



Оригинальная статья

Оценка диагностической точности системы автоматического анализа цифровых рентгенограмм легких при выявлении округлых образований

Гаврилов П.В.¹ • Смольникова У.А.¹

Актуальность. Большинство данных об эффективности систем анализа цифровых рентгенологических изображений предоставлено самими разработчиками и нуждается в качественной проверке на базах данных, подготовленных независимо от разработчика.

Цель – проанализировать информативность автоматического распознавания округлых образований в легких при цифровой рентгенографии с использованием одного из общедоступных диагностических алгоритмов на публично недоступных эталонных наборах данных.

Материал и методы. Исследование основано на распознавании и анализе цифровых рентгенологических изображений из двух публично недоступных эталонных наборов данных, имеющих государственную регистрацию (Российская Федерация), посредством одного из общедоступных диагностических алгоритмов FutureMed Analyzer. Работа выполнена на примере двух моделей рентгенологического скрининга: модель 1 состояла из 100 рентгенограмм легких с соотношением «норма:патология» 94%:6%, модель 2 состояла из 5150 рентгенограмм легких с соотношением «норма:патология» 97%:3%.

Результаты. Анализ результатов интерпретации рентгенограмм диагностической системой

показал: из рентгенограмм модели 1 было верно интерпретировано 98% случаев, модели 2 – 95%. При этом 83% случаев из модели 1 и 69% из модели 2 были интерпретированы как рентгенограммы с наличием патологических изменений в легких. Количество правильных ответов при разделении рентгенограмм легких на две категории – «норма» и «патология» – в отношении моделей 1 и 2 составило 95 и 98% соответственно. Чувствительность в выявлении патологических образований колебалась от 69 до 83%. Специфичность составила 99% для рентгенограмм из модели 1 и 96% для рентгенограмм из модели 2. Был получен довольно низкий показатель гиподиагностики: для модели 1 – 17%, для модели 2 – 31%. Параметр «площадь под кривой» для модели 1 был равен 0,91, для модели 2 – 0,85.

Заключение. Диагностическая эффективность автоматического анализа изображений на основе сверточных нейронных сетей приближается к аналогичным показателям врачей-рентгенологов. Эта система автоматического выявления патологических изменений не смогла решить наиболее сложные проблемы выявления округлых образований с низкими плотностными характеристиками (согласно данным компьютерной томографии – по типу «матового стекла»)

и так называемую проблему суммации теней при локализации патологических изменений в таких затруднительных для интерпретации местах, как верхушки легких, ключицы, ребра и др. Для выбора подходящей системы медицинским учреждениям необходимо выполнять предварительное тестирование на собственных моделях, эквивалентных исследованиям, которые проводятся в данном учреждении (параметры выполнения рентгенографии, характер и частота выявляемой патологии).

Ключевые слова: рентгенография легких, округлые образования, скрининг, искусственный интеллект, сверточные нейронные сети

Для цитирования: Гаврилов П.В., Смольникова У.А. Оценка диагностической точности системы автоматического анализа цифровых рентгенограмм легких при выявлении округлых образований. Альманах клинической медицины. 2021;49(6):359–364. doi: 10.18786/2072-0505-2021-49-035.

Поступила 11.05.2021; доработана 25.06.2021; принята к публикации 28.06.2021; опубликована онлайн 15.07.2021

Цифровые рентгенологические исследования легких занимают значимую (более 30%) долю от всех лучевых исследований в Российской Федерации [1, 2]. По данным многих авторов, существенную проблему в выявлении патологических изменений на рентгенограммах органов грудной клетки представляют сложность трактовки суммационных изображений и высокие требования к квалификации врачей-рентгенологов [3–5]. Эти факторы объясняют большой интерес разработчиков программного обеспечения к созданию диагностических алгоритмов, позволяющих проводить автоматический анализ рентгенологических

изображений легких. Применение таких алгоритмов предполагает более эффективное обнаружение патологических изменений [6–8].

В настоящее время большинство данных об эффективности систем анализа цифровых рентгенологических изображений предоставлено самими разработчиками и нуждается в качественной проверке на базах данных, подготовленных независимо от разработчика. При этом следует учитывать, что разработанные базы должны быть эквивалентны тем исследованиям, в которых будет применяться конкретная система распознавания и анализа рентгенограмм в реальной практике.



Целью исследования было проанализировать информативность автоматического распознавания округлых образований в легких при цифровой рентгенографии с использованием одного из общедоступных диагностических алгоритмов на публично недоступных эталонных наборах данных.

Материал и методы

Настоящее исследование основано на результатах распознавания и анализа цифровых рентгенограмм легких из тестовых баз посредством автоматической системы анализа медицинских изображений, ядром которой служит ансамбль из 10 нейронных сетей, созданных на основе архитектуры DenseNet-121 (свидетельство о регистрации программы для ЭВМ RU-2019665266, торговое название – FutureMed Analyzer). Она позиционируется как программное обеспечение в виде системы поддержки принятия врачебных решений, предоставляющее врачу-рентгенологу возможность получить «второе мнение» [9].

Для тестирования были сформированы две модели рентгенологического скрининга на основе зарегистрированных баз данных:

- «Базы цифровых рентгенограмм легких в передней проекции без патологических изменений» (свидетельство о регистрации RU-2019622406) – 5000 рентгенограмм;
- «Результаты лучевых исследований пациентов с периферическими образованиями легких» (свидетельство о регистрации RU-2019621712) – 150 рентгенограмм.

Для формирования «Базы цифровых рентгенограмм легких в передней проекции без патологических изменений» исполняли рентгенограммы легких, выполненные пациентам без жалоб с целью скрининга туберкулеза. Критерием отбора было согласованное мнение 5 врачей-рентгенологов об отсутствии патологических изменений на рентгенограмме. Критерием исключения из базы служило подозрение хотя бы одного рентгенолога на наличие патологических образований в легких.

База «Результаты лучевых исследований пациентов с периферическими образованиями легких» составлена на основе структурированных деперсонализированных данных лучевых исследований (рентгенограмм в передней проекции и компьютерных томограмм) 150 пациентов с различными верифицированными периферическими образованиями в легких. Структура патологических изменений, представленных в базе, была следующей: туберкулез легких – 33% (n = 50),

Гаврилов Павел Владимирович – канд. мед. наук, вед. науч. сотр., руководитель направления «Лучевая диагностика»; ORCID: <https://orcid.org/0000-0003-3251-4084>
✉ 191036, г. Санкт-Петербург, Лиговский пр-т, 2–4, Российская Федерация.
Тел.: +7 (812) 775 75 55.
E-mail: spbniifrentgen@mail.ru

Смольникова Ульяна Алексеевна – аспирант отделения лучевой диагностики¹; ORCID: <https://orcid.org/0000-0001-9568-3577>.
E-mail: ulammonika@mail.ru

рак легких – 49% (n=74), доброкачественные образования легких – 14% (n=20), другое – 3% (n=6).

Модель 1 состояла из 100 рентгенограмм легких с соотношением «норма:патология» 94%:6% (94 человека без значимой рентгенологической патологии и 6 человек с подтвержденным наличием синдрома округлого образования в легочной ткани).

Для более детальной оценки возможности системы машинного обучения и анализа цифровых рентгенологических изображений в качестве программного продукта, направленного на обнаружение округлых образований в легких на цифровых рентгенограммах, и для определения совокупности параметров их диагностической точности была сформирована модель 2 с использованием всего объема информации из ранее указанных баз данных (5150 цифровых рентгенограмм). Соотношение «норма:патология» в модели 2 составило 97%:3%.

Для сопоставления эффективности автоматического распознавания образований в легких при цифровой рентгенографии относительно референс-теста (баз данных) применялся классический набор показателей, характеризующих диагностическую результативность (чувствительность, специфичность, отношение правдоподобия положительного результата, отношение правдоподобия отрицательного результата, прогностическая ценность положительного результата, прогностическая ценность отрицательного результата, точность) [10, 11].

Результаты

Согласно полученным результатам интерпретации цифровых рентгенограмм посредством использования системы машинного обучения и анализа цифровых рентгенологических изображений были определены основные критерии информативности диагностического теста, сформированы и проанализированы четырехпольные таблицы сопряженности (табл. 1, 2).

По результатам анализа, диагностической системой из модели 1 было верно интерпретировано 98% рентгенограмм, из модели 2 – 95%.

При анализе результатов интерпретации рентгенограмм с наличием округлого образования в легком из моделей 1 и 2 были интерпретированы как рентгенограммы с наличием патологических изменений в легких 83 и 69% случаев соответственно, при этом лишь на 33 и 55% снимков была верно указана локализация патологии на получаемой тепловой карте.

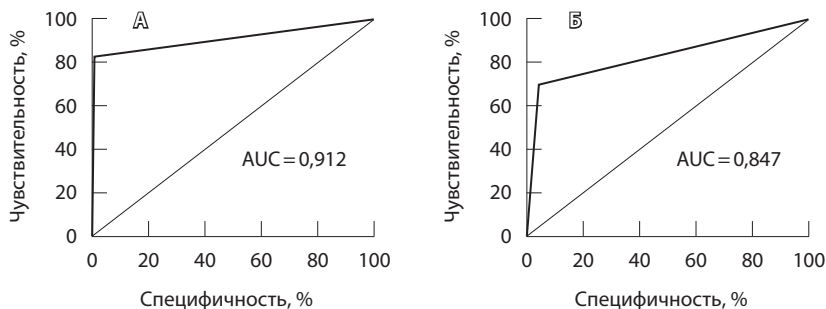
¹ ФГБУ «Санкт-Петербургский научно-исследовательский институт физиопульмонологии» Минздрава России; 191036, г. Санкт-Петербург, Лиговский пр-т, 2–4, Российская Федерация



Таблица 1. Результаты интерпретации цифровых рентгенограмм системой автоматического анализа

Показатель	Классификатор			
	определено как патология		определено как норма	
	TP	FP	FN	TN
Модель 1 (n = 100), n (%)	5 (83)	1 (1)	1 (17)	93 (99)
Модель 2 (n = 5150), n (%)	104 (69)	219 (4)	46 (31)	4781 (96)

FN – false negative (классификатор неверно утверждает, что объект не принадлежит к рассматриваемому классу), FP – false positive (классификатор неверно отнес объект к рассматриваемому классу), TN – true negative (классификатор верно утверждает, что объект не принадлежит к рассматриваемому классу), TP – true positive (классификатор верно отнес объект к рассматриваемому классу)



Характеристические кривые: **А** – для модели 1, **Б** – для модели 2; AUC – area under the curve (площадь под кривой)

Количество правильных ответов при разделении рентгенограмм легких из моделей 1 и 2 на норму и патологию составило 95 и 98 соответственно. Чувствительность в выявлении патологических образований колебалась от 69 до 83%. Специфичность составила 99% для рентгенограмм из модели 1 и 96% для рентгенограмм из модели 2, что коррелирует с показателем прогностической ценности отрицательного результата (99%).

Дополнительно нами были построены графики, позволяющие оценить качество бинарной классификации, – характеристические кривые (ROC-кривые). Параметр «площадь под кривой» для модели 1 составил 0,91, для модели 2 – 0,85

(рисунок), что служит признаком хорошего качества моделей [10, 12, 13].

Отношение правдоподобия положительно-го результата по итогам анализа рентгенограмм из модели 1 составило 78,3, из модели 2 – 15,8. Прогностическая ценность положительного результата была больше при интерпретации рентгенограмм из выборки 1 – 83%. В целом получен довольно низкий показатель гиподиагностики: для модели 1 – 17%, для модели 2 – 31%. Все это можно рассматривать как положительный результат в отношении использования подобных программных продуктов на выборке с заранее известным преобладающим числом рентгенограмм без патологических изменений.

Мы также проанализировали влияние рентгенологических характеристик округлых образований в легких на качество их выявления системой машинного обучения и анализа цифровых рентгенологических изображений из модели 2. В 31% случаев снимки были интерпретированы как рентгенограммы без патологических изменений, при этом на 19% рентгенограмм была неверно указана локализация патологических изменений. Среди случаев гиподиагностики 46% рентгенограмм были с верифицированным раком легкого.

Дополнительно изучена взаимосвязь между типом округлого образования, его размерами и частотой пропуска патологических изменений при использовании автоматической системы анализа медицинских изображений. Программа не выявила ни одного образования по типу «матового стекла». Среди рентгенограмм с наличием образования солидного типа количество ошибочных интерпретаций составило 29% (n = 38), среди рентгенограмм с образованиями субсолидного типа – 38% (n = 6).

Количество случаев гипердиагностики при анализе рентгенограмм с образованиями размерами до 10 мм составило 80% (n = 4), размерами от 10 до 30 мм – 33% (n = 32), размерами более 30 мм – 21% (n = 10).

Оказалось, что наибольшие сложности с выявлением патологии возникали в случае локализации изменений в С1, С2 правого легкого

Таблица 2. Результативность выявления патологических изменений в легких

Показатель	Чувствительность, %	Специфичность, %	Отношение правдоподобия положительного результата	Отношение правдоподобия отрицательного результата	Прогностическая ценность положительного результата, %	Прогностическая ценность отрицательного результата, %	Точность, %
Модель 1 (n = 100)	83	99	78,33	0,17	83	99	98
Модель 2 (n = 5150)	69	96	15,83	0,32	32	99	95



(31 и 28% соответственно) и в С1+2 левого легкого (53%). Это может быть ассоциировано с локализацией патологических изменений за тенью ребер и в верхушках легких в результате суммации тени округлого образования и тени 1-го ребра или ключиц на рентгенограмме.

Обсуждение

Полученные показатели критериев диагностической эффективности интерпретации цифровых рентгенограмм легких с синдромом округлого образования в легочной ткани свидетельствуют о перспективности данного метода интерпретации цифровых рентгенологических изображений, а также о возможности допуска программного продукта к клинической валидации [10].

При сравнении результатов нашего исследования с данными других работ отмечено, что показатели чувствительности, специфичности, отношения правдоподобия положительного результата, отношения правдоподобия отрицательного результата, прогностической ценности положительного результата, прогностической ценности отрицательного результата находятся в одном диапазоне, но колеблются в пределах 5–10%, что может быть обусловлено как качеством обучения программных продуктов, так и разностью выборок, примененных для тестирования. Это требует дальнейшего исследования на большем количестве выборок [6, 8, 14].

Если сравнивать ресурсы данного продукта с возможностями врача-рентгенолога в выявлении аналогичной патологии, можно говорить о сопоставимых результатах. В нашей предыдущей работе мы провели тестирование врачей-рентгенологов на возможность выявления округлых образований в легких при цифровой рентгенографии. Были получены следующие показатели диагностической эффективности: чувствительность – 75,35%, специфичность – 72,28%, отношение правдоподобия положительного результата – 2,71, отношение правдоподобия отрицательного результата – 0,34, прогностическая ценность положительного результата – 54,02%, прогностическая ценность отрицательного результата – 87,15%, точность – 73,2% [5]. Схожие данные опубликованы Y. Sim и соавт.: результаты тестирования системы машинного обучения и анализа цифровых рентгенологических изображений были аналогичны таковым, полученным при тестировании врачей-рентгенологов, при этом на этапе повторного анализа рентгенограмм врачом-рентгенологом уже с использованием результатов анализа рентгенограмм программным

продуктом средняя чувствительность возможностей рентгенологов повысилась с 65,1 до 70,3%, в свою очередь чувствительность программы автоматического анализа в среднем составила 67,3% (от 56,1 до 82,7%) [15].

Вместе с тем подчеркнем, что при использовании модели 2 с большим объемом рентгенограмм в исследовании и уменьшением соотношения между нормой и патологией мы видим снижение чувствительности программного продукта, прогностической ценности положительного результата, отношения правдоподобия положительного результата и площади ROC-кривой. При этом специфичность, прогностическая ценность отрицательного результата и точность снижаются крайне незначительно (в пределах 2–3%). Все это может быть обусловлено как большей вариативностью различных патологических процессов, предложенных системе, так и размерами выборки наблюдений и соотношением между нормой и патологией.

Отметим: данная система автоматического обнаружения патологических изменений не смогла решить наиболее сложные проблемы выявления при рентгенографии округлых образований, встречающихся в клинической практике (образования с низкими плотностными характеристиками, по данным компьютерной томографии относящиеся к типу «матового стекла»), и при локализации патологических изменений за тенью ребер и в верхушках легких – так называемую проблему суммации теней.

Заключение

Диагностическая эффективность исследуемой системы автоматического анализа изображений на основе сверточных нейронных сетей приближается к аналогичным показателям врачей-рентгенологов.

В случае принятия решения о целесообразности рассмотрения системы в качестве вспомогательного второго мнения для врачей-рентгенологов медицинским учреждениям для выбора подходящей системы необходимо проводить предварительное тестирование на собственных моделях, эквивалентных исследованиям, которые проводятся в данном учреждении (параметры выполнения рентгенографии, характер и частота выявляемой патологии).

Медицинскому сообществу предстоит широкая клиническая апробация разработанных систем искусственного интеллекта на независимых наборах данных, и эти результаты могут изменить показатели диагностической точности данных алгоритмов. ©



Дополнительная информация

Финансирование

Исследование выполнено в рамках диссертационной работы У.А. Смольниковой, финансирование осуществлялось ФГБУ «СПб НИИФ» Минздрава России за счет средств, выделяемых на проведение научно-исследовательских работ.

Конфликт интересов

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

Участие авторов

П.В. Гаврилов – концепция и дизайн исследования, анализ и интерпретация результатов исследования, редактирование текста, утверждение итогового варианта текста рукописи; У.А. Смольникова – анализ и интерпретация результатов исследования, написание и редактирование текста, статистическая обработка данных. Оба автора прочли и одобрили финальную версию статьи перед публикацией, согласны нести ответственность за все аспекты работы и гарантируют, что ими надлежащим образом были рассмотрены и решены вопросы, связанные с точностью и добросовестностью всех частей работы.

Литература

1. Тюрин ИЕ. Лучевая диагностика в Российской Федерации в 2016 г. Вестник рентгенологии и радиологии. 2017;98(4):219–226. doi: 10.20862/0042-4676-2017-98-4-219-226.
2. Трофимова ТН, Козлова ОВ. Лучевая диагностика 2018 в цифрах и фактах. Лучевая диагностика и терапия. 2019;(3):100–102. doi: 10.22328/2079-5343-2019-10-3-100-102.
3. Yerushalmy J, Harkness JT, Cope JH, Kennedy BR. The role of dual reading in mass radiography. American Review of Tuberculosis. 1950;61:443–464.
4. Nakamura K, Ohmi A, Kurihara T, Suzuki S, Tadera M. [Studies on the diagnostic value of 70 mm radiophotograms by mirror camera and the reading ability of physicians]. Kekkaku. 1970;45(4):121–128. Japanese.
5. Гаврилов ПВ, Ушков АД, Смольникова УА. Выявление округлых образований в легких при цифровой рентгенографии: роль опыта работы врача-рентгенолога. Медицинский альянс. 2019;(2):51–56.
6. Lakhani P, Sundaram B. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. Radiology. 2017;284(2):574–582. doi: 10.1148/radiol.2017162326.
7. Jaeger S, Karargyris A, Candemir S, Folio L, Siegelman J, Callaghan F, Zhiyun Xue, Palaniappan K, Singh RK, Antani S, Thoma G, Yi-Xiang Wang, Pu-Xuan Lu, McDonald CJ. Automatic tuberculosis screening using chest radiographs. IEEE Trans Med Imaging. 2014;33(2):233–245. doi: 10.1109/TMI.2013.2284099.
8. Морозов СП, Владзимирский АВ, Ледихова НВ, Соколова ИА, Кульберг НС, Гомболевский ВА. Оценка диагностической точности системы скрининга туберкулеза легких на основе искусственного интеллекта. Туберкулез и болезни легких. 2018;96(8):42–49. doi: 10.21292/2075-1230-2018-96-8-42-49.
9. Падалко МА, Наумов АМ, Назариков СИ, Лушников АА. Применение технологий искусственного интеллекта для диагностики туберкулеза и онкологических заболеваний. Туберкулез и болезни легких. 2019;97(11):62. doi: 10.21292/2075-1230-2019-97-11-62-62.
10. Морозов СП, Владзимирский АВ, Кляшторный ВГ, Андрейченко АЕ, Кульберг НС, Гомболевский ВА, Сергунова КА. Клинические испытания программного обеспечения на основе интеллектуальных технологий (лучевая диагностика): методические рекомендации. Серия «Лучшие практики лучевой и инструментальной диагностики». Вып. 57. М.; 2019. 53 с.
11. Васильев АЮ, Малый АЮ, Серова НС. Анализ данных лучевых методов исследования на основе принципов доказательной медицины: учебное пособие. М.: ГЭОТАР-Медиа; 2008. 32 с.
12. Fawcett T. An introduction to ROC analysis. Pattern recognition letters. 2006;27(8):861–874. doi: 10.1016/j.patrec.2005.10.010.
13. Macskassy SA, Provost F, Rosset S. ROC confidence bands: An empirical evaluation. In: ICML '05: Proceedings of the 22nd international conference on Machine learning. 2005. p. 537–544. doi: 10.1145/1102351.1102419.
14. Nam JG, Park S, Hwang EJ, Lee JH, Jin KN, Lim KY, Vu TH, Sohn JH, Hwang S, Goo JM, Park CM. Development and Validation of Deep Learning-based Automatic Detection Algorithm for Malignant Pulmonary Nodules on Chest Radiographs. Radiology. 2019;290(1):218–228. doi: 10.1148/radiol.2018180237.
15. Sim Y, Chung MJ, Kotter E, Yune S, Kim M, Do S, Han K, Kim H, Yang S, Lee DJ, Choi BW. Deep Convolutional Neural Network-based Software Improves Radiologist Detection of Malignant Lung Nodules on Chest Radiographs. Radiology. 2020;294(1):199–209. doi: 10.1148/radiol.2019182465.
6. Lakhani P, Sundaram B. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. Radiology. 2017;284(2):574–582. doi: 10.1148/radiol.2017162326.
7. Jaeger S, Karargyris A, Candemir S, Folio L, Siegelman J, Callaghan F, Zhiyun Xue, Palaniappan K, Singh RK, Antani S, Thoma G, Yi-Xiang Wang, Pu-Xuan Lu, McDonald CJ. Automatic tuberculosis screening using chest radiographs. IEEE Trans Med Imaging. 2014;33(2):233–245. doi: 10.1109/TMI.2013.2284099.
8. Morozov SP, Vladzimirskiy AV, Ledikhova NV, Sokolina IA, Kulberg NS, Gombolevskiy VA. [Evaluation of diagnostic accuracy of the system for pulmonary tuberculosis screening based on artificial neural networks]. Tuberculosis and Lung Diseases. 2018;96(8):42–49. Russian. doi: 10.21292/2075-1230-2018-96-8-42-49.
9. Padalko MA, Naumov AM, Nazarikov SI, Lushnikov AA. [Using artificial intellect for diagnosis of tuberculosis and oncologic diseases]. Tuberculosis and Lung Diseases. 2019;97(11):62. Russian. doi: 10.21292/2075-1230-2019-97-11-62-62.
10. Morozov SP, Vladzimirskiy AV, Klyashstorny V, Andreychenko AE, Kul'berg NS, Gombolevskiy VA, Sergunova KA. [Clinical Trials of Intelligent Software (radiation diagnostics): guidelines]. Series "Best Practices in Radiation and Instrumental Diagnostics". Issue 57. Moscow; 2019. 53 p. Russian.
11. Vasil'ev AY, Malyy AY, Serova NS. [Analysis of data from radiation research methods based on the principles of evidence-based medicine]. Moscow: GEOTAR-Media; 2008. 32 p. Russian.
12. Fawcett T. An introduction to ROC analysis. Pattern recognition letters. 2006;27(8):861–874. doi: 10.1016/j.patrec.2005.10.010.



13. Macskassy SA, Provost F, Rosset S. ROC confidence bands: An empirical evaluation. In: ICML '05: Proceedings of the 22nd international conference on Machine learning. 2005. p. 537–544. doi: 10.1145/1102351.1102419.
14. Nam JG, Park S, Hwang EJ, Lee JH, Jin KN, Lim KY, Vu TH, Sohn JH, Hwang S, Goo JM,

Park CM. Development and Validation of Deep Learning-based Automatic Detection Algorithm for Malignant Pulmonary Nodules on Chest Radiographs. *Radiology*. 2019;290(1): 218–228. doi: 10.1148/radiol.2018180237.

15. Sim Y, Chung MJ, Kotter E, Yune S, Kim M, Do S, Han K, Kim H, Yang S, Lee DJ, Choi BW.

Deep Convolutional Neural Network-based Software Improves Radiologist Detection of Malignant Lung Nodules on Chest Radiographs. *Radiology*. 2020;294(1):199–209. doi: 10.1148/radiol.2019182465.

Evaluation of diagnostic accuracy of the automatic system for the analysis of digital lung X-ray for detection of spherical masses

P.V. Gavrilov¹ • U.A. Smolnikova¹

Rationale: Most data on the effectiveness of systems for the analysis of digital X-ray images have been provided by their developers and require a high-quality validation in databases prepared independently of the developer.

Aim: To analyze the information content of automatic identification of spherical lung masses with digital X-ray imaging using one of the widely available diagnostic algorithms on publicly inaccessible reference datasets.

Materials and methods: The study was based on the recognition and analysis of digital X-ray images from two publicly inaccessible reference datasets that have the state registration (Russian Federation) with one of the publicly available diagnostic algorithms (FutureMed Analyzer). The study was performed using two models of X-ray screening as examples: Model 1 consisted of 100 X-ray images of the lungs with a normal: abnormal ratio of 94%: 6%; Model 2 consisted of 5150 chest X-ray images with a normal: abnormal ratio of 97%: 3%.

Results: According to the results of the analysis of the X-ray images with the diagnostic system, 98% of the images were correctly interpreted with Model 1 and 95% of the images, with Model 2. 83% of the cases from Model 1 and 69% from Model 2 were interpreted as images with lung abnormalities. The percentage of correct answers for differentiation of the chest X-ray images into two categories (normal vs. abnormal) for Model 1 and Model 2 was 95% and 98%, respectively. The sensitivity for detection of abnormal masses ranged from 69% to 83%. The specificity was 99%

for the Model 1 chest X-ray images and 96% for the Model 2 chest X-ray images. The underdiagnosis rate was quite low ranging for Model 1 – 17%, and for Model 2 – 31%. The area under the curve for Model 1 was 0.91 and for Model 2 0.85.

Conclusion: The diagnostic efficiency of the automatic image analysis based on the convolutional neuronal networks approaches that of the radiologists. This system of automatic identification of abnormalities was unable to solve the most complex problems of detecting low density spherical masses (like "ground glass" area on computed tomography) and that of shadow summation for abnormalities located in such difficult to interpret zones as lung apices, clavicles, ribs, etc. To select a suitable system, medical institutions need to conduct preliminary testing in their own models equivalent to the studies performed in a given institution (parameters for radiography, nature and frequency of abnormalities).

Key words: lung X-ray, lung mass, screening, artificial intelligence, convolutional neuronal networks

For citation: Gavrilov PV, Smolnikova UA. Evaluation of diagnostic accuracy of the automatic system for the analysis of digital lung X-ray for detection of spherical masses. *Almanac of Clinical Medicine*. 2021;49(6):359–364. doi: 10.18786/2072-0505-2021-49-035.

Received 11 May 2021; revised 25 June 2021; accepted 28 June 2021; published online 15 July 2021

Pavel V. Gavrilov – MD, PhD, Leading Research Fellow, Head of Radiology Area¹; ORCID: <https://orcid.org/0000-0003-3251-4084>
✉ 2–4 Ligovskiy prospekt, Saint Petersburg, 191036, Russian Federation. Tel.: +7 (812) 775 75 55. E-mail: spbniifrentgen@mail.ru

Uliana A. Smolnikova – Postgraduate Student, Department of Radiology¹; ORCID: <https://orcid.org/0000-0001-9568-3577>. E-mail: ulamonika@mail.ru

Funding

The study was performed as a part of the thesis by U.A. Smolnikova, financed by the Saint Petersburg Research Institute for Phthisiopulmonology from the research budget.

Conflict of interests

The authors declare that there is no conflict of interests regarding the publication of this article.

Authors' contributions

P.V. Gavrilov, the study concept and design, analysis and interpretation of the results, text editing, approval of the final version of the manuscript; U.A. Smolnikova, analysis and interpretation of the results, text writing and editing, statistical analysis. Both authors have read and approved the final version of the manuscript before submission, agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work have been appropriately investigated and resolved.

¹Saint Petersburg Research Institute for Phthisiopulmonology; 2–4 Ligovskiy prospekt, Saint Petersburg, 191036, Russian Federation