

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

BACKGROUND

Radiology reports contain textual medical information, including a preliminary diagnosis, clinical data, descriptive characteristics of changes in organs and systems examined a radiologic diagnosis or a conclusion, and follow-up recommendations [1, 2]. This information can be used in complex diagnostics and treatment, outcome prediction, and condition monitoring and for organizational, statistical, or research purposes.

Radiology protocols have several features, including various narrative styles, using telegraphic speech, lexical and terminological variations, various word orders, abbreviations, and acronyms [3]. Special mention should be made of a characteristic of any medical information, such as the use of terminology, which is often impossible to be assessed by a person without special education. Russian protocols have also several specific properties, such as less strict syntax and lexical diversity. Radiologists use nonstandard abbreviations, complex grammatical constructions, long and difficult-to-interpret phrases, and various options to denote negation [4]. Lexical variations are typical for radiology in general; however, in Russian radiology, this diversity is even wider (e.g., “shadow” can be described as “shading,” “infiltrate,” “area of reduced transparency,” “area of increased density,” “area of reduced airiness,” “focus,” “compaction,” and various other options even for this group of changes alone). On the contrary, in English radiology, such variability is regulated by rules, recommendations, etc. Therefore, radiology reports contain a lot of textual, unstructured, and specialized information, which poses some difficulties when using exclusively automated methods.

Studies have focused on assessing the current use of natural language processing (NLP) tools for structuring and standardizing reports, highlighting the information necessary for clinical specialists, ensuring the automatic replacement of specific terminology, and including the use of patient-friendly language, more understandable vocabulary, or translation of information into other languages [1, 2]. Identifying reports containing certain data to extract can be useful for solving clinical issues [1]. Some studies have proposed ways to identify reports describing the musculoskeletal system with signs of bone fractures, computed tomography (CT) signs of pulmonary embolism, pulmonary nodules, etc. [3, 5, 6].

An algorithm for machine processing of Russian reports must be developed for the use and analysis of large amounts of data to evaluate and describe medical images and prepare conclusions.

The purpose of the study was to evaluate the opportunities and limitations of using text-processing tools to search for various abnormalities in radiology reports.

MATERIALS AND METHODS

Development of a tool for evaluating text in radiology reports

This study was performed as part of a study previously approved by the ethics committee (Extract from Protocol

No. 2 of the Independent Ethics Committee of the Moscow Regional Branch of the Russian Society of Roentgenologists and Radiologists [RSRR] dated February 20, 2020, Clinical trials Registration ID: NCT04489992).

The tool for evaluating text radiology protocols was developed as part of the Moscow experiment on the use of innovative computer vision technologies to analyze medical images and compare the results of assessing medical images for abnormalities by artificial intelligence (AI) services and radiologists.

Mammography, chest radiography and fluorography, CT, and low-dose CT (LDCT) reports were evaluated. All findings were obtained from healthcare facilities of the Department of Health of Moscow in 2020. Anonymized radiology reports were used.

The main purpose was to create an automated algorithm for the automatic analysis of radiography reports for abnormal changes of interest. The target abnormality selection and corresponding glossary development were based on the general requirements for AI data (<https://mosmed.ai/>).

For chest radiography and fluorography, target abnormalities included pleural effusion, pneumothorax, atelectasis, lesion, infiltration/consolidation, dissemination, cavity with degradation or fluid, calcification, and non-integrity of the cortical layer (fracture). For CT and LDCT, target abnormalities included solid and subsolid nodules larger than 100 mm³. For chest CT, another group of abnormalities included changes that correlated with signs of coronavirus disease 2019 (COVID-19). The classification by severity was used according to the interim guidelines “Prevention, Diagnosis, and Treatment of a New Coronavirus Infection (COVID-19)” of the Ministry of Health of the Russian Federation and guidelines of the State Budgetary Healthcare Institution “Scientific and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Department of Health of Moscow” and “Radiology Diagnostics of Coronavirus Disease (COVID-19): Organization, Methodology, Interpretation” [7, 8]. Mammography was performed using the breast imaging reporting and data system (BI-RADS) 3–6 system for analyzing and recording results of breast imaging [9].

Software and hardware solutions for pre-labeling text reports

Software and hardware solutions have been developed for pre-labeling text reports for each type of examination based on NLP methods in combination with the expert opinion of radiologists. Pre-labeling algorithms were developed iteratively with several milestones for each modality.

1. A primary set of key features was defined to search for indications of certain abnormalities. Among others, keywords and phrases, size designations and, if necessary, stop words and phrases were considered. The primary set of keywords compiled by a radiologist included the generally accepted and most frequently used terms by radiologists. Stop words included non-target

abnormalities or non-normal findings, such as changes in other organs in the scan area and anatomical variations.

2. Those feature sets were translated into machine language using the high-level programming language Python.

In this study, Python 3.8 was used with pandas libraries = 1.1.3, numpy = 1.19.2, re = 2.2.1, and nltk = 3.5. The input program data are text reports in tabular formats (.csv, .xlsx) containing both the examination description and conclusion. The results are presented with the initial data labeled "0" or "1," where "0" means the absence of a symptom/abnormality, and "1" means the presence of a symptom/abnormality.

The module for searching for COVID-19 signs using textual chest CT conclusions is based on NLP methods and a classifier from the family of machine-learning algorithms. The output indicates the presence or absence of COVID-19 signs and CT degree of lung damage.

The module for searching for breast cancer signs in mammography reports with conclusion and description detects breast cancer signs according to the BI-RADS classification, giving the BI-RADS class and binary classification (like the screening scale) as a response. The requirements for mammography result description define the mandatory classification of examinations according to the BI-RADS scale. On the contrary, BI-RADS 1 spelling has numerous variants. HCPs may use different cases, punctuations, and, most importantly, layouts. Thus, text parsing is not always effective, leading to false omissions. For this task, the purpose of NLP is limited to extracting information.

The module for searching for lung cancer signs according to text chest CT/LDCT reports (with description and conclusion) identifies cancer signs using a combination of keywords (keys) and parameters (sizes) and is based on NLP.

The module for searching for various abnormal signs in text description and conclusion for chest radiography and fluorography identifies abnormal signs according to the keyword glossary ("sign RG/FLG": "keyword1," "keyword2," "keyword3"...).

3. Report labeling by the pre-labeling program created in Step 2.
4. Selection of unique diverse reports (by abnormalities) using a formatted sample in Step 3. In this study, we had several opportunities to prove that radiologists use template formulations. Training NLP algorithms using template formulations will allow quick retraining of the model. We shall ensure the widest possible implementation of the proposed algorithm to enrich the training data set with different rare (unique) formulations.
5. Manual verification of machine labeling and estimated accuracy of the pre-labeling algorithm. Machine labeling accuracy was evaluated as the percentage of correctly labeled protocols. Manual verification was conducted many times for each task by radiologists.
6. A list of adjustments was labeled, including additional stop words and phrases, keywords, and other

recommendations to improve the quality of the pre-labeling program.

7. Adding verified protocols to the database.

In this step, diverse, unique, and formatted samples were formed while maintaining the class balance. When forming the study samples, the sizes are given below. All examinations were considered for a limited period (January 2019–August 2020). The class balance corresponded to that of the general population. The expected abnormality distribution was as follows: CT for COVID-19, 20% normal, 80% abnormal; mammography, 95% normal, 5% abnormal; fluorography, 95% normal, 5% abnormal; chest radiography, 75% normal, 25% abnormal.

Steps 2–7 were repeated iteratively until the pre-labeling program is 98% accurate. This value was chosen based on the maximum accuracy level of NLP for individual clinical problems found in the analysis of medical literature, which was 97% [10].

In total, during the development of pre-labeling algorithms, the radiologist analyzed 977 reports for mammography, 3196 for radiography, 1608 for fluorography, 4074 for chest CT, and 398 for chest LDCT. The study included all examination reports that were sent to AI services as part of an experiment on the use of innovative computer vision technologies for the analysis of medical images and further implementation in the healthcare system of Moscow (<https://mosmed.ai/>). Only reports with incomplete description and conclusion were excluded.

To improve the quality and speed up the automatic labeling of text protocols, machine-learning methods were used to consider the complex semantic structures of sentences in the reports to search for signs of COVID-19 pneumonia according to CT data. In the future, intelligent algorithms should be developed to search for abnormal signs in mammography, CT (for lung cancer), fluorography, and radiography reports.

The module for processing COVID-19 CT reports was designed for three functions: (1) search for conclusions in input data using an already labeled report database, (2) labeling of the remaining reports using a regular expression, and (3) labeling of the remaining reports using k-nearest neighbor (kNN) models. These functions were implemented sequentially. When developing this tool, one of the conditions was performance optimization. Earlier, we mentioned the frequent use of template expressions in reports prepared by radiologists. Thus, most of the COVID-19 CT reports have the same form. In addition, a team of authors initiated manual labeling of chest CT reports with COVID-19 as the target pathology. This simplifies and speeds up the algorithm through the use of a simple logical comparison function, allowing the protocol to be compared with those available in the database. Some reports that are not included in the database of previously labeled examinations remain; hence, a much slower function of text analysis using regular expressions is launched. If there are protocols for which

the regular expression does not find the target pattern, the machine-learning model based on the kNN is launched.

The architecture developed has an optimal combination of speed and accuracy compared with the use of machine-learning alone. Without machine-learning, it was impossible to cover all the reports. The training sample included 4,074 pre-labeled protocols. Such a number of reports was necessary to ensure the required level of accuracy and was obtained in several iterations of model testing and retraining. For the functioning of this module, the Sklearn library (Scikit-learn) was also imported. The trained algorithm was evaluated on a test set and showed high accuracy (99.6%).

RESULTS

A list of keywords and stop words was developed for the selected modalities and abnormalities, considering some special characteristics of reports. Based on Moscow reports, the best results based on the developed glossary were achieved in the search for signs of COVID-19 in pneumonia chest CT reports with an accuracy of 0.996, sensitivity of 0.998, and specificity of 0.989 (true negative [TN]^{*} = 1115; false positive [FP]^{**} = 6; false negative [FN][#] = 2; true positive [TP]^{##} = 3,837) and breast cancer mammography with an accuracy of 1.0, sensitivity of 1.0, and specificity of 1.0 (TN = 461; FP = 0; FN = 0; TP = 571). When looking for lung cancer signs in chest CT and LDCT, the following parameters were obtained: accuracy, 0.895; sensitivity, 0.829; specificity, 0.936 (TN = 619; FP = 42; FN = 72; TP = 349). When looking for abnormal changes in chest radiography and fluorography, the above-mentioned parameters were 0.912; 1.000, and 0.844, respectively (TN = 259; FP = 48; FN = 0; TP = 237).

^{*}TN: prediction of negative class as a negative class (number), the true value is 1, and prediction is 1.

^{**}FP: prediction of a negative class as a positive class (number), the true value is 1, and prediction is 0.

[#]FN: prediction of a positive class as a negative class (number), the true value is 0, and prediction is 1.

^{##}TP: prediction of a positive class as a positive class (number), the true value is 0, and prediction is 0.

DISCUSSION

Mammography glossary

A mammography glossary was the simplest to compile and use because of state-of-art standardized protocols. Mammography reports are the most structured ones. They require BI-RADS and are subject to control because of the high significance of the detected pathology. In most cases, the BI-RADS category is indicated in the conclusion, and changes to set the category are presented in the description. The presence of BI-RADS categories in the glossary ensures achieving a high accuracy of 1.0 in the automated processing of reports. This tool can be used in other regions of the Russian Federation, if required, for example, for reports of

a different structure. Moreover, adding key defining words ("tumor" and "c-r") is possible.

Tool limitations may be associated with the absence of the BI-RADS category in the report; such protocols often contain information about the impossibility of assessing the condition of the mammary glands because of inadequate image quality or other reasons. Since such examinations are classified into a separate group, their targeted search is possible, for example, for quality control and selection of patients requiring additional examination. The proposed tool, subject to further improvement, can be used for other purposes, for example, assessment of the compliance of the report with the standard and comparison of the description and conclusion for audit purposes.

Glossary for COVID-19 pneumonia findings by severity

High levels of accuracy were also achieved when using the tool to analyze COVID-19 pneumonia findings by severity, which is also related to the structure of reports and unambiguity of the compiled glossary. The glossary used is based on standard grades: RT0 = no evidence of viral pneumonia; RT1–RT4 = mild-to-critical viral pneumonia; and OTHER = other changes not associated with viral pneumonia. During the pandemic, reports contained information about the absence or presence of viral pneumonia signs, degree of spread associated with the severity, and, in most cases, likelihood of viral pneumonia. Provided that certain parameters are met, the text conclusion and glossary based on the degree of spelling variations are sufficient for the severity assessment. In most cases, the descriptive parts of the reports have common features and common terminology.

Despite the common features of most protocols, slight variations are present in the conclusions, which are associated with various construction options, terminology, spelling, punctuation, lexical features of the HCP language, and personal experience with CT. Sometimes, diagnostic inaccuracies were associated with the CT features of viral pneumonia. If the describing physician believes that changes such as frosted glass, reticular striation, etc., may correspond to other diseases, or there is comorbidity, the protocol and conclusion may contain sentences that are atypical in construction and terminology. In these situations, there may be uncertainty in the tool operation, which allows focusing on such examinations, conducting a targeted audit, identifying inaccuracies in the terminology use, and using it for clinical purposes to identify comorbidities that require special attention and need further monitoring. Different classifications and protocols can be used for describing viral pneumonia in different regions and healthcare facilities.

Glossary for looking up lung cancer signs

Developing keywords and stop words to search for lung cancer signs in chest CT and LDCT reports is difficult;

thus, the algorithm was less accurate. LDCT protocols use the Lung-RADS classification, and its use could simplify searching for suspicious nodes as much as possible [11]. However, when analyzing Moscow findings, the search for Lung-RADS categories does not allow the thorough evaluation of available protocols because not all reports contain such a category in the conclusion. In addition, 8.3% of the reports contain discrepancies between the description and the conclusion [12].

The development of a glossary for searching for signs of various lung nodules and neoplasms is still an urgent task and is associated with several issues and limitations. Despite the use of templates and methodological recommendations for the description, chest CT and LDCT reports have quite a variety of options for structures and sequences. Many radiologists do not use the standard recommended terminology (e.g., 4-mm nodules are denoted by the term “mass”), which leads to the misuse of terms [13].

Based on current recommendations and required terminology, keywords were used to search for suspicious changes, corresponding to solid lung nodules/foci and masses >6 mm (lesions >3 cm) [10–12]. These criteria have several limitations, which can lead to FP or FN algorithm results. Thus, when performing a chest CT, randomly found solid lung nodules are proposed to be evaluated using the Fleischner recommendations. However, their use requires assessing personal and clinical information, risk factors, comorbidities, including neoplasms [14].

The use of the size criterion and main set of keywords made it impossible to completely exclude benign changes. For example, to exclude benign nodes with structural calcification, its distribution, which is often not specified in reports, should be considered. Moreover, calcifications can be described in the cancer structure. Large foci can be described by HCPs as part of the description of other diseases such as tuberculosis, sarcoidosis, and bronchiolitis of various etiologies.

To compare protocol data and results of AI processing, the capabilities of current algorithms are adequate. After further improvement ensuring increased accuracy, this tool can be used in other regions, including for developing another useful tool for different tasks.

The modified tool can be used to create a more accurate algorithm considering necessary risk factors, such as the presence of immunodeficiency, inflammatory processes, clinical information, and referral diagnosis. This improvement may be important when evaluating lung nodules in patients with cancer. Cancer information can be obtained from the description of the report and in electronic medical records. It could also be promising to use the tool for estimating changes in nodule size and comparing findings with recommendations for the management of pulmonary nodules to improve tool functions in accordance with improving computer vision models.

For chest CT, some special limitations are notable, which are associated with a wide and difficult-to-cover list of “stop” words. This is related to the examination characteristics: the scan area includes the abdomen, neck, and other chest organs. For most organs (e.g., thyroid, liver, kidneys, and adrenal glands), such “stop” words include names of such organs. However, anatomical chest structures such as the mediastinum, pericardium, ribs and thoracic vertebrae, soft tissues of the chest wall, and diaphragm cannot be used as independent “stop” words because of cases when lung neoplasms have invasive growth and affect adjacent tissues, which is described by radiologists as a summary (“mass extending into the mediastinum”). Moreover, various independent neoplasms of the listed organs and anatomical structures are often revealed, which leads to many FP algorithm results.

Glossary of keywords for chest radiography and fluorography

The development and use of such a tool for chest radiography and fluorography are challenging. Radiography and fluorography reports have many variations in form, structure, size, and characteristics used, while the terminology varies significantly [15, 16].

In addition to the generally accepted terms for abnormalities, the proposed glossary of keywords for chest radiography and fluorography included specific radiological terms such as “darkening,” “focus,” and “shadow.” This leads to several issues because such terms can be used for non-target abnormalities or anatomical structures (“rib shadow”) and additional medical devices (“pacemaker shadow” and “drainage tube shadow”).

To define the pathology according to the binary classification (normal/abnormal) considering the listed issues, the high accuracy of radiology examination is required. However, it is necessary to classify abnormalities (e.g., effusion, pneumothorax, atelectasis, and focus), even if radiologists consider it to be findings difficult to classify by groups, which is also associated with the limitations of radiography. In addition, the same keywords are used to refer to completely different radiological findings.

When developing stop words, different variants of normal spellings were considered using various lexical and syntactic variants of negation (“no shadows the in lungs,” “no abnormal shadow effects in the lungs,” etc.). As new data become available, the glossary shall be constantly updated. Current accuracy indicators for this type of diagnostics allow us to solve the main problem of comparing the results of AI models and HCP findings.

CONCLUSION

With high accuracy achieved, machine-learning methods can be used to automatically classify the texts

of mammography and chest CT reports to search for viral pneumonia signs because of the structured and standardized description of findings.

When searching for lung cancer signs in chest CT and LDCT reports and abnormal changes in chest radiography and fluorography reports, the achieved accuracy is adequate for the successful use of the tool to automatically compare HCP and AI findings in radiology departments. Less accuracy is related to the less strict structure of reports and their diagnostic, lexical, and terminological features.

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

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

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.

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, Vladimirovskiy 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).

tional 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-нCoV. Временные методические рекомендации 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., Гомболевский 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 Thoracic 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. Сеницын В.Е., Комарова М.А., Мершина Е.А. Протокол рентгенологического описания: прошлое, настоящее, будущее // *Вестник рентгенологии и радиологии.* 2014. № 3. С. 35–40.

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

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

* Corresponding author / Автор, ответственный за переписку