision, machine learning, artificial neural networks, detection of distortions and sub-images of images, etc. He has 40 publications, including 20 scientific papers and 1 monograph.

E-mail: kuznetsoff.andrey@gmail.com. Web-page: <http://nil97.ssau.ru/employee/detail.php?ID=35>

Evgeny Nikolaevich Surovtsev (b. 1987). Studied at Samara State Medical University. Works as an assistant at Radiation Diagnostics and Radiation Therapy department. FSBEI HE SamSMU of the Ministry of Health of Russia and the head of Magnetic Resonance Imaging department of LLC MDTs MIBS - Togliatti. Research interests: radiation diagnostics, magnetic resonance imaging, oncology, image processing. E-mail: evgeniisurovcev@mail.ru.

Artem Vladimirovich Nikonorov (b. 1979). He graduated from SSAU in 2002. He received his PhD degree in 2016. Now he works as a professor at Supercomputing and Computer Science sub-department of Samara University and as a head of Laboratory of Video Mining in IPSI RAS. He has more than 70 publications, two monographs. Field of scientific interest: pattern recognition and image analysis, system identification, parallel and distributed programming, GPGPU programming. E-mail: artniko@gmail.com.

Received August 1, 2022. The final version – October 10, 2022.
