

АНАЛИЗ ТЕКСТОВ ОПИСАНИЙ КТ-ИССЛЕДОВАНИЙ ГОЛОВНОГО МОЗГА С ПРИЗНАКАМИ ВНУТРИЧЕРЕПНЫХ КРОВОИЗЛИЯНИЙ С ПОМОЩЬЮ АЛГОРИТМА ДЕРЕВА РЕШЕНИЙ

DOI: 10.17691/stm2022.14.6.04

УДК 616.831-005-073.756.8

Поступила 07.09.2022 г.

**А.Н. Хоружая**, младший научный сотрудник отдела инновационных технологий¹;**Д.В. Козлов**, младший научный сотрудник отдела медицинской информатики, радиомики и радиогеномики¹;**К.М. Арзамасов**, к.м.н., руководитель отдела медицинской информатики, радиомики и радиогеномики¹;**Е.И. Кремнева**, к.м.н., ведущий научный сотрудник отдела инновационных технологий¹; старший научный сотрудник²¹Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы, ул. Петровка, 24, стр. 1, Москва, 127051;²Научный центр неврологии, Волоколамское шоссе, 80, Москва, 125367

Цель исследования — разработать, обучить и протестировать алгоритм анализа текстовых описаний КТ-исследований головного мозга с использованием модели дерева решений для задачи простой бинарной классификации наличия/отсутствия признаков внутричерепного кровоизлияния (ВЧК).

Материалы и методы. Исходные данные представляют собой выгрузку из системы ЕРИС ЕМИАС, содержащей 34 188 исследований, полученных в результате проведения бесконтрастной КТ головного мозга в 56 медицинских организациях стационарной медицинской помощи. Анализ данных, а также их предварительную обработку осуществляли с использованием библиотеки для символьной и статистической обработки естественного языка NLTK (Natural Language Toolkit, v. 3.6.5) и библиотеки для машинного обучения, содержащей инструменты для решения задач классификации scikit-learn. По 14 подобранным ключевым словам, имеющим отношение к ВЧК, а также 33 стоп-фразам, содержащим ключевые слова, наличие которых в тексте описания подразумевало отсутствие ВЧК, выполняли автоматический отбор КТ-исследований и их последующую экспертную верификацию. Получена выборка из 3980 протоколов описаний, из которой сформированы два класса данных: содержащих описание ВЧК и без такового. В качестве модели, с помощью которой решали задачу бинарной классификации, выбран алгоритм решающего дерева. Для оценки производительности модели КТ-исследования были разделены случайным образом на две выборки в соотношении 7:3. Из 3980 протоколов 2786 были отнесены к обучающему набору данных, 1194 — к тестовому.

Результаты. По результатам тестирования чувствительность разработанного и обученного алгоритма при бинарной классификации описаний КТ-исследований «с признаками ВЧК» и «без признаков ВЧК» составила 0,94, специфичность — 0,88, F-мера — 0,83.

Заключение. Разработанный и обученный алгоритм анализа текстовых протоколов КТ-описаний показал высокую точность работы при исследованиях головного мозга с признаками ВЧК. Он может применяться для решения задач бинарной классификации и создания соответствующих наборов данных. Ограничением алгоритма является необходимость ручного пересмотра КТ-исследований с целью обеспечения контроля качества.

Ключевые слова: компьютерная томография; диагностические описания; внутричерепное кровоизлияние; обработка медицинских текстов; машинное обучение; алгоритм дерева решений.

Как цитировать: Khoruzhaya A.N., Kozlov D.V., Arzamasov K.M., Kremneva E.I. Text analysis of radiology reports with signs of intracranial hemorrhage on brain CT scans using the decision tree algorithm. *Sovremennye tehnologii v medicine* 2022; 14(6): 34, <https://doi.org/10.17691/stm2022.14.6.04>

Для контактов: Хоружая Анна Николаевна, e-mail: a.khoruzhaya@npcmr.ru

English

Text Analysis of Radiology Reports with Signs of Intracranial Hemorrhage on Brain CT Scans Using the Decision Tree Algorithm

A.N. Khoruzhaya, Junior Researcher, Department of Innovative Technologies¹;

D.V. Kozlov, Junior Researcher, Department of Medical Informatics, Radiomics and Radiogenomics¹;

K.M. Arzamasov, MD, PhD, Head of the Department of Medical Informatics, Radiomics and Radiogenomics¹;

E.I. Kremneva, MD, PhD, Leading Researcher, Department of Innovative Technologies¹; Senior Researcher²

¹Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department, Bldg 1, 24 Petrovka St., Moscow, 127051, Russia;

²Research Center of Neurology, 80 Volokolamskoye Shosse, Moscow, 125367, Russia

The aim of the study is to create, train, and test the algorithm for the analysis of brain CT text reports using a decision tree model to solve the task of simple binary classification of presence/absence of intracranial hemorrhage (ICH) signs.

Materials and Methods. The initial data is a download from the Unified Radiological Information Service of the Unified Medical Information and Analytical System (URIS UMIAS) containing 34,188 studies obtained by a non-contrast CT of the brain in 56 inpatient medical settings. Data analysis and preprocessing were carried out using NLTK (Natural Language Toolkit, version 3.6.5), a library for symbolic and statistical processing of natural language, and scikit-learn, a machine learning library containing tools for classification tasks. According to 14 selected ICH-related key words, as well as 33 stop-phrases with key words denoting absence of ICH, an automatic selection of the CT investigations and their subsequent expert verification were carried out. Two classes of investigations were formed based on the sample from 3980 protocol descriptions: containing descriptions of ICH and without them. The problem of binary classification was solved using the decision tree algorithm as a model. To evaluate the performance of the model, the CT investigations were divided randomly into samples in the ratio of 7:3. Of 3980 protocols, 2786 were assigned to the training data set, 1194 — to the test one.

Results. According to the test results, the designed and trained algorithm in the binary classification of the CT reports “with signs of ICH” and “without signs of ICH” has shown sensitivity of 0.94, specificity of 0.88, F-score of 0.83.

Conclusion. The developed and trained algorithm for the analysis of radiology reports has demonstrated high accuracy in relation to brain CT with signs of intracranial hemorrhage and can be used to solve binary classification problems and create appropriate data sets. However, it is limited by the need for manual revision of CT studies to ensure quality control.

Key words: computed tomography; diagnostic reports; intracranial hemorrhage; natural language processing; machine learning; decision tree algorithm.

Введение

Современные медицинские учреждения генерируют и накапливают огромные объемы информации разных категорий: это текстовые данные медицинских карт пациентов с описаниями жалоб и анамнезом, направления на обследование, эпикризы, текстовые и цифровые результаты лабораторных и инструментальных исследований, цифровые медицинские изображения и т.д. Основная доля этого массива информации приходится на неструктурированные текстовые медицинские данные. Тем не менее в подавляющем числе случаев именно в них содержится ценная информация, которая может лечь в основу разработки новых инструментов цифровой медицины: систем поддержки принятия врачебных решений, различных электронных медицинских помощников, моделей прогнозирования развития заболеваний и других атрибутов происходящей в данный момент цифровой трансформации здравоохранения [1].

Количество описаний диагностических исследований (например, компьютерных томограмм), генерируемое в течение года, составляет более сотни тысяч

и с каждым годом лишь возрастает. Так, по данным Единого радиологического информационного сервиса Единой медицинской информационно-аналитической системы (ЕРИС ЕМИАС), за 9 мес 2017 г. в амбулаторно-поликлинических учреждениях Департамента здравоохранения Москвы было проведено 111487 КТ-исследований [2], а в 2021 г. их количество составило уже 777402. Своевременное извлечение необходимой информации, полученной в результате рентгенологических исследований, и ее последующая аналитика посредством алгоритмов машинного обучения могут способствовать принятию быстрых и эффективных решений при диагностике той или иной патологии и обеспечить повышение качества соответствующей медицинской помощи. Особую актуальность это приобретает в сфере экстренной и неотложной медицинской помощи, в частности при своевременной диагностике внутричерепных кровоизлияний (ВЧК) [3–5].

При разработке качественных алгоритмов машинного обучения для анализа медицинских изображений необходимо создание качественных наборов данных. Первичный отбор таких данных из всего массива ис-

следований может осуществляться вручную, что требует больших временных затрат, а может быть автоматизирован на основе анализа неструктурированных текстовых протоколов рентгенологических описаний. Методы машинной обработки естественного языка (natural language processing, NLP) преобразовывают подобный неструктурированный текст в структурированную форму, из которой можно извлекать информацию, несущую в себе необходимую смысловую нагрузку [6]. Таким образом, автоматическая обработка протоколов описаний позволяет отбирать рентгенологические исследования с искомыми признаками.

Эффективность машинного обучения во многом зависит от того, насколько хорошо были размечены данные в обучающей выборке, что в свою очередь требует значительных усилий врачей-экспертов, обладающих узкоспециализированными знаниями. Чтобы снизить временные затраты с их стороны и упростить процесс разметки без ухудшения итогового результата, существует стратегия слабого контроля для алгоритмов машинного обучения на слабо размеченных обучающих данных [7]. Она нашла широкое применение в биомедицинских областях именно для задач классификации [8, 9]. Суть ее такова: на начальном этапе обучения алгоритма автоматически создаются слабые метки, которые анализирует врач-эксперт. Таким образом формируется набор данных, на котором уже происходит дальнейшее обучение. В конечном итоге получается обученная модель для извлечения информации из неструктурированного клинического текста. Предложенная схема была оценена в задачах бинарной классификации и продемонстрировала высокую точность срабатывания — до 0,97 [10].

Для выполнения подобных задач следует выбирать простые модели с возможностью автоматического обучения. Одной из таких моделей, сочетающей в себе оба эти качества, является алгоритм дерева решений [11]. Кроме того, немаловажным фактором представляется интерпретируемость модели, а самая высокая интерпретируемость выявляется у алгоритма дерева решений [12]. Поэтому в данной работе мы остановились на его применении.

Данных об эффективности применения алгоритмов на основе решающих деревьев для классификации медицинских текстов, в частности протоколов описаний КТ-исследований головного мозга, в литературе очень мало. При этом более широко представлены разработки в области классификации текстов на иностранных языках, в частности на английском [13] и китайском [14], но практически отсутствуют работы по анализу русскоязычных медицинских текстов. Вместе с тем решение данной задачи позволило бы проводить аналитику потока диагностических исследований по частоте встречаемости патологий, а также осуществлять выборку исследований для подготовки и контроля качества наборов данных при обучении алгоритмов компьютерного зрения, предназначенных для анализа медицинских изображений, хотя этими

задачами применение подобных алгоритмов в лучевой диагностике не ограничивается [15].

Цель исследования — разработать, обучить и протестировать алгоритм анализа текстовых описаний КТ-исследований головного мозга с использованием модели дерева решений для задачи простой бинарной классификации наличия/отсутствия признаков внутричерепного кровоизлияния.

Материалы и методы

Исходные данные представляют собой выгрузку из системы ЕРИС ЕМИАС, содержащую 34 188 исследований, которые были получены в результате проведения бесконтрастной КТ головного мозга в 56 медицинских организациях стационарной медицинской помощи. Каждая строка такого набора данных содержит следующую информацию: уникальный идентификатор, возраст, пол, дату проведения диагностики, список медицинских учреждений, участвующих в исследовании, а также описание и заключение по исследованию.

В качестве критериев исключения КТ-исследований из выборки были приняты следующие факторы: отсутствие протоколов описаний с заключениями (пустые поля в данных строках), возраст до 18 лет, отсутствие информации о возрасте или его аномальные значения из-за неправильного ввода даты (976, 1000 лет), полные дубликаты строк. По этой причине количество исследований, данные которых вошли в выборку, составило 29 682. Соответственно критериями включения в выборку стали заполненные поля с текстом описания и заключениями, отсутствие аномальных значений возраста пациентов и дублирование информации.

Все исследования были сделаны в период с 00:00:00 01.01.2020 по 00:00:00 31.12.2020. В число пациентов вошли 14 895 женщин и 14 787 мужчин. Минимальный возраст составил 18 лет, максимальный — 99.

Рассматриваемая задача оценки наличия ВЧК по результатам текстовых протоколов описаний КТ-исследований головного мозга, независимо от локализации их проведения, представляет собой задачу бинарной классификации: кровоизлияние есть/нет. Анализ данных, а также их предварительную обработку осуществляли с использованием библиотеки для символьной и статистической обработки естественного языка NLTK (Natural Language Toolkit, v. 3.6.5) и библиотеки для машинного обучения, содержащей инструменты для решения задач классификации, — scikit-learn. Используемые библиотеки, а также последующий алгоритм написаны на языке программирования Python (v. 3.9.7).

В качестве исходной выборки для машинного обучения были отобраны КТ-исследования, содержащие в описании и заключении 14 ключевых слов, имеющих отношение к ВЧК: кровоизлияние(я), гематома(ы), геморрагический(ие), внутримозговой(ая, ые), субарах-

ноидальные, эпидуральные, субдуральные, внутримозговые, паренхиматозные, эписубдуральные; САК (субарахноидальное кровоизлияние), ЭДК (эпидуральное кровоизлияние), СДК (субдуральное кровоизлияние), ВМК (внутричерепное кровоизлияние). Эти ключевые слова были выбраны на основании экспертного мнения специалиста — врача-рентгенолога со стажем в этой области более трех лет. На рис. 1 приведено распределение ключевых слов в исходной выборке.

На данном этапе отбор осуществляли «механически», т.е. по факту присутствия данного ключа (слова) в тексте, без учета окружающих слов. Количество КТ-исследований после этого этапа составило 5889 (рис. 2).

Однако отбор по ключевым словам не стал для нас решением задачи. Дело в том, что ключевое слово в сочетании с отрицанием (стоп-слово или стоп-фраза) означает отсутствие искомой патологии. И врачи-рентгенологи часто употребляют подобные стоп-фразы в протоколах описаний КТ-исследований. По этой причине поиск только лишь ключевых слов не позволяет корректно получить нужные данные. В результате сформированная таким образом выборка (5889) была передана для верификации трем врачам-рентгенологам со стажем работы более трех лет. С их участием был составлен список из 33 стоп-фраз, содержание которых в протоколе подразумевало отсутствие любых ВЧК в исследовании. Примеры данных стоп-фраз:

очагов патологической плотности вещества головного мозга не выявлено;

КТ-данных за внутричерепную гематому и ушиб головного мозга не получено;

признаков внутричерепного кровоизлияния не выявлено;

на полученных изображениях очагов патологической плотности в веществе головного мозга не определяется;

КТ-признаков внутричерепной гематомы, перелома костей черепа, других очаговых и объемных изменений вещества мозга не получено.

На следующем этапе был проведен повторный

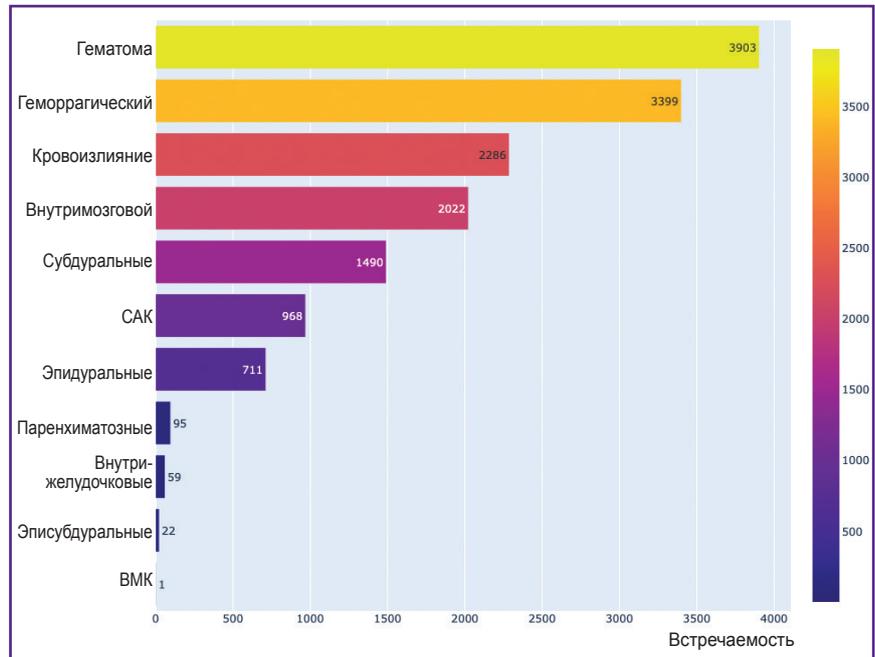


Рис. 1. Распределение частоты встречаемости (горизонтальная ось) ключевых слов (вертикальная ось), имеющих отношение к внутричерепному кровоизлиянию в исходной выборке.

Здесь: ВМК — внутричерепное кровоизлияние; САК — субарахноидальное кровоизлияние

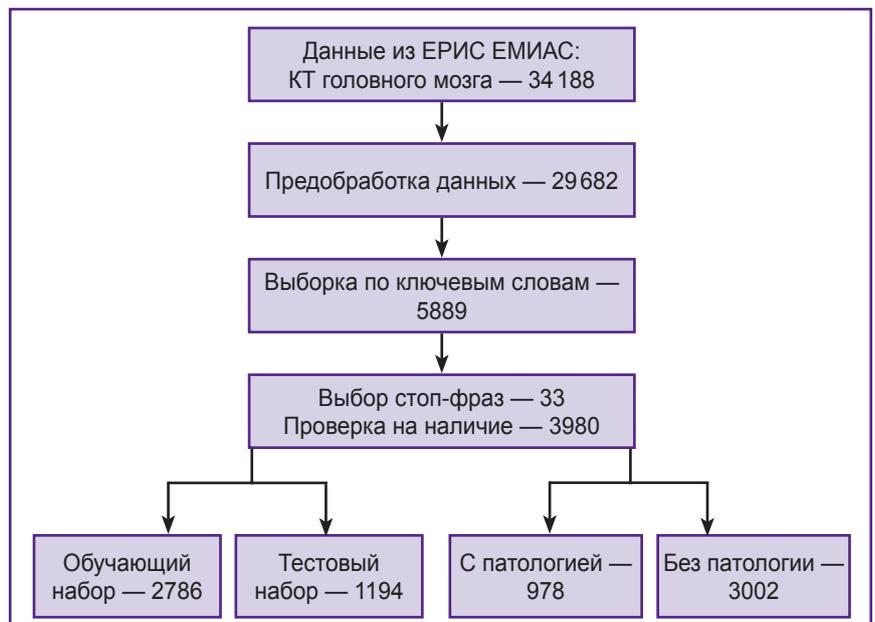


Рис. 2. Дизайн исследования

автоматический отбор КТ-исследований, в которых присутствовали и ключевые слова, и стоп-фразы. Полученные в итоге 3980 исследований были разделены на два класса: содержащие описание ВЧК (978) и без такового (3002). В качестве модели, с помощью которой решалась задача бинарной классификации, был выбран алгоритм решающего дерева. Дерево решений (DecisionTreeClassifier) является одним из

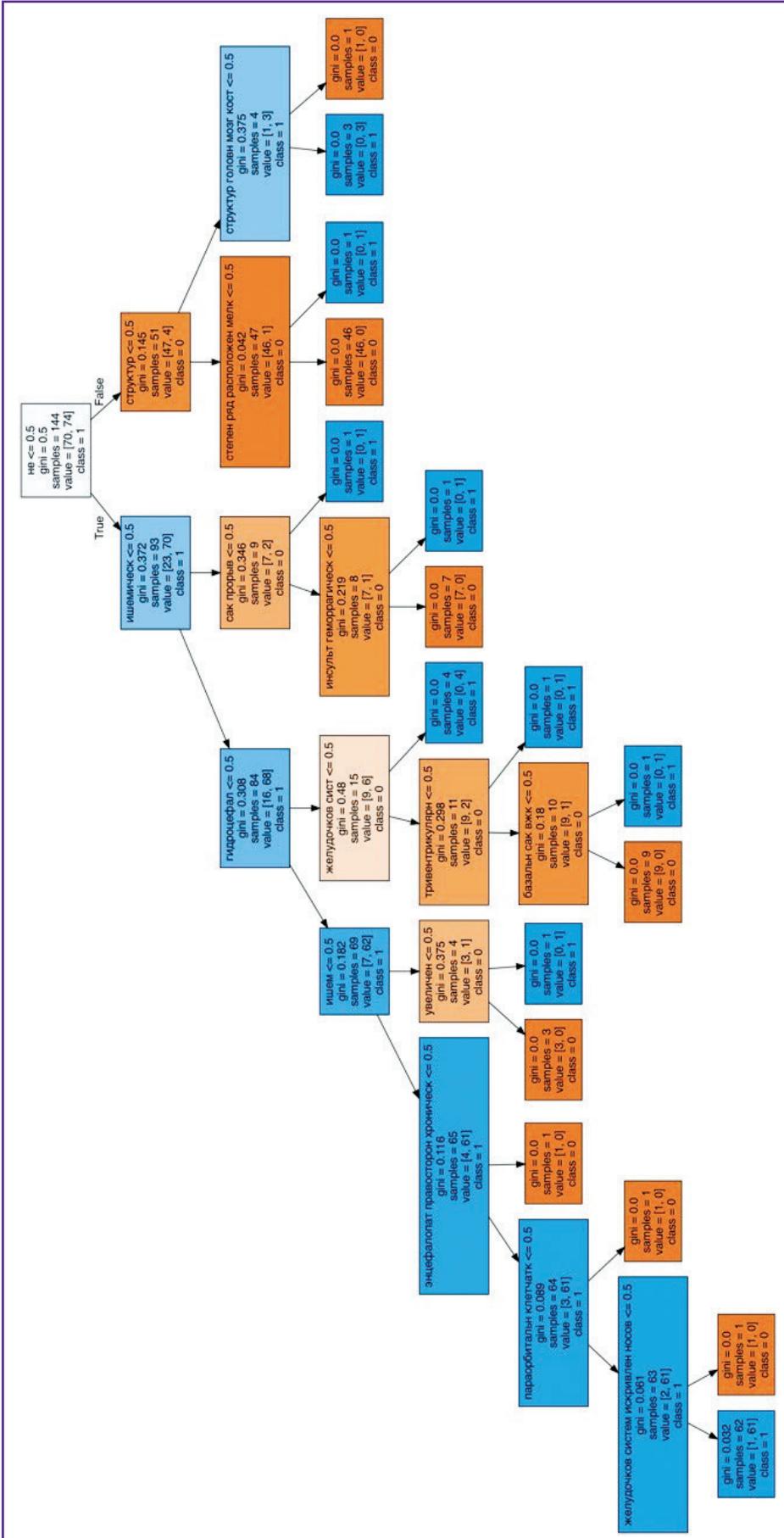


Рис. 3. Разработанный алгоритм анализа текстовых описаний КТ-исследований головного мозга с признаками внутримозговых кровоизлияний

методов классификации в машинном обучении с помощью библиотеки scikit-learn. Максимальная глубина дерева решений была подобрана эмпирически и составила 15 уровней. Оценка качества работы алгоритма проводилась при помощи функции classification_report (рис. 3).

Для оценки производительности модели КТ-исследования были разделены случайным образом на выборки в соотношении 7:3, поскольку именно такое соотношение обучающего/тестировочного набора данных позволяет получить наиболее оптимальные метрики качества работы алгоритма [16]. Из 3980 протоколов 2786 были отнесены к обучающему набору данных, 1194 — к тестовому. Из 1194 тестовых наборов 927 не содержали признаков ВЧК, 267 имели такие признаки.

Результаты и обсуждение

По результатам тестирования чувствительность обученного алгоритма при бинарной классификации текстовых протоколов КТ-исследований «с признаками ВЧК» и «без признаков ВЧК» составила 0,94 (95% CI: 0,942–0,939), специфичность — 0,88 (95% CI: 0,841–0,919). Положительная прогностическая значимость составила 0,96, т.е. с вероятностью 96% выбранное алгоритмом КТ-исследование с меткой «патология» будет иметь признаки таковой. В свою очередь отрицательная прогностическая значимость оказалась равной 0,81, следовательно, с вероятностью 81% алгоритм даст верный ответ об отсутствии признаков патологии в тексте, где она действительно не описана. F-мера при этом составила 0,83. Этот показатель представляет собой взвешенное гармоническое среднее и объединяет полноту и точность исследуемого алгоритма. Для наглядности полученных результатов приведем четырехпольную таблицу — матрицу ошибок (рис. 4).

Данные, полученные в ходе нашего исследования, в целом сопоставимы с теми, которые приводятся в литературных источниках. Так, в исследовании I.C. Hostettler и соавт. [12] применение методов естественной обработки языка показывает достаточно высокую эффективность в выявлении и исходе заболеваний. Клинический исход при субарахноидальном кровоизлиянии на 1-, 3- и 7-й дни по данным лабораторных исследований 548 пациентов был предсказан с помощью алгоритма решающего дерева. Модель имела наибольшую точность на первый день. Чувствительность при прогнозировании летального исхода составила 83,1%, а специфичность — 75,3%. Однако следует отметить, что в данной работе анализировали лабораторные показатели, представление которых обладает высокой степенью стандартизации.

J.L. Warner и соавт. [17] при помощи алгоритма решающего дерева оценивали, насколько точно он может предсказать стадию рака легких у пациентов, опи-

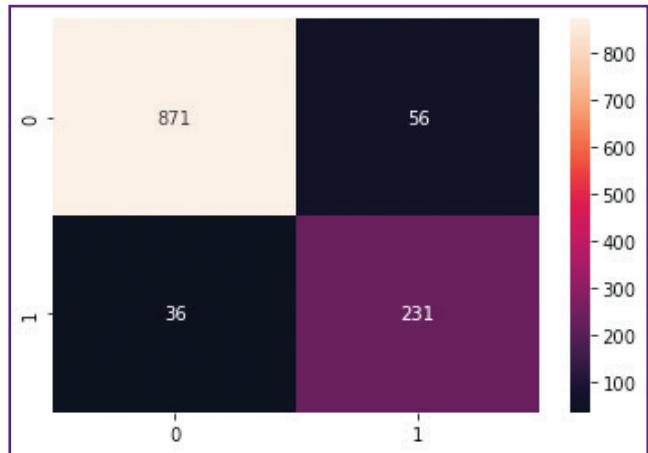


Рис. 4. Матрица ошибок:

по вертикали — истинная оценка КТ-исследования: 0 — исследование без признаков внутримозговых кровоизлияний (истинно-отрицательный результат); 1 — исследование с признаками кровоизлияний (истинно-положительный результат);

по горизонтали — оценки с помощью разработанного алгоритма: 0 — наличие патологии выявлено ошибочно (ложноположительный результат); 1 — отсутствие патологии указано ошибочно (ложноотрицательный результат)

раясь на имеющиеся неструктурированные протоколы диагностических описаний. В результате анализа методами машинного обучения было классифицировано 751 880 текстовых медицинских документов от 2327 пациентов. Несмотря на наличие значительных вариаций описаний в документации, точность оценки стадии рака легких алгоритмом получилась достаточно высокой — 0,906 (95% CI: 0,873–0,939).

D.A. Szlosek и J.M. Ferretti [18] рассматривали возможность использования алгоритмов машинного обучения для обработки естественного языка с целью автоматизации оценки систем поддержки принятия клинических решений в электронных системах медицинской документации. Набор данных содержал информацию о результатах КТ головного мозга 3621 пациента, которые получили легкую черепно-мозговую травму. Классификатор, построенный на основе алгоритма дерева решений, продемонстрировал чувствительность, равную 57,75%, но гораздо более высокую специфичность — 98,68%.

Разработанный нами алгоритм также показал чуть более низкую чувствительность, что проявилось в большем количестве ложноположительных оценок, нежели ложноотрицательных. По большей части это обусловлено использованием в протоколах КТ-исследования и стоп-фраз, и описания внечерепной патологии с помощью ключевых слов. Например, стоп-фраза «свежих кровоизлияний или ишемических изменений в головном мозге не выявлено» приводилась в текстовом протоколе вместе с описанием кровоизлияния в мягкие ткани лица. Тем не менее

полученные нами значения чувствительности выше представленных в литературе.

Ложноотрицательные срабатывания алгоритма связаны с «конфликтом» одновременно присутствующей в протоколе стоп-фразы, которая трактуется алгоритмом как признак отсутствия патологии (например, «данных за внутричерепную гематому при данном исследовании не выявлено»), и ключевых слов, употребляемых для описания, к примеру, небольшой зоны геморрагического пропитывания тканей мозга.

Основная проблема, которая возникает при использовании инструментов методов NLP на основе машинного обучения, в частности с применением алгоритмов дерева решений, — это отсутствие стандартизации описаний. Такой вид машинного обучения предполагает простую классификацию, и для его успешного использования подходит тот тип данных, который подвергнут унификации, например численные значения лабораторных исследований или категории в системе анализа и протоколирования результатов лучевых исследований молочной железы BI-RADS [19].

Полученные в настоящем исследовании метрики диагностической точности указывают на возможность практического применения разработанного алгоритма в соответствии с требованиями методических рекомендаций «Клинические испытания программного обеспечения на основе интеллектуальных технологий (лучевая диагностика)» [20].

Данный алгоритм может быть использован на первом этапе подготовки наборов данных для первичного, грубого отбора КТ-исследований с необходимыми признаками из большого массива информации, например из прямой выгрузки всех КТ-исследований головного мозга за год. После этого полученный набор КТ-исследований можно отправлять на дальнейший анализ, например с применением нейронных сетей, для более тонкого отбора по признакам (если, к примеру, нужны только случаи, где не проводилось хирургическое вмешательство, или же КТ-снимки с конкретным типом кровоизлияния). Первичные наборы данных могут быть использованы также для обучения или тестирования диагностических сервисов на основе искусственного интеллекта.

Разработанный алгоритм может с успехом применяться на этапе после оказания медицинской помощи в любом из типов медицинских учреждений для контроля качества работы врачей и упрощения подготовки статистических отчетов.

Однако также следует отметить ограничения представленной работы. На данный момент она является пилотной для классификации текстовых заключений КТ головного мозга с признаками ВЧК. У разработанного алгоритма дерева решений выявлены следующие недостатки: ложные срабатывания, сложности с классификацией неструктурированного текста с множеством вариаций описательных значений наличия и

отсутствия патологии, необходимость ручного пересмотра исследований для обеспечения контроля качества. Эти недостатки указывают на необходимость усложнения классификатора и использования других подходов машинного обучения, в том числе нейронных сетей.

Заключение

Разработанный и обученный алгоритм анализа текстовых протоколов описаний на основе модели дерева решений показал высокую точность работы при отборе КТ-исследований головного мозга с признаками внутричерепных кровоизлияний. Он может применяться для решения задач бинарной классификации и оптимизации создания соответствующих наборов диагностических исследований, которые будут использованы для обучения и валидации медицинских сервисов на основе искусственного интеллекта, направленных на диагностику кровоизлияний по данным КТ головного мозга. Кроме того, после соответствующего обучения он может быть применен для анализа и бинарной классификации любых других медицинских текстов, а также для контроля диагностики и медицинской помощи.

Финансирование исследования. Публикация подготовлена при поддержке гранта Российского научного фонда №22-25-20231, <https://rscf.ru/project/22-25-20231/>.

Конфликт интересов. Авторы подтверждают отсутствие конфликтов интересов.

Литература/References

1. Белолипецкая А.Е., Головина Т.А., Полянин А.В. Цифровая трансформация сферы здравоохранения: компетентностный подход. *Проблемы социальной гигиены, здравоохранения и истории медицины* 2020; 28(S): 694–700, <https://doi.org/10.32687/0869-866x-2020-28-s1-694-700>.
2. Belolipetskaya A.E., Golovina T.A., Polyinin A.V. Digital transformation of healthcare: a competency-based approach. *Problemy sotsialnoi gigieny, zdravookhraneniya, i istorii meditsiny* 2020; 28(S): 694–700, <https://doi.org/10.32687/0869-866x-2020-28-s1-694-700>.
3. Полищук Н.С., Ветшева Н.Н., Косарин С.П., Морозов С.П., Кузьмина Е.С. Единый радиологический информационный сервис как инструмент организационно-методической работы Научно-практического центра медицинской радиологии Департамента здравоохранения г. Москвы (аналитическая справка). *Радиология — практика* 2018; 1: 6–17.
4. Polishchuk N.S., Vetsheva N.N., Kosarin S.P., Morozov S.P., Kuz'mina E.S. Unified Radiological Information Service as a key element of organizational and methodical work of Research and Practical Center of Medical Radiology. *Radiologia — praktika* 2018; 1: 6–17.
5. Buchlak Q.D., Milne M.R., Seah J., Johnson A., Samarasinghe G., Hachey B., Esmaili N., Tran A., Leveque J.C., Farokhi F., Goldschlager T., Edelstein S., Brothie P. Charting

the potential of brain computed tomography deep learning systems. *J Clin Neurosci* 2022; 99: 217–223, <https://doi.org/10.1016/j.jocn.2022.03.014>.

4. Kuo W., Häne C., Mukherjee P., Malik J., Yuh E.L. Expert-level detection of acute intracranial hemorrhage on head computed tomography using deep learning. *Proc Natl Acad Sci U S A* 2019; 116(45): 22737–22745, <https://doi.org/10.1073/pnas.1908021116>.

5. Ginat D.T. Analysis of head CT scans flagged by deep learning software for acute intracranial hemorrhage. *Neuroradiology* 2020; 62(3): 335–340, <https://doi.org/10.1007/s00234-019-02330-w>.

6. Pons E., Braun L.M.M., Hunink M.G.M., Kors J.A. Natural language processing in radiology: a systematic review. *Radiology* 2016; 279(2): 329–343, <https://doi.org/10.1148/radiol.16142770>.

7. Wang Y., Sohn S., Liu S., Shen F., Wang L., Atkinson E.J., Amin S., Liu H. A clinical text classification paradigm using weak supervision and deep representation. *BMC Med Inform Decis Mak* 2019; 19(1): 1, <https://doi.org/10.1186/s12911-018-0723-6>.

8. Vo T.H., Nguyen N.T.K., Kha Q.H., Le N.Q.K. On the road to explainable AI in drug-drug interactions prediction: a systematic review. *Comput Struct Biotechnol J* 2022; 20: 2112–2123, <https://doi.org/10.1016/j.csbj.2022.04.021>.

9. Chen J., Druhl E., Polepalli Ramesh B., Houston T.K., Brandt C.A., Zulman D.M., Vimalananda V.G., Malkani S., Yu H. A natural language processing system that links medical terms in electronic health record notes to lay definitions: system development using physician reviews. *J Med Internet Res* 2018; 20(1): e26, <https://doi.org/10.2196/jmir.8669>.

10. Chen P.H. Essential elements of natural language processing: what the radiologist should know. *Acad Radiol* 2020; 27(1): 6–12, <https://doi.org/10.1016/j.acra.2019.08.010>.

11. Sysoev O., Bartoszek K., Ekström E.C., Ekholm Selling K. PSICA: decision trees for probabilistic subgroup identification with categorical treatments. *Stat Med* 2019; 38(22): 4436–4452, <https://doi.org/10.1002/sim.8308>.

12. Hostettler I.C., Muroi C., Richter J.K., Schmid J., Neidert M.C., Seule M., Boss O., Pangalu A., Germans M.R., Keller E. Decision tree analysis in subarachnoid hemorrhage: prediction of outcome parameters during the course of aneurysmal subarachnoid hemorrhage using decision tree analysis. *J Neurosurg* 2018; 129(6): 1499–1510, <https://doi.org/10.3171/2017.7.jns17677>.

13. He B., Guan Y., Dai R. Classifying medical relations in clinical text via convolutional neural networks. *Artif Intell Med* 2019; 93: 43–49, <https://doi.org/10.1016/j.artmed.2018.05.001>.

14. Qing L., Linhong W., Xuehai D. A novel neural network-based method for medical text classification. *Future Internet* 2019; 11(12): 255, <https://doi.org/10.3390/fi11120255>.

15. Donnelly L.F., Grzeszczuk R., Guimaraes C.V. Use of natural language processing (NLP) in evaluation of radiology reports: an update on applications and technology advances. *Semin Ultrasound CT MR* 2022; 43(2): 176–181, <https://doi.org/10.1053/j.sult.2022.02.007>.

16. Vrigazova B. The proportion for splitting data into training and test set for the bootstrap in classification problems. *Bus Syst Res* 2021; 12(1): 228–242, <https://doi.org/10.2478/bsrj-2021-0015>.

17. Warner J.L., Levy M.A., Neuss M.N. ReCAP: feasibility and accuracy of extracting cancer stage information from narrative electronic health record data. *J Oncol Pract* 2016; 12(2): 157–158, <https://doi.org/10.1200/jop.2015.004622>.

18. Szlosek D.A., Ferretti J.M. Using machine learning and natural language processing algorithms to automate the evaluation of clinical decision support in electronic medical record systems. *EGEMS (Wash DC)* 2016; 4(3): 1222, <https://doi.org/10.13063/2327-9214.1222>.

19. Davidson E.M., Poon M.T.C., Casey A., Grivas A., Duma D., Dong H., Suárez-Paniagua V., Grover C., Tobin R., Whalley H., Wu H., Alex B., Whiteley W. The reporting quality of natural language processing studies: systematic review of studies of radiology reports. *BMC Med Imaging* 2021; 21(1): 142, <https://doi.org/10.1186/s12880-021-00671-8>.

20. Морозов С.П., Владимирский А.В., Кляшторный В.Г., Андрейченко А.Е., Кульберг Н.С., Гомболевский В.А., Сергунова К.А. Клинические испытания программного обеспечения на основе интеллектуальных технологий (лучевая диагностика). Серия «Лучшие практики лучевой и инструментальной диагностики». М; 2019; 51 с.

Morozov S.P., Vladimirskiy A.V., Klyashtornyy V.G., Andreychenko A.E., Kul'berg N.S., Gombolevskiy V.A., Sergunova K.A. *Klinicheskie ispytaniya programmogo obespecheniya na osnove intellektual'nykh tekhnologiy (luhevaya diagnostika)*. Seriya "Luchshie praktiki luhevoy i instrumental'noy diagnostiki" [Clinical trials of software based on intelligent technologies (diagnostic radiology). Series "Best practices of radiological and instrumental diagnostics"]. Moscow; 2019; 51 p.