

DOI: <https://doi.org/10.17816/DD101099>

# Возможности и ограничения использования инструментов машинной обработки текстов в лучевой диагностике

Д.Ю. Кокина, В.А. Гомболевский, К.М. Арзамасов, А.Е. Андрейченко, С.П. Морозов

Научно-практический клинический центр диагностики и телемедицинских технологий, Москва, Российская Федерация

## АННОТАЦИЯ

**Обоснование.** В радиологии важную информацию содержат не только медицинские изображения, но и сопровождающие их текстовые описания, создаваемые врачами-рентгенологами. Идентификация протоколов исследований, содержащих определённые данные, и извлечение этих данных может быть полезным в первую очередь для клинических задач, однако, учитывая большой объём таких данных, необходима разработка машинных алгоритмов анализа.

**Цель** — оценить возможности и ограничения использования инструментов машинной обработки текстов для поиска патологий в протоколах лучевых исследований.

**Материалы и методы.** Для создания первого прототипа алгоритма автоматического анализа протоколов были выбраны исследования молочных желёз (маммография) и органов грудной клетки (рентгенография, флюорография, компьютерная томография и низкодозная компьютерная томография), выполненные в лечебно-профилактических учреждениях Москвы, которые участвовали в эксперименте по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений. Для каждого вида исследований был первоначально составлен словарь ключевых слов, соответствующий наличию или отсутствию целевых патологий. После первичной автоматической разметки протоколов разработанным инструментом производились выборочная оценка и валидация результатов врачом-рентгенологом. Количество протоколов, проанализированных врачом для обучения и валидации алгоритмов, составило 977 для маммографии, 3196 для рентгенографии, 1608 для флюорографии, 4074 для компьютерной и 398 для низкодозной компьютерной томографии органов грудной клетки. Для окончательного тестирования разработанных алгоритмов были дополнительно размечены тестовые датасеты из 1032 исследований для маммографии, 544 для флюорографии/рентгенографии, 5000 для компьютерной и 1082 для низкодозной компьютерной томографии органов грудной клетки.

**Результаты.** Наилучшие результаты достигнуты в поиске признаков вирусной пневмонии по протоколам компьютерной томографии органов грудной клетки (точность 0,996, чувствительность 0,998, специфичность 0,989) и рака молочной железы по протоколам маммографии (точность 1,0, чувствительность 1,0, специфичность 1,0). При поиске алгоритмом признаков рака лёгкого метрики получились следующими: точность 0,895, чувствительность 0,829, специфичность 0,936, а при поиске патологических изменений органов грудной клетки в протоколах рентгенографии и флюорографии точность составила 0,912, чувствительность — 1,000, специфичность — 0,844.

**Заключение.** Машинные методы с высокой точностью могут быть использованы с целью автоматической классификации текстов рентгенологических протоколов маммографии и компьютерной томографии органов грудной клетки для поиска вирусной пневмонии. Для поиска признаков рака лёгкого в модальности компьютерной и низкодозной компьютерной томографии, а также патологических изменений в протоколах рентгенографии и флюорографии органов грудной клетки достигнутой точности достаточно для успешного применения в целях автоматизированного сравнения работы врачей и моделей искусственного интеллекта.

**Ключевые слова:** протоколы рентгенологических исследований; COVID-19-пневмония; рак лёгкого; рак молочной железы; обработка естественного языка.

## Как цитировать

Кокина Д.Ю., Гомболевский В.А., Арзамасов К.М., Андрейченко А.Е., Морозов С.П. Возможности и ограничения использования инструментов машинной обработки текстов в лучевой диагностике // *Digital Diagnostics*. 2022. Т. 3, № 4. С. 374–383. DOI: <https://doi.org/10.17816/DD101099>

Рукопись получена: 18.02.2022

Рукопись одобрена: 24.11.2022

Опубликована: 19.12.2022

DOI: <https://doi.org/10.17816/DD101099>

# Possibilities and limitations of using machine text-processing tools in Russian radiology reports

Daria.Yu. Kokina, Victor A. Gombolevskiy, Kirill M. Arzamasov, Anna E. Andreychenko, Sergey P. Morozov

Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies, Moscow, Russian Federation

## ABSTRACT

**BACKGROUND:** In radiology, important information can be found not only in medical images, but also in the accompanying text descriptions created by radiologists. Identification of study protocols containing certain data and extraction of these data can be useful primarily for clinical problems; however, given the large amount of such data, the development of machine analysis algorithms is necessary.

**AIM:** To estimate the possibilities and limitations of using a tool for machine processing of radiology reports to search for pathological findings.

**MATERIALS AND METHODS:** To create an algorithm for automatic analysis of radiology reports, use cases were selected that participated in the experiment on the use of innovative technologies in the computer vision for the analysis of medical images in 2020. Mammography, chest X-ray, chest computed tomography (CT), and LDCT, were among the use cases performed in Moscow. A dictionary of keywords has been compiled. After the automatic marking of the reports by the developed tool, the results were assessed by a radiologist. The number of protocols analyzed by the radiologist for training and validation of the algorithms was 977 for mammography, 4,804 for all chest X-ray scans, 4,074 for chest CT, and 398 for chest LDCT. For the final testing of the developed algorithms, test datasets of 1,032 studies for mammography, 544 for chest X-ray, 5,000 for CT of the chest, and 1,082 studies for the LDCT of the chest were additionally labeled.

**RESULTS:** The best results were achieved in the search for viral pneumonia in chest CT reports (accuracy 0.996, sensitivity 0.998, and specificity 0.989) and breast cancer in mammography reports (accuracy 1.0, sensitivity 1.0, and specificity 1.0). When searching for signs of lung cancer by the algorithm, the metrics were as follows: accuracy 0.895, sensitivity 0.829, and specificity 0.936, when searching for pathological changes in the chest organs in radiography and fluorography protocols (accuracy 0.912, sensitivity 1.000, and specificity 0.844).

**CONCLUSIONS:** Machine methods with high accuracy can be used to automatically classify the radiology reports of mammography and chest CT with viral pneumonia. The achieved accuracy is sufficient for successful application to automatically compare the conclusions of physicians and artificial intelligence models when searching for signs of lung cancer in chest CT and LDCT, pathological findings in chest X-ray.

**Keywords:** radiology reports, COVID-19 pneumonia, lung cancer, breast cancer, natural language processing

## To cite this article

Kokina DYu, Gombolevskiy VA, Arzamasov KM, Andreychenko AE, Morozov SP. Possibilities and limitations of using machine text-processing tools in Russian radiology reports. *Digital Diagnostics*. 2022;3(4):374–383. DOI: <https://doi.org/10.17816/DD101099>

Received: 18.02.2022

Accepted: 24.11.2022

Published: 19.12.2022

DOI: <https://doi.org/10.17816/DD101099>

# 在辐射诊断中使用文本机床的可能性和局限性

Daria.Yu. Kokina, Victor A. Gombolevskiy, Kirill M. Arzamasov, Anna E. Andreychenko, Sergey P. Morozov

Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies, Moscow, Russian Federation

## 简评

**论证。**在放射学中，重要信息不仅包括在医学图像中，还包括在放射科医生创建的随附文本描述中。包含某些数据的研究方案的识别和这些数据的提取首先可能对临床问题有用，但是，鉴于大量此类数据，机器分析算法的开发是必要的。

**研究目的**是评估使用文本处理工具在放射学协议中搜索病理的可能性和局限性。

**材料与方法。**为了创建自动协议分析算法的第一个原型，选择了参与使用计算机视觉领域的创新技术进行医学图像分析的实验的研究。这些研究包括在莫斯科医疗机构进行的乳房X光检查、胸部X光摄影、胸部X线间接照相、胸部CT和LDCT。对于每种类型的研究，最初都编制了一个关键词词典，对应于目标病理学的存在与否。在使用开发的工具对协议进行初始自动标记之后，放射科医生对结果进行了选择性评估和验证。医生为训练和验证算法而分析的协议数量为977个乳房X线照相术、3196个射线照相术、1608个荧光照相术、4074个胸部CT和398个胸部LDCT。为了对开发的算法进行最终测试，额外标记了1032项乳房 X线照相术研究、544项荧光照相/射线照相术、5000项胸部CT研究和1082项胸部LDCT研究的测试数据集。

**结果。**最好结果是根据胸部CT协议（精确度0.996，灵敏度0.998，特异性0.989）和乳房X光检查协议（精确度1.0，灵敏度1.0，特异性1.0）分别在寻找病毒性肺炎迹象和寻找乳腺癌迹象的方面取得的。当通过该算法搜索肺癌征兆时，指标如下：精确度0.895，灵敏度0.829，特异性0.936，以及在射线照相和荧光照相术协议中搜索胸部器官的病理变化时为精确度0.912，灵敏度1.000，特异性0.844。

**结论。**机器方法可用于乳腺X线检查和胸部CT检查文本的自动分类，以寻找病毒性肺炎。在胸部CT和LDCT模式中寻找肺癌征象，在胸部X线摄影和荧光摄影协议中寻找病理变化，所达到的准确性足以成功应用于医生和人工智能模型工作的自动比较。

**关键词：**X射线协议；COVID-19肺炎；肺癌；乳腺癌；自然语言处理。

## To cite this article

Kokina DYu, Gombolevskiy VA, Arzamasov KM, Andreychenko AE, Morozov SP. 在辐射诊断中使用文本机床的可能性和局限性. *Digital Diagnostics*. 2022;3(4):374–383. DOI: <https://doi.org/10.17816/DD101099>

收到: 18.02.2022

接受: 24.11.2022

发布日期: 19.12.2022

## ОБОСНОВАНИЕ

Протоколы лучевых исследований содержат в себе текстовую медицинскую информацию, включающую предварительный диагноз, клинические данные, описательную часть с характеристиками изменений исследуемых органов и систем, рентгенологический диагноз или заключение и рекомендации по наблюдению [1, 2]. Эта информация может быть использована в комплексной диагностике и лечении заболеваний, прогнозировании исходов и мониторинга состояния, а также вне клинических целей, например для организационных, статистических или исследовательских задач.

Протоколы лучевых исследований обладают рядом особенностей, среди которых разнообразие повествовательных стилей, использование телеграфной речи, лексические и терминологические вариации, разнообразный порядок слов, многообразие сокращений, акронимов [3]. Отдельной особенностью, характерной для всей медицинской информации, является использование терминологии, оценка которой зачастую невозможна без наличия профильного образования. Кроме того, русскоязычные протоколы имеют также ряд специфических свойств, например, более свободные правила синтаксиса, лексическое разнообразие; врачи-рентгенологи используют нестандартные аббревиатуры, сложные грамматические конструкции, объёмные фразы, затрудняющие интерпретацию, множество вариантов обозначения отрицания [4]. Лексические вариации характерны для лучевой диагностики в целом, однако в русскоязычной лучевой диагностике это разнообразие ещё более широко (например, «затемнение» может описываться как «затенение», «тень», «инфильтрат», «участок пониженной прозрачности», «участок повышенной плотности», «участок пониженной воздушности», «фокус», «уплотнение» и огромное количество других вариантов для одной группы изменений). Напротив, вариативность в англоязычной лучевой диагностике регулируется нормативными документами, рекомендациями и т.д. Таким образом, протоколы лучевых исследований содержат большое количество текстовой, неструктурированной и специализированной информации, что приводит к определённым трудностям в её использовании исключительно автоматизированными методами.

В настоящее время исследуется использование инструментов обработки естественного языка для структуризации и стандартизации протоколов, выделения необходимой для клинических специалистов информации, автоматической замены специфической терминологии, в том числе для пациентов, с использованием более понятной лексики или переводом на другие языки [1, 2]. Идентификация протоколов исследований, содержащих определённые данные, и извлечение этих данных может быть полезным в первую очередь для клинических задач [1]. Так, имеются работы, посвящённые выявлению

протоколов описания костно-мышечной системы с признаками переломов костей, компьютерной томографии с признаками тромбоэмболии лёгочных артерий, узлов лёгких и т.д. [3, 5, 6].

Разработка алгоритма для машинной обработки русскоязычных протоколов необходима для работы с большим объёмом данных и использования результатов такого анализа для оценки медицинских изображений в совокупности с описанием и заключением врача.

**Цель исследования** — оценить возможности и ограничения использования инструментов машинной обработки текстов для поиска патологий различного характера в протоколах лучевых исследований.

## МАТЕРИАЛЫ И МЕТОДЫ

### Разработка инструмента по оценке текстовых протоколов лучевых исследований

Настоящая работа выполнена в рамках ранее одобренного этическим комитетом исследования; выписка из протокола № 2 Независимого этического комитета Московского регионального отделения Российского общества рентгенологов и радиологов (НЭК МРО РОПР) от 20 февраля 2020 г., также зарегистрированного на ClinicalTrials (NCT04489992).

Инструмент по оценке текстовых протоколов лучевых исследований разрабатывался в рамках московского эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и применялся для сопоставления результатов оценки наличия патологии на медицинских изображениях сервисами искусственного интеллекта и врачами-рентгенологами.

Протоколы исследований включали маммографию, рентгенографию и флюорографию органов грудной клетки (ОГК), компьютерную томографию (КТ) и низкодозную компьютерную томографию (НДКТ) ОГК; все исследования были выполнены в лечебно-профилактических учреждениях Департамента здравоохранения Москвы в 2020 г. В работе использовались анонимизированные протоколы лучевых исследований.

Основной целью было создание автоматизированного алгоритма для автоматического анализа протоколов рентгенологических исследований на предмет наличия или отсутствия искомых патологических изменений. Основой для выделения целевых патологий и разработки словаря были базовые требования к результатам сервисов искусственного интеллекта (<https://mosmed.ai/>).

Для рентгенографии и флюорографии ОГК целевые патологические признаки включали плевральный выпот, пневмоторакс, ателектаз, очаг, инфильтрацию/консолидацию, диссеминацию, полость с распадом или уровнем жидкости, кальцинат и нарушение целостности коркового слоя (перелом). Для КТ и НДКТ целевая патология включала солидные и субсолидные узлы объёмом более

100 мм<sup>3</sup>. Другой группой патологии для КТ ОГК были изменения, коррелирующие с признаками коронавирусной инфекции (COVID-19); использована классификация по степени тяжести, согласно временным методическим рекомендациям «Профилактика, диагностика и лечение новой коронавирусной инфекции (COVID-19)» Министерства здравоохранения Российской Федерации, а также методическим рекомендациям ГБУЗ НПКЦ ДиТ ДЗМ «Лучевая диагностика коронавирусной болезни (COVID-19): организация, методология, интерпретация результатов» [7, 8]. Для маммографии использовалась система анализа и протоколирования результатов лучевых исследований молочной железы BI-RADS 3–6 (Breast Imaging Reporting and Data System) [9].

### Программно-технические решения для предразметки текстовых протоколов

Были разработаны программно-технические решения для предразметки текстовых протоколов для каждого типа исследований, основанные на методах обработки естественного языка в комбинации с экспертным мнением врачей-рентгенологов. Разработка алгоритмов-предразметчиков осуществлялась итеративным образом в несколько основных этапов для каждой модальности.

1. Формирование первичного набора ключевых признаков для поиска указаний на определённые патологии. Среди признаков рассматривались ключевые слова и фразы, указания на размеры и, если это необходимо, «стоп»-слова и фразы. Первичный набор ключевых слов, составленный врачом-рентгенологом, включал общепринятые и наиболее часто употребляемые врачами-рентгенологами термины. «Стоп»-слова включали нецелевые патологические изменения или непатологические находки (изменения других органов в зоне сканирования, анатомические варианты и т.д.).
2. Перевод наборов признаков на машинный язык с помощью высокоуровневого языка программирования Python.

В настоящей работе мы использовали Python 3.8, для работы требуются библиотеки pandas=1.1.3, numpy=1.19.2, re=2.2.1, nltk=3.5. Входными данными для разработанной программы являются текстовые отчёты в табличной форме (.csv, .xlsx), содержащие как описание исследования, так и его заключение. Результаты работы представлены исходными данными с меткой «0» или «1», где «0» означает отсутствие признака/патологии, «1» — наличие признака/патологии.

Модуль для поиска признаков COVID-19 по текстовым заключениям КТ ОГК основан на методах обработки естественного языка (natural language processing, NLP) и классификаторе из семейства алгоритмов машинного обучения. В качестве выхода указывает наличие или отсутствие признаков COVID-19, а также КТ-степень поражения лёгких.

Модуль для поиска признаков злокачественных новообразований молочной железы по данным протоколов маммографии, включая заключение и описание, выявляет признаки злокачественных новообразований молочных желёз по классификации BI-RADS, выдавая в качестве ответа класс BI-RADS и бинарную классификацию (аналогичную скрининговой шкале). Требования к описанию результатов маммологического исследования регламентируют обязательную классификацию исследования по шкале BI-RADS; с другой стороны, написание формулировки BI-RADS 1 на практике имеет широкий вариатив. Врачи могут использовать написание с разным регистром, разными пунктуационными знаками и, что самое главное, с использованием различных раскладок. По этой причине обычный анализ (parsing) текста оказывается не всегда эффективным, приводя к ложным пропускам. Применение NLP для данной задачи сводится к функции извлечения информации.

Модуль для поиска признаков злокачественных новообразований лёгкого по текстовым протоколам (описанию и заключению) КТ/НДКТ ОГК идентифицирует признаки злокачественных новообразований по сочетанию ключевых слов (ключей) и параметров (размеров); основан на NLP.

Модуль для поиска различных признаков патологии в текстах описания и заключения для рентгенографических и флюорографических исследований ОГК идентифицирует признаки патологии по словарю ключевых слов ('признак RG/FLG': 'ключевое слово1', 'ключевое слово2', 'ключевое слово3'...).

3. Разметка протоколов программой-предразметчиком, созданной на этапе 2.
4. Отбор уникальных разнообразных (по патологическим признакам) протоколов из размеченной выборки на этапе 3. В ходе нашей работы мы неоднократно убеждались в использовании шаблонных формулировок врачами-рентгенологами. Обучение алгоритмов NLP на шаблонных формулировках позволит быстро переобучить модель. Нами была поставлена задача максимально широкого применения разрабатываемого алгоритма, благодаря обогащению обучающего набора данных разнообразными редкими (уникальными) формулировками.
5. Ручная верификация машинной разметки и оценка точности алгоритма-предразметчика. Точность машинной разметки оценивалась как процент правильно размеченных протоколов. Ручная верификация проводилась многократно для каждой отдельной задачи врачом-рентгенологом.
6. Формирование списка корректировок, включая дополнительные «стоп»-слова и фразы, ключевые слова и прочие рекомендации с целью улучшения качества предразметчика.
7. Добавление верифицированных протоколов в базу. На данном этапе происходило формирование разнообразных, уникальных и размеченных выборок с сохранением

баланса классов. При формировании выборок исследований, размеры которых приведены ниже, брались все исследования за ограниченный промежуток времени (с января 2019 по август 2020 г.). Баланс классов соответствовал таковому в генеральной совокупности. Ожидаемое распространение по патологиям: КТ COVID-19: 20% — норма, 80% — патология; маммография: 95% — норма, 5% — патология; флюорография: 95% — норма, 5% — патология; рентгенография ОГК: 75% — норма, 25% — патология.

8. Повторение этапов 2–7 итеративным образом до тех пор, пока точность предразметчика не достигнет 98%. Данное значение было выбрано исходя из максимального уровня точности NLP для отдельных медицинских задач, обнаруженного при анализе медицинской литературы и составлявшего 97% [10].

В общей сложности количество протоколов, проанализированных врачом-рентгенологом при разработке алгоритмов-предразметчиков, составляет 977 для маммографии, 3196 для рентгенографии, 1608 для флюорографии, 4074 для КТ ОГК и 398 НДКТ ОГК. В исследование включались все протоколы исследований, которые направлялись сервисам искусственного интеллекта в рамках эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы (<https://mosmed.ai/>). Исключению подвергались только те протоколы, которые не имели заполненных разделов «описание» и «заключение».

Для повышения качества и ускорения автоматической разметки текстовых протоколов были применены методы машинного обучения, позволяющие учитывать сложные семантические структуры предложений в протоколах, для поиска признаков COVID-19-пневмонии по данным КТ ОГК. В будущем планируется разработка интеллектуализированных алгоритмов для поиска признаков патологий в протоколах маммографии, КТ (для поиска рака лёгкого), флюорографии и рентгенографии.

Модуль обработки протоколов КТ COVID-19 состоял из трёх функций: (1) поиск заключений во входных данных по уже размеченной базе протоколов; (2) разметка оставшихся протоколов с помощью регулярного выражения; (3) разметка оставшихся протоколов, используя модели *k*-ближайших соседей (*k*-Nearest Neighbors, *k*NN). Работа этих функций была реализована последовательно. Одним из условий при разработке нашего инструмента была оптимизация быстродействия. Ранее мы уже упоминали частое использование шаблонных выражений при составлении протокола врачом-рентгенологом. По этой причине большая часть протоколов по КТ COVID-19 имеет одинаковый вид. Кроме того, ранее коллективом авторов была проведена большая работа по ручной разметке протоколов КТ ОГК с целевой

патологией COVID-19. Это упростило и ускорило работу алгоритма благодаря использованию простой функции логического сравнения, позволяя сравнивать протокол с имеющимися в базе данных. Далее остаётся часть протоколов, которых нет в базе данных ранее размеченных исследований, и для этого запускается существенно более медленная функция анализа текста при помощи регулярных выражений. При наличии протоколов, для которых регулярное выражение не обнаружило целевого образца (паттерна), запускается модель машинного обучения на основе *k*NN.

Разработанная нами архитектура обладает оптимальным сочетанием быстродействия и точности по сравнению с отдельным применением машинного обучения. Без машинного обучения невозможно было охватить весь объём протоколов. Обучающая выборка — 4074 предварительно размеченных протоколов. Такое количество протоколов было необходимо для обеспечения требуемого уровня точности, и было получено за несколько итераций тестирования и дообучения модели. Для функционирования данного модуля дополнительно импортировалась библиотека Sklearn (Scikit-learn). Обученный алгоритм был оценён на тестовой выборке и показал достаточно высокую точность (99,6%).

## РЕЗУЛЬТАТЫ

Создан перечень ключевых слов и «стоп»-слов по выбранным модальностям и патологиям с учётом особенностей протоколов. На протоколах исследований, выполненных в Москве, наилучшие результаты работы инструмента на основании разработанного словаря достигнуты в поиске признаков вирусной пневмонии COVID-19 по данным протоколов КТ ОГК: точность 0,996, чувствительность 0,998, специфичность 0,989 (TN<sup>\*</sup>=1115; FP<sup>\*\*</sup>=6; FN<sup>#</sup>=2; TP<sup>##</sup>=3837) и рака молочной железы по данным маммографических заключений: точность 1,0, чувствительность 1,0, специфичность 1,0 (TN=461; FP=0; FN=0; TP=571). При поиске признаков рака лёгкого в модальности КТ и НДКТ ОГК получились следующие метрики: точность 0,895, чувствительность 0,829, специфичность 0,936 (TN=619; FP=42; FN=72; TP=349), а при поиске патологических изменений ОГК в протоколах рентгенографии и флюорографии — 0,912; 1,000 и 0,844 соответственно (TN=259; FP=48; FN=0; TP=237).

\* TN (true negative — истинно отрицательный): предсказать отрицательный класс как отрицательный класс (количество), истинное значение — 1, прогноз — 1.

\*\* FP (false positive — ложноположительный): прогнозировать отрицательный класс как положительный класс (количество), истинное значение — 1, прогноз — 0.

# FN (false negative — ложноотрицательный): прогнозировать положительный класс как отрицательный класс (количество), истинное значение — 0, прогноз — 1.

## TP (true positive — истинно положительный): прогнозировать положительный класс как положительный класс (количество), истинное значение — 0, прогноз — 0.

## ОБСУЖДЕНИЕ

### Словарь для маммографических исследований

Наиболее простым для составления и использования оказался словарь для маммографических исследований, что связано с особенностями современных стандартизированных протоколов. Протоколы описания маммографических исследований наиболее структурированы, для них обязательно использование BI-RADS, они подлежат контролю в связи с высокой значимостью выявляемой патологии. В абсолютном большинстве случаев в заключении указана категория BI-RADS, в описательной части — изменения, на основании которых выставлена та или иная категория. Наличие в словаре категорий BI-RADS позволяет достигать высокой точности в автоматизированной обработке протоколов, достигающей точности 1.0. Данный инструмент может использоваться в других регионах (субъектах Российской Федерации), при необходимости, например для протоколов отличающейся структуры, возможно добавление ключевых определяющих слов («опухоль», «с-г»).

Ограничения в работе инструмента могут быть связаны с отсутствием категории BI-RADS в протоколе: обычно такие протоколы содержат информацию о невозможности оценки состояния молочных желёз в связи с ненадлежащим качеством снимков или другими причинами. Так как такие исследования выносятся в отдельную группу, возможен их целенаправленный поиск, например для осуществления контроля качества, отбора пациентов, нуждающихся в дообследовании. Разработанный инструмент при условии дальнейшего совершенствования может быть использован с другими целями, например для оценки соответствия протокола стандарту, сопоставления описательной части и заключения с целью аудита.

### Словарь для исследований на COVID-19-пневмонию по степеням тяжести

Высокие показатели точности были достигнуты также при применении инструмента для анализа исследований на предмет COVID-19-пневмонии по степеням тяжести, что также обусловлено структурой протоколов и однозначностью составленного словаря. Используемый словарь основан на стандартных степенях: КТ0 — отсутствие признаков вирусной пневмонии, КТ1–КТ4 — вирусная пневмония от лёгкой степени до критической, ПРОЧЕЕ — прочие изменения, не связанные с вирусной пневмонией. В период пандемии заключения содержали информацию об отсутствии или наличии признаков вирусной пневмонии, степени распространения, ассоциированной со степенью тяжести, а также в большинстве случаев вероятности

вирусной пневмонии. В условиях соблюдения указанных параметров для оценки инструментом достаточно текста заключения и словаря на основании вариаций написания степени поражения. Кроме того, описательные части протоколов в подавляющем большинстве имеют общие черты и единую терминологию.

Несмотря на общие черты большинства протоколов, существует небольшое количество вариаций написания заключения, связанных с разнообразными вариантами построения, использования терминологии, орфографическими, пунктуационными, лексическими особенностями языка врача, а также личным опытом работы с КТ-исследованиями. Иногда возникают диагностические неточности, связанные с особенностями КТ-картины вирусной пневмонии. В ситуациях, когда изменения по типу матового стекла, ретикулярной исчерченности и т.д. в представлении описывающего врача могут соответствовать другим заболеваниям, или существует другая сопутствующая патология, протокол и в особенности заключение могут содержать нетипичные по построению и терминологии предложения. В перечисленных ситуациях возможно наличие неопределённости в срабатывании инструмента, что позволяет сконцентрировать внимание на таких исследованиях и провести целенаправленный аудит, выявить неточности использования терминологии, а также использовать в клинических целях для определения сопутствующих заболеваний, требующих особого внимания и нуждающихся в дальнейшем наблюдении. Следует отметить, что в разных регионах и лечебных учреждениях возможно использование различных классификаций и протоколов описания вирусной пневмонии.

### Словарь для поиска признаков рака лёгкого

Более сложной оказалась задача по разработке ключевых слов и «стоп»-слов для поиска признаков рака лёгкого по данным протоколов КТ и НДКТ ОГК, что стало причиной меньшей точности алгоритма. Для протоколов НДКТ существует классификация Lung-RADS, использование которой могло бы максимально упростить задачу поиска подозрительных узлов [11]. Однако поиск категорий Lung-RADS в работе с исследованиями г. Москвы не позволяет оценить имеющиеся протоколы в полном масштабе, так как не все протоколы содержат такую категорию в заключении; кроме того, в 8,3% протоколов содержатся расхождения между описательной частью и заключением [12].

Разработка словаря для поиска признаков узелков и новообразований лёгких различного характера по-прежнему является актуальной задачей, сопряжённой с рядом сложностей и ограничений. Протоколы КТ и НДКТ ОГК, несмотря на наличие шаблонов и методических рекомендаций по описанию, имеют достаточно разнообразные варианты структуры, последовательности. Многие врачи-рентгенологи не используют традиционную рекомендуемую терминологию (например, узелки 4 мм

обозначены термином «образование»), что приводит к неправомерному использованию терминов [13].

С учётом существующих рекомендаций и желательной терминологии для поиска подозрительных изменений были приняты ключевые слова, соответствующие солидным узлам/очагам лёгких от 6 мм, а также образованиям (изменения размером более 3 см) [10–12]. Эти критерии имеют ряд ограничений, что может привести к ложноположительным или ложноотрицательным срабатываниям алгоритма. Так, например, при выполнении КТ ОГК случайно найденные солидные узлы лёгких рекомендуется оценивать с использованием рекомендаций Флейшнеровского сообщества (Fleischner), однако для их применения необходимы оценка паспортных данных, клинической информации, оценка факторов риска, наличия других заболеваний, в том числе неопластического характера [14].

При использовании критерия размера и основного набора ключевых слов невозможно полностью исключить доброкачественные изменения. Например, для исключения доброкачественных узлов с кальцификацией в структуре необходимо учитывать характер её распределения, зачастую не указанный в протоколах. В то же время кальцинаты могут быть описаны в структуре злокачественных новообразований. Крупные очаги могут быть описаны врачом в рамках других заболеваний, например туберкулёза, саркоидоза, бронхолитов различной этиологии.

Для выполнения задачи по сопоставлению данных протокола и результата обработки сервисом искусственного интеллекта существующих возможностей разработанного алгоритма достаточно. Дальнейшее улучшение инструмента с повышением его точности позволит не только использовать его в других регионах, но и поможет создать полезный инструмент для других задач.

С доработкой инструмента возможно создание более точного алгоритма, учитывающего необходимые факторы риска (наличие иммунодефицита, воспалительных процессов из паспортной части протокола, клинической информации, направительного диагноза и т.д.). Особенно важным данное усовершенствование может стать в оценке узелков лёгких у онкологических больных. Информацию о наличии онкологического заболевания можно получить из описательной части протокола, а также в электронной истории болезни. Оценка динамики изменения размера узелков с помощью инструмента и сопоставление с данными рекомендаций по ведению лёгочных узелков также могли бы стать перспективным направлением усовершенствования инструмента для соответствия его функциональных возможностей совершенствующимся моделям компьютерного зрения.

Отдельная группа ограничений, особенно актуальная в группе КТ ОГК, связана с широким и сложно

охватываемым списком «стоп»-слов. Это связано с особенностями исследования: в зону сканирования попадают органы брюшной полости, шеи, а также другие органы грудной клетки. Для большинства органов (например, щитовидной железы, печени, почек, надпочечников и т.д.) такие «стоп»-слова включают название самих органов. Однако такие анатомические структуры грудной клетки, как средостение, перикард, рёбра и грудные позвонки, мягкие ткани грудной стенки, диафрагма, не могут быть использованы как самостоятельные «стоп»-слова, так как нередко случаи, когда злокачественные новообразования лёгких обладают инвазивным ростом и поражают прилежащие ткани, что описывается врачом в совокупности («образование, распространяющееся в средостение»). В то же время часто встречаются различные самостоятельные новообразования перечисленных органов и анатомических структур, что приводит к большому количеству ложноположительных срабатываний алгоритма.

### Словарь ключевых слов для рентгенографии и флюорографии органов грудной клетки

Сложной областью для разработки и применения инструмента анализа текстов проколов оказались рентгенография и флюорография ОГК. Протоколы рентгенографии и флюорографии имеют множество вариаций формы, структуры, размера, используемых характеристик, при этом значительно варьирует терминология [15, 16].

Разработанный словарь ключевых слов для рентгенографии и флюорографии ОГК помимо общепринятых обозначений патологии включал специфические рентгенологические термины, например «затемнение», «фокус», «тень». Это сопряжено с рядом трудностей, так как такие термины могут использоваться для нецелевой патологии или обозначения анатомических структур («тень ребра»), дополнительных медицинских приспособлений («тень электрокардиостимулятора», «тень дренажной трубки»).

Определение патологии по бинарной классификации (норма/патология) с условием перечисленных затруднений возможно с достаточно высокой точностью, однако при необходимости разделения патологии на многочисленные классы (например, выпот, пневмоторакс, ателектаз, очаг и т.д.) существуют проблемы классификации находок по группам даже для самих врачей-рентгенологов, что связано и с ограниченными возможностями рентгеновского метода. Кроме того, одни и те же ключевые слова используются для обозначения совершенно разных рентгенологических находок.

В процессе разработки «стоп»-слов были учтены многочисленные варианты написания нормы с использованием различных лексических и синтаксических вариантов отрицания («в лёгких без затемнений», «без патологических тенеобразований в лёгких» и т.д.). С поступлением



новых данных необходимо постоянное расширение словаря. В настоящий момент полученные показатели точности для данного типа исследований позволяют решать основную задачу сопоставления результатов моделей искусственного интеллекта и врача.

## ЗАКЛЮЧЕНИЕ

С высокой точностью машинные методы могут быть использованы для автоматической классификации текстов рентгенологических протоколов маммографии и КТ ОГК для поиска вирусной пневмонии, что связано со структурированностью и наличием стандартизации описаний для данных исследований.

Для поиска признаков рака лёгкого в протоколах исследований КТ и НДКТ ОГК и патологических изменений в протоколах рентгенографии и флюорографии ОГК достигнутой точности достаточно для успешного применения в целях автоматизированного сравнения работы врачей и моделей искусственного интеллекта в отделениях лучевой диагностики. Меньшая точность обусловлена более свободной структурой протоколов, их диагностическими, лексическими и терминологическими особенностями.

## СПИСОК ЛИТЕРАТУРЫ

1. Sorin V., Barash Y., Konen E., Klang E. Deep learning for natural language processing in radiology: Fundamentals and a systematic review // *J Am College Radiol*. 2020. Vol. 17, N 5. P. 639–648. doi: 10.1016/j.jacr.2019.12.026
2. Monshi M.M., Poon J., Chung V. Deep learning in generating radiology reports: A survey // *Artif Intell Med*. 2020. N 106. P. 101878. doi: 10.1016/j.artmed.2020.101878
3. Banerjee I., Chen M.C., Lungren M.P., Rubin D.L. Radiology report annotation using intelligent word embeddings: Applied to multi-institutional chest CT cohort // *J Biomed Inform*. 2018. N 77. P. 11–20. doi: 10.1016/j.jbi.2017.11.012
4. Kivotova E., Maksudov B., Kuleev R., Ibragimov B. Extracting clinical information from chest x-ray reports: A case study for Russian language // *Conference: International Conference Nonlinearity, Information and Robotics (NIR)At: Innopolis, Russia, 2020*. P. 1–6. doi: 10.1109/NIR50484.2020.9290235
5. Lee C., Kim Y., Kim Y.S., Jang J. Automatic disease annotation from radiology reports using artificial intelligence implemented by a recurrent neural network // *Am J Roentgenol*. 2019. Vol. 212, N 4. P. 734–740. doi: 10.2214/AJR.18.19869
6. Yuan J., Zhu H., Tahmasebi A. Classification of pulmonary nodular findings based on characterization of change using radiology reports // *AMIA Jt Summits Transl Sci Proc*. 2019. Vol. 2019. P. 285–294.
7. Морозов С.П., Проценко Д.Н., Сметанина С.В., и др. Лучевая диагностика коронавирусной болезни (COVID-19): организация, методология, интерпретация результатов: препринт № ЦДТ 2020-II. Версия 2. Москва, 2020. 78 с.
8. Профилактика, диагностика и лечение новой коронавирусной инфекции 2019-nCoV. Временные методические ре-

## ДОПОЛНИТЕЛЬНО

**Источник финансирования.** Авторы заявляют об отсутствии внешнего финансирования при проведении исследования.

**Конфликт интересов.** Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

**Вклад авторов.** Все авторы подтверждают соответствие своего авторства международным критериям ICMJE (все авторы внесли существенный вклад в разработку концепции, проведение исследования и подготовку статьи, прочли и одобрили финальную версию перед публикацией).

## ADDITIONAL INFORMATION

**Funding source.** This study was not supported by any external sources of funding.

**Competing interests.** The authors declare that they have no competing interests.

**Authors' contribution.** All authors made a substantial contribution to the conception of the work, acquisition, analysis, interpretation of data for the work, drafting and revising the work, final approval of the version to be published and agree to be accountable for all aspects of the work.

комендации Министерства здравоохранения Российской Федерации // *Пульмонология*. 2019. Т. 29, № 6. С. 655–672. doi: 10.18093/0869-0189-2019-29-6-655-672

9. D'Orsi C.J., Sickles E.A., Mendelson E.B., et al. ACR BI-RADS atlas, breast imaging reporting and data system. Reston, VA, American College of Radiology; 2013.

10. Caliskan D., Zierk J., Kraska D., et al. First steps to evaluate an NLP tool's medication extraction accuracy from discharge letters // *Stud Health Technol Inform*. 2021. N 278. P. 224–230. doi: 10.3233/SHTI210073

11. American College of Radiology Committee on Lung-RADS. Lung-RADS Assessment Categories version 1.1. Режим доступа: <https://www.acr.org/-/media/ACR/Files/RADS/Lung-RADS/LungRADSAssessmentCategoriesv1-1.pdf>. Дата обращения: 01.01.2020.

12. Морозов С.П., Владимирский А.В., Гомбоевский В.А., и др. Искусственный интеллект: автоматизированный анализ текста на естественном языке для аудита радиологических исследований // *Вестник рентгенологии и радиологии*. 2018. Т. 99, № 5. С. 253–258.

13. Hansell D.M., Bankier A.A., MacMahon H., et al. Fleischner Society: glossary of terms for thoracic imaging // *Radiology*. 2008. Vol. 246, N 3. P. 697–722. doi: 10.1148/radiol.2462070712

14. MacMahon H., Naidich D.P., Goo J.M., et al. Guidelines for management of incidental pulmonary nodules detected on CT images: from the fleischner society 2017 // *Radiology*. 2017. Vol. 284, N 1. P. 228–243. doi: 10.1148/radiol.2017161659

15. Callister M.E., Baldwin D.R., Akram A.R., et al. British Thoracic Society Pulmonary Nodule Guideline Development Group; British Thoracic Society Standards of Care Committee. British Tho-

racic Society guidelines for the investigation and management of pulmonary nodules // *Thorax*. 2015. Vol. 70, Suppl 2. P. ii1–ii54. doi: 10.1136/thoraxjnl-2015-207168

16. Синицын В.Е., Комарова М.А., Мершина Е.А. Протокол рентгенологического описания: прошлое, настоящее, будущее // *Вестник рентгенологии и радиологии*. 2014. № 3. С. 35–40.

## REFERENCES

1. Sorin V, Barash Y, Konen E, Klang E. Deep learning for natural language processing in radiology: Fundamentals and a systematic review. *J Am Coll Radiol*. 2020;17(5):639–648. doi: 10.1016/j.jacr.2019.12.026
2. Monshi MM, Poon J, Chung V. Deep learning in generating radiology reports: A survey. *Artif Intell Med*. 2020;(106):101878. doi: 10.1016/j.artmed.2020.101878
3. Banerjee I, Chen MC, Lungren MP, Rubin DL. Radiology report annotation using intelligent word embeddings: Applied to multi-institutional chest CT cohort. *J Biomed Inform*. 2018;(77):11–20. doi: 10.1016/j.jbi.2017.11.012
4. Kivotova E, Maksudov B, Kulee R, Ibragimov B. Extracting clinical information from chest X-ray reports: A case study for Russian language. Conference: International Conference Nonlinearity, Information and Robotics (NIR)At: Innopolis, Russia; 2020. P. 1–6. doi: 10.1109/NIR50484.2020.9290235
5. Lee C, Kim Y, Kim YS, Jang J. Automatic disease annotation from radiology reports using artificial intelligence implemented by a recurrent neural network. *Am J Roentgenol*. 2019;212(4):734–740. doi: 10.2214/AJR.18.19869
6. Yuan J, Zhu H, Tahmasebi A. Classification of pulmonary nodular findings based on characterization of change using radiology reports. *AMIA Jt Summits Transl Sci Proc*. 2019;2019:285–294.
7. Morozov SP, Protsenko DN, Smetanina SV, et al. Radiation diagnostics of coronavirus disease (COVID-19): Organization, methodology, interpretation of results: Preprint, 2020-II. Version 2. Moscow; 2020. 78 p. (In Russ).
8. The prevention, diagnosis and treatment of the new coronavirus infection 2019-nCoV. Temporary guidelines Ministry of Health of the Russian Federation. *Pulmonologiya*. 2019;29(6):655–672. (In Russ). doi: 10.18093/0869-0189-2019-29-6-655-672

9. D'Orsi CJ, Sickles EA, Mendelson EB, et al. ACR BI-RADS Atlas, Breast Imaging Reporting and Data System. Reston, VA, American College of Radiology; 2013.
10. Caliskan D, Zierk J, Kraska D, et al. First steps to evaluate an NLP tool's medication extraction accuracy from discharge letters. *Stud Health Technol Inform*. 2021;(278):224–230. doi: 10.3233/SHTI210073
11. American College of Radiology Committee on Lung-RADS. Lung-RADS Assessment Categories version 1.1. Available from: <https://www.acr.org/-/media/ACR/Files/RADS/Lung-RADS/LungRADSAssessmentCategoriesv1-1.pdf>. Accessed: 01.01.2020.
12. Morozov SP, Vladzimirskiy AV, Gombolevskiy VA, et al. Artificial intelligence: natural language processing for peer-review in radiology. *J Radiol Nuclear Med*. 2018;99(5):253–258. (In Russ). doi: 10.20862/0042-4676-2018-99-5-253-258
13. Hansell DM, Bankier AA, MacMahon H, et al. Fleischner Society: glossary of terms for thoracic imaging. *Radiology*. 2008;246(3):697–722. doi: 10.1148/radiol.2462070712
14. MacMahon H, Naidich DP, Goo JM, et al. Guidelines for management of incidental pulmonary nodules detected on CT images: From the fleischner society 2017. *Radiology*. 2017;284(1):228–243. doi: 10.1148/radiol.2017161659
15. Callister ME, Baldwin DR, Akram AR, et al.; British Thoracic Society Pulmonary Nodule Guideline Development Group; British Thoracic Society Standards of Care Committee. British Thoracic Society guidelines for the investigation and management of pulmonary nodules. *Thorax*. 2015;70(Suppl 2):ii1–ii54. doi: 10.1136/thoraxjnl-2015-207168
16. Sinityn VE, Komarova MA, Mershina EA. Radiology report: past, present and future. *J radiol nuclear med*. 2014;(3):35–40. (In Russ).

## ОБ АВТОРАХ

### \* Кокина Дарья Юрьевна;

адрес: Россия, 127051, Москва, ул. Петровка, д. 24, стр. 1;  
ORCID: <https://orcid.org/0000-0002-1141-8395>;  
eLibrary SPIN: 9883-4656; e-mail: d.kokina@npcmr.ru

### Гомболевский Виктор Александрович, к.м.н.;

ORCID: <https://orcid.org/0000-0003-1816-1315>;  
eLibrary SPIN: 6810-3279; e-mail: g\_victor@mail.ru

### Арзамасов Кирилл Михайлович, к.м.н.;

ORCID: <http://orcid.org/0000-0001-7786-0349>;  
eLibrary SPIN: 3160-8062; e-mail: k.arzamasov@npcmr.ru

### Андрейченко Анна Евгеньевна, канд. физ.-мат. наук;

ORCID: <https://orcid.org/0000-0001-6359-0763>;  
eLibrary SPIN: 6625-4186; e-mail: a.andreychenko@npcmr.ru

### Морозов Сергей Павлович, д.м.н., профессор;

ORCID: <https://orcid.org/0000-0001-6545-6170>;  
eLibrary SPIN: 8542-1720; e-mail: spmoroz@gmail.com

## AUTHORS' INFO

### \* Daria Yu. Kokina;

address: 24-1 Petrovka str., 127051, Moscow, Russia;  
ORCID: <https://orcid.org/0000-0002-1141-8395>;  
eLibrary SPIN: 9883-4656; e-mail: d.kokina@npcmr.ru

### Victor A. Gombolevskiy, MD, Cand. Sci. (Med.);

ORCID: <https://orcid.org/0000-0003-1816-1315>;  
eLibrary SPIN: 6810-3279; e-mail: g\_victor@mail.ru

### Kirill M. Arzamasov, MD, Cand. Sci. (Med.);

ORCID: <http://orcid.org/0000-0001-7786-0349>;  
eLibrary SPIN: 3160-8062; e-mail: k.arzamasov@npcmr.ru

### Anna E. Andreychenko, Cand. Sci. (Phys.-Math.);

ORCID: <https://orcid.org/0000-0001-6359-0763>;  
eLibrary SPIN: 6625-4186; e-mail: a.andreychenko@npcmr.ru

### Sergey P. Morozov, MD, Dr. Sci. (Med.), Professor;

ORCID: <https://orcid.org/0000-0001-6545-6170>;  
eLibrary SPIN: 8542-1720; e-mail: spmoroz@gmail.com

\* Автор, ответственный за переписку / Corresponding author