

*Я.А. Пчелинцев¹, А.В. Хвостиков², А.С. Крылов³,
Л.Е. Паролина⁴, Н.А. Никифорова⁵, Л.П. Шепелева⁶,
Е.С. Прокопьев⁷, М. Фариас⁸, Д. Йонг⁹*

АНАЛИЗ ЖЕСТКОСТИ РЕНТГЕНОВСКИХ СНИМКОВ ДЛЯ НЕЙРОСЕТЕВОЙ ДИАГНОСТИКИ ТУБЕРКУЛЕЗА*

Введение

Одной из актуальных задач при применении методов глубокого обучения в медицинской диагностике является анализ и предобработка входных данных. Необходим контроль соответствия входной информации и применяемого обученного (обучаемого) метода глубокого обучения. Это, например, требуется при контроле наличия адверсативных атак на данные [1].

При изучении рентгеновских снимков и, в частности, при диагностике туберкулёза важным фактором является жёсткость снимка, так как она напрямую влияет на его информативность [2, 3]. При условии правильного контрастирования снимка уровень жёсткости можно определить визуально, подсчитав число отчётливо видимых на снимке верхних грудных позвонков: оптимальному уровню соответствуют 3–4 видимых позвонка, меньшее или большее число свидетельствует о том, что снимок слишком мягкий или жёсткий [3, 4]. Примеры снимков разного уровня жёсткости представлены на Рис. 1.

Контроль качества рентгеновских снимков грудной клетки на основании разных параметров важен для полноценного анализа снимка и постановки правильного диагноза. Автоматическое определение контроля пространственных условий снимка (позы пациента, положения грудной клетки в кадре и т.п.) рассматривается в работах [5, 6, 7]. Садре Р.

¹ Факультет ВМК МГУ имени М.В.Ломоносова, e-mail: yapchelintsev@live.cs.msu.ru

² Факультет ВМК МГУ имени М.В.Ломоносова, e-mail: khvostikov@cs.msu.ru

³ Факультет ВМК МГУ имени М.В.Ломоносова, e-mail: kryl@cs.msu.ru

⁴ «НМИЦ фтизиопульмонологии и инфекционных заболеваний» Минздрава России, e-mail: parolina@nmrc.ru

⁵ ГБУ РС(Я) НПЦ "Фтизиатрия" имени Е.Н.Андреева, e-mail: nikiada@mail.ru

⁶ ГБУ РС(Я) НПЦ "Фтизиатрия" имени Е.Н.Андреева, e-mail: shepelevalp@mail.ru

⁷ ГБУ РС(Я) НПЦ "Фтизиатрия" имени Е.Н.Андреева, e-mail: Prokopeves@ftiz14.ru

⁸ Университет Бразилиа, Бразилия, e-mail: mylene@ene.unb.br

⁹ Чжэцзянский университет, Ханчжоу, Китай e-mail: dingyong09@zju.edu.cn

* Работа выполнена при поддержке гранта РФФИ 19-5780014 БРИКС_т (BRICS2019-394).

с соавторами в своих исследованиях изучают влияние качества снимка на результаты автоматической диагностики COVID-19 [8].



а) мягкий

б) нормальный

в) жёсткий

Рис. 1. Примеры рентгеновских снимков разной жёсткости

В нашей работе контроль входных рентгеновских снимков используется для проверки соответствия уровню облучения, требуемому для качественной диагностики туберкулеза легких.

Данная работа является развитием [9], где было показано, что оптимизация алгоритма диагностики заболеваний по рентгеновскому снимку грудной клетки для работы с изображениями близкой жёсткости вместе с автоматическим контролем качества рентгеновских изображений, позволяет достичь точности классификации выше, чем у алгоритма, созданного в расчёте на обработку разнородных по уровню жёсткости изображений.

В данной работе рассматриваются: 1) задача автоматического определения уровня жёсткости рентгеновского снимка грудной клетки с помощью нейросетевого алгоритма; 2) влияние предварительной фильтрации обучающей и валидационной выборок на качество работы алгоритма классификации в задаче диагностики туберкулёза по рентгеновским снимкам грудной клетки.

Используемые наборы данных

Для создания и тестирования алгоритма определения жёсткости рентгеновских снимков использовался набор из 1298 рентгеновских снимков больных туберкулёзом пациентов, собранный в нескольких медицинских учреждениях Республики Саха (Якутия). Примеры снимков этого набора представлены на Рис. 2. Набор был размечен врачом-рентгенологом: для каждого изображения имеется число отчётливо видимых на этом снимке верхних грудных позвонков. Распределение изображений по такому показателю жёсткости представлено на Рис. 3. Далее этот набор будем именовать «SakhaTB».

Так как в данной работе этапу обучения модели в задаче диагностики туберкулёза предшествует этап фильтрации снимков на основании уровня жёсткости, то, подобно предыдущей работе [9], в целях поддержания размера обучающей, валидационной и тестовой выборки на достаточном уровне для задачи диагностики использовался набор данных из двух частей, описание которых следует ниже.



Рис. 2. Примеры изображений из набора SakhaTB

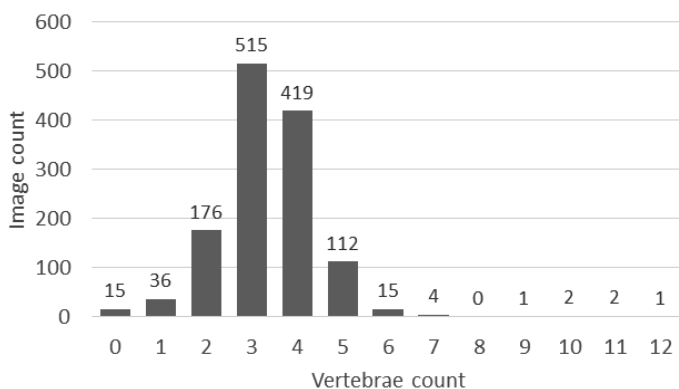


Рис. 3. Гистограмма распределения снимков набора SakhaTB по числу отчётливо видимых позвонков

В качестве первой части были выбраны наиболее часто используемые в задачах обработки рентгеновских снимков больных туберкулёзом [10] открытые наборы данных Montgomery County и Shenzhen [11, 12]. Оба предоставляются Национальной библиотекой медицины Национальных институтов здравоохранения США и собраны Министерством здравоохранения и социальных служб США в округе Монтгомери штата Мэриленд и Медицинским колледжем провинции Гуандун в Народном госпитале № 3 г. Шэньчжэнь в КНР соответственно. Они содержат рентгеновские снимки грудной клетки в оттенках серого с глубиной цвета 8 бит и их метки соответствия 2 классам: здоровые (далее – NORMAL) и больные туберкулёзом (далее – TB) пациенты. Размеры снимков разнятся и примерно равны 3000x3000 и 4000x4900 пикселей. Количество снимков каждого класса представлено в Табл. 1. Так как эти два набора данных в большинстве работ используются совместно [10], то было принято решение так же объединить их, получившийся набор данных далее будет обозначаться «Montgomery-Shenzhen» или «MC-SZ». Примеры снимков из него представлены на Рис. 4.

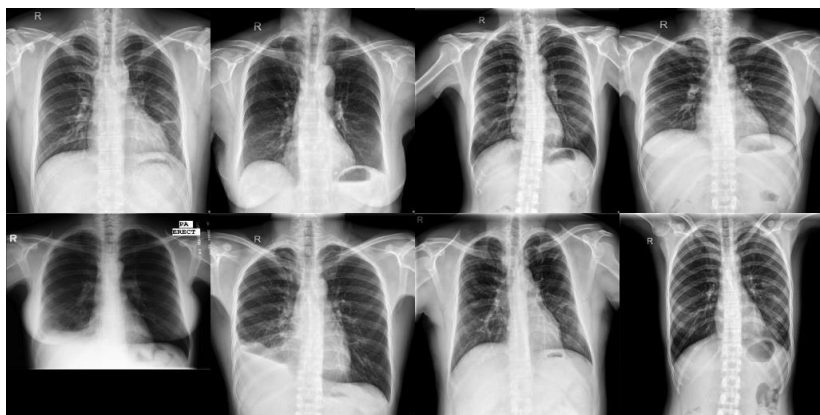
Вторая часть используемых для решения задачи диагностики туберкулёза данных представляет собой подмножество набора TBX11K [13], подготовленного в Нанькайском университете г. Тяньцзинь в КНР. TBX11K содержит 11200 рентгеновских снимков грудной клетки в оттенках серого с глубиной цвета 8 бит размера 512x512 пикселей. Из них для 8400 изображений доступны метки принадлежности к одному из 3 классов (здоровые, больные туберкулёзом и больные иными заболеваниями пациенты) и границы поражённых областей лёгких. Количество снимков, имеющих метки принадлежности к рассматриваемым в данной работе классам, представлено в Табл. 1, все они были включены в итоговую выборку. Примеры снимков представлены на Рис. 5.

Табл. 1. Размеры использованных наборов данных

Название набора	Число снимков класса NORMAL	Число снимков класса TB	Общее число снимков
Montgomery	80	58	138
Shenzhen	326	336	662
TBX11K	3800	800	4600
Всего	4206	1194	5400



а) здоровые пациенты (класс NORMAL)

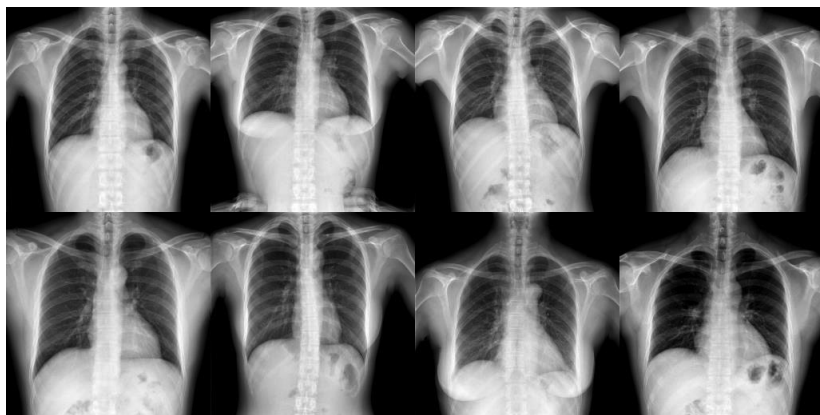


б) пациенты, больные туберкулёзом (класс TB)

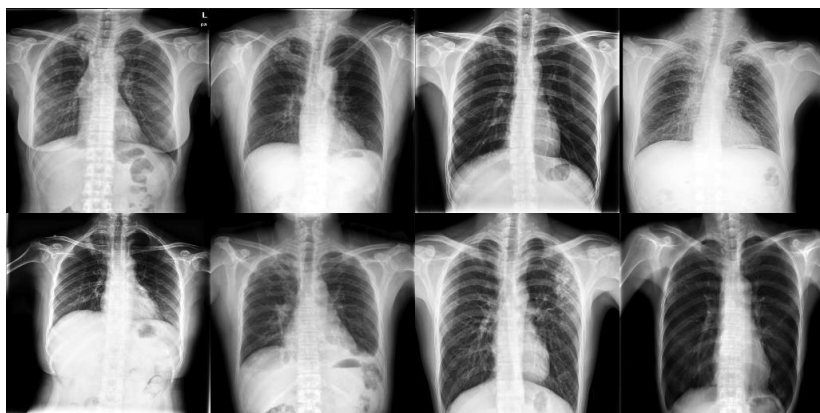
Рис. 4. Примеры изображений из набора Montgomery-Shenzhen (MC-SZ), соответствующие разным классам

Метод определения жёсткости снимков

Так как уровень жёсткости рентгеновского снимка — это упорядоченная величина, для сохранения отношений упорядоченности между классами задача его определения рассматривалась как задача порядковой регрессии (также иногда называется порядковой классификацией).



а) здоровые пациенты (класс NORMAL)



б) пациенты, больные туберкулёзом (класс TB)

Рис. 5. Примеры изображений из набора TBX11K, соответствующие разным классам

Как видно из Рис. 3, набор изображений значительно несбалансирован и для ряда подуровней жёсткости содержит крайне мало примеров. Исходя из этого было принято решение, используя определённое врачом число отчётливо видимых на снимках верхних грудных позвонков, разделить все снимки по уровню жёсткости на 3 группы, ориентируясь на медицинские критерии: мягкие (видно менее 3 позвонков), нормальные (видно 3-4 позвонка) и жёсткие (видно более 4 позвонков) [3, 4]. Количество снимков в сформированных группах

представлено на Рис. 6. Именно эти три класса рассматривались в качестве допустимых значений целевой переменной для задачи порядковой регрессии.

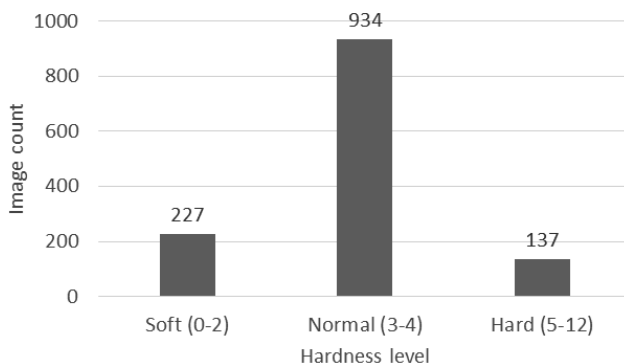


Рис. 6. Гистограмма распределения снимков набора SakhaTB по уровню жёсткости

Для решения задачи автоматического определения жёсткости рентгеновского снимка грудной клетки был разработан нейросетевой метод. На вход алгоритма поступает рентгеновский снимок грудной клетки и подвергается предобработке, затем при помощи нейронной сети ему ставится в соответствие вещественное число на отрезке $[0,1]$, которое является внутренним безразмерным показателем жёсткости снимка, на основании которого после сравнения с настраиваемыми порогами изображение относится к одному из рассматриваемых классов жёсткости. Пороги являются частью модели и настраиваются вместе с весами слоёв нейронной сети в процессе обучения. Преимуществом такого подхода является то, что с помощью внутреннего показателя жёсткости можно ранжировать изображения относительно друг друга даже в том случае, когда обрабатываемое изображение значительно отличается от обучающей выборки и для такого снимка пороги разделения классов жёсткости могут быть настроены неверно.

Для решения задач обработки медицинских изображений и, в частности, диагностики заболеваний широко используются [10] свёрточные нейронные сети семейств ResNet [14], DenseNet [15] и др.

В данной работе для решения задачи определения жёсткости снимков, вследствие малого размера выборки, была использована компактная сеть ResNet-18 с меньшим числом параметров и, следовательно, меньшей склонностью к переобучению по сравнению с более крупными сетями этой же архитектуры или представителями других вышеупомянутых архитектур.

Также было проведено сравнение качества решения этой задачи с таковым у компактной сети более новой архитектуры свёрточных нейронных сетей EfficientNetV2-S [16], так как в задачах классификации изображений она проявляет себя лучше перечисленных выше. Её основное отличие от них заключается в оптимизации работы свёрточных слоёв с помощью их пропорционального масштабирования, смены порядка операций разной размерности, уменьшения размера ядер свёртки и отказа от «тяжёлых» слоёв, что позволяет уменьшить потребление памяти и задействовать освободившиеся ресурсы на увеличение глубины и обобщающей способности нейронной сети.

Для ускорения процесса обучения за счёт наличия готовых низкоуровневых фильтров при решении обеих задач в качестве начального состояния нейронной сети использовались веса соответствующей модели, обученной на решение задачи классификации изображений реального мира ImageNet-1K [17]. Последний слой был заменён на полносвязный слой с 1 выходом и 2 настраиваемыми в процессе обучения параметрами-порогами пороговой модели порядковой регрессии [18].

Хотя финальным критерием при определении уровня жёсткости рентгеновского снимка является число чётко контурируемых верхних грудных позвонков [3, 4], для определения уровня контраста перед исследованием требуется обращать внимание на видимость других областей грудной клетки, (например, на элементы лёгочного рисунка) и органов [4]. На основании этого было принято решение не ограничивать область исследования на снимках пределами грудной клетки, а рассматривать изображения целиком.

Этап предобработки заключается в выполнении последовательности следующих шагов:

1. автоматическое контрастирование изображения:

$$h(x) = 255 * \frac{x - p_{0.5}}{p_{99.5} - p_{0.5}},$$

где $p_{0.5}(x)$ и $p_{99.5}(x)$ – 0.5%- и 99.5%-перцентили значений интенсивности пикселей изображения;

2. автоматическое гамма-преобразование интенсивности пикселей изображения:

$$\begin{cases} g(x) = 255 \left(\frac{x}{255} \right)^\gamma, \\ \gamma = \log_\mu 0.5, \end{cases}$$

где x – значение интенсивности пикселя входного изображения, μ – средняя интенсивность всего изображения;

3. уменьшение изображения до входного разрешения нейронной сети (512x512 пикселей для ResNet-18 и 384x384 пикселей для EfficientNetV2-S);
4. опциональный шаг глобальной или локальной (CLANE [19]) эквализации гистограммы. Размер в пикселях стороны окна, используемого в методе локальной эквализации гистограммы, прямо зависит от размера стороны изображения и составляет $\frac{1}{2^n}$ от него, где $n \in \mathbb{N}$ – параметр метода. Влияние наличия этого шага и размера окна на качество работы алгоритма будет показано ниже.

В качестве функции потерь была выбрана функция потерь пороговой порядковой регрессии в виде суммы слагаемых, число которых зависит от числа классов (all-threshold) [18]:

$$\begin{cases} L(z, y) = \sum_{k=1}^{K-1} f(s(k, y)(\theta_k - z)), \\ s(k, y) = \begin{cases} -1, & k < y, \\ +1, & k \geq y, \end{cases} \end{cases}$$

где z – выход нейронной сети (безразмерный показатель жёсткости), принимающий значения из отрезка $[0,1]$ (чем ближе к 1, тем выше жёсткость снимка), y – истинный класс соответствующего этому выходу изображения, K – общее число классов, $\theta_1 < \theta_2 < \dots < \theta_{K-1}$ – пороги, делящие действительную прямую на K частей, а $f(x)$ – базовая функция потерь бинарной классификации, в качестве которой использовалась логистическая функция потерь:

$$f(x) = \ln \frac{1}{1 + e^{-x}}.$$

Так как даже после сведения количества классов в задаче порядковой регрессии к трём выборка всё равно осталась сильно несбалансированной, то применялось взвешивание значений функции потерь для каждого примера с весами, обратно пропорциональными количеству изображений соответствующего класса. В роли меры качества порядковой регрессии выступала сбалансированная по классам средняя абсолютная ошибка (macro-averaged MAE, далее – mMAE) [20].

Также в целях получения базовой оценки качества была обучена модель для простой классификации изображений на 3 класса. Последний слой был заменён на полносвязный с 3 выходами. Роль функции потерь выполняла кросс-энтропия:

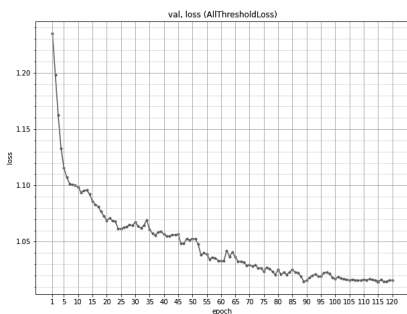
$$\left\{ \begin{array}{l} CE(z, y) = \sum_{k=1}^K \mathbb{I}[y = k] \cdot \ln(\text{softmax}(z)_k) \\ \text{softmax}(z)_k = \frac{e^{z_k}}{\sum_{i=1}^K e^{z_i}}, \\ \mathbb{I}[y = k] = \begin{cases} 1, y = k, \\ 0, y \neq k, \end{cases} \end{array} \right. ,$$

а роль меры качества – сбалансированная точность (далее – BalAcc) [21].

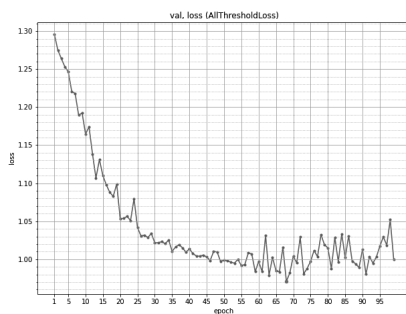
Для обучения рассмотренных моделей весь набор данных делился на обучающую, валидационную и тестовую выборку в соотношении 64:16:20 с предварительным случайным перемешиванием изображений и стратификацией для сохранения пропорций между классами. Обучающая выборка подвергалась случайным преобразованиям:

- поворотам (в пределах 15 градусов в каждую сторону),
- масштабированию (коэффициент выбирался случайно из отрезка [0.8,1.2]),
- сдвигам (до 30% размера изображения по каждой оси),
- изменению яркости и контраста (до 20% в каждую сторону).

Для оптимизации функции потерь применялся алгоритм градиентного спуска AdamW [22] с параметрами $lr = 5 \cdot 10^{-6}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\lambda = 0.01$. Размер пакета (англ. batch) изображений был равен 64 для модели на основе ResNet-18 и 16 для модели на основе EfficientNetV2-S; в обоих случаях применялось накопление градиента на протяжении 8 и 2 итераций соответственно (до достижения размера «виртуального пакета» в 128 объектов). В конце каждой эпохи качество модели замерялось на валидационной выборке, при отсутствии уменьшения значения функции потерь на ней в течение 10 эпох шаг градиентного спуска уменьшался в 5 раз, а при отсутствии улучшения в течение 31 эпохи обучение прекращалось. Переобучение контролировалось замерами функции потерь и меры качества, однако значительного ухудшения качества работы модели на валидационной выборке с течением времени не наблюдалось (см. Рис. 7; можно заметить небольшой рост значений функции потерь у модели на основе EfficientNetV2-S, что свидетельствует о переобучении модели), поэтому в качестве финального состояния принималось значение весов в конце последней эпохи.



а) ResNet-18



б) EfficientNetV2-S

Рис. 7. Примеры графиков зависимости функции потерь от количества эпох на валидационной выборке в задаче определения уровня жёсткости снимков

Итоговые значения мер качества, полученные на тестовой выборке, представлены в Табл. 2. Модели для решения задачи порядковой регрессии содержат «ord» в своём названии, модель для решения задачи классификации – «clf». Модель «ord-eff», была создана на основе EfficientNetV2-S, а остальные модели – на основе ResNet-18. В столбце «Эквализация гистограммы» стоит прочерк (не применялась), «глобальная» в случае использования глобальной эквализации или размер окна локальной эквализации гистограммы в виде доли от размера всего изображения.

Табл. 2. Значения мер качества работы алгоритмов определения жёсткости на тестовой выборке набора SakhaTB

Модель	Эквализация гистограммы	BalAcc	mMAE
clf	-	0.563	0.452
ord-eff	-	0.623	0.399
ord	-	0.609	0.452
ord-glob	глобальная	0.609	0.452
ord-clahe2	$\frac{1}{2}$ стороны изображения	0.636	0.398
ord-clahe4	$\frac{1}{4}$ стороны изображения	0.610	0.449
ord-clahe8	$\frac{1}{8}$ стороны изображения	0.593	0.468
ord-clahe16	$\frac{1}{16}$ стороны изображения	0.600	0.468

На основании сбалансированной точности и сбалансированной MAE лучшей моделью оказалась модель порядковой регрессии с локальной эквализацией гистограммы на этапе предобработки снимков с окном, длина стороны которого была равна $\frac{1}{2}$ длины стороны изображения. Далее для краткости она будет обозначена «ord-clahe2». Дальнейший анализ изображений проводится с использованием этой модели.

На Рис. 8 представлена зависимость вероятностей классов «Жёсткий» (Hard) и «Мягкий» (Soft), предсказанных моделью «clf» для решения задачи простой классификации, а также безразмерной величины жёсткости, предсказанной моделью «ord-clahe2» для задачи порядковой регрессии, и истинного значения уровня жёсткости объектов тестовой выборки. Можно заметить, что разделение классов далеко от идеального.

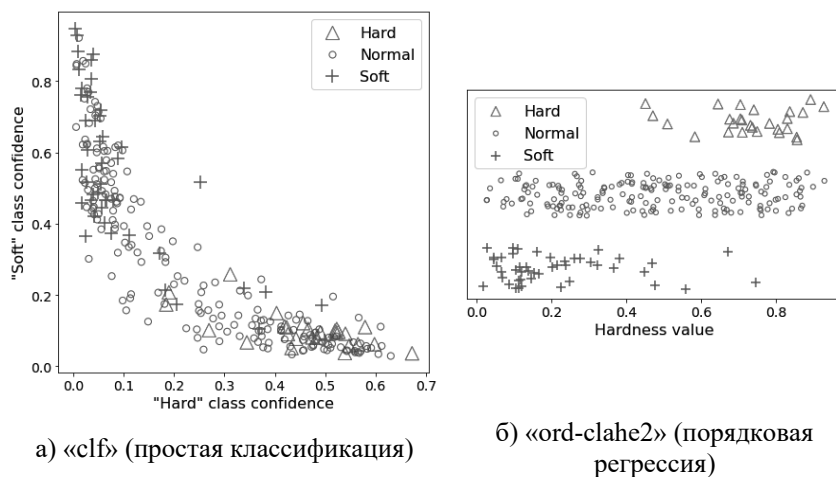


Рис. 8. Зависимость предсказаний моделей в зависимости от истинного класса объекта для набора SakhaTV

Как выяснилось после проверки ошибочно классифицированных снимков, причиной тому послужила шумность использованного набора данных из-за неоднозначности и недостаточной определённости критериев разметки. Подтверждением этого предположения также служит близость значений меры качества на обучающей выборке к таковым на тестовой выборке: сбалансированная точность около 0.70 и 0.67 у моделей «clf» и «ord-clahe2» соответственно.

Рассмотрение задачи определения жёсткости рентгеновского снимка как задачи порядковой регрессии позволяет, используя полученную для её решения модель, с некоторой точностью ранжировать изображения по жёсткости на основе внутреннего показателя жёсткости нейронной сети. В

качестве меры качества ранжирования был выбран коэффициент ранговой корреляции Спирмена [23], так как он позволяет обнаруживать в том числе нелинейные зависимости величин. Замеры делались относительно указанных врачом количества чётко видимых на снимке позвонков и относительно классов жёсткости снимков. Его значения приведены в Табл. 3. Распределение снимков из тестовой выборки и из наборов TBX11K и Montgomery-Shenzhen по безразмерному показателю жёсткости, предсказанному моделью «ord-clahe2», представлен на Рис. 9. Однако следует осторожно относиться к разделению снимков из наборов TBX11K и Montgomery-Shenzhen на классы жёсткости, так как истинные пороги разделения классов жёсткости для этих данных могут сильно отличаться от полученных сетью из-за их визуального отличия от изображений обучающей выборки.

Табл. 3. Значения мер качества ранжирования тестовой выборки набора SakhaTB алгоритмами определения жёсткости

Модель	Spearman (позвонки)	Spearman (жёсткость)
ord-eff	0.564	0.457
ord	0.576	0.497
ord-glob	0.599	0.514
ord-clahe2	0.606	0.534
ord-clahe4	0.602	0.519
ord-clahe8	0.588	0.498
ord-clahe16	0.596	0.507

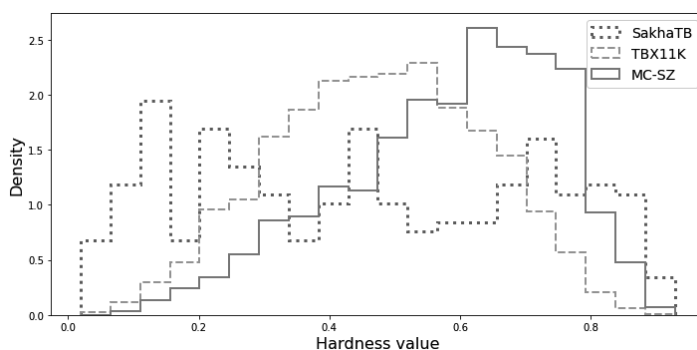
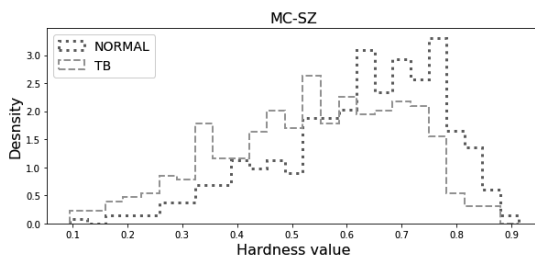


Рис. 9. Гистограммы распределения снимков 3 используемых наборов изображений по предсказанному моделью «ord-clahe2» показателю жёсткости

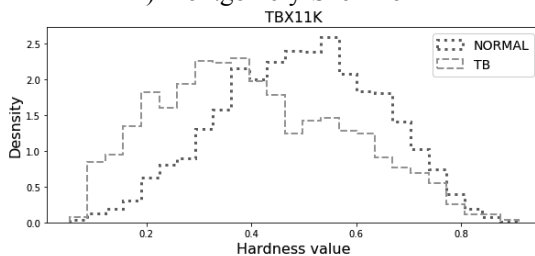
Была проведена оценка качества работы модели «ord-clahe2» на наборе MC-SZ. Для этого он также был размечен аналогично набору SakhaTB: для каждого изображения было измерено число отчётливо видимых на этом снимке верхних грудных позвонков. Результаты сравнения качества работы алгоритма на наборе MC-SZ и на тестовой выборке набора SakhaTB приведены в Табл. 4. Отдельные гистограммы для классов NORMAL и TB для наборов MC-SZ и TBX11K представлены на Рис. 10. Небольшие различия гистограмм классов совпадают с визуальной разницей между снимками этих классов: класс TB в обоих из них содержит больше мягких снимков, а классы NORMAL – больше жёстких (см. Рис. 4 и Рис. 5).

Табл. 4. Сравнение качества работы модели «ord-clahe2» на наборах SakhaTB и MC-SZ

Набор данных	BalAcc	mMAE	Spearman (позвонки)	Spearman (жёсткость)
SakhaTB	0.636	0.398	0.606	0.534
MC-SZ	0.546	0.565	0.325	0.203



a) Montgomery-Shenzhen



б) TBX11K

Рис. 10. Гистограммы распределения снимков каждого класса использованных в задаче диагностики туберкулёза наборов изображений по предсказанному моделью «ord-clahe2» показателю жёсткости

Нейросетевая диагностика туберкулеза с учетом жёсткости снимков

На основании предсказанных полученной на предыдущем этапе моделью величин жёсткости было проведено удаление из наборов одинаковой доли самых жёстких и самых мягких снимков (то есть с обеих сторон гистограммы распределения). После этого качество обученной на прореженной обучающей выборке модели замерялось на прореженной тестовой выборке и сравнивалось с качеством работы на такой же тестовой выборке модели, настроенной на непрореженной обучающей выборке.

Исходя из визуального различия всех трёх наборов данных и малого размера набора MC-SZ, было принято решение прореживать не объединённый набор изображений, а каждую из двух его частей отдельно, чтобы изменение их пропорций в итоговой выборке не повлияло на качество работы. Такой «осторожный» подход обусловлен необходимостью правильного контрастирования снимков перед определением жёсткости, однако решение этой задачи выходит за рамки данной работы.

Помимо прореживания выборок с обеих сторон гистограммы уровня жёсткости был рассмотрен и случай отбрасывания только самых жёстких снимков, так как на мягких снимках ещё могут сохраняться некоторые детали лёгочной ткани, в то время как на жёстких они могут быть полностью утеряны.

Алгоритм диагностики туберкулеза выглядел следующим образом: на вход поступает рентгеновский снимок грудной клетки и подвергается предобработке, затем при помощи нейронной сети ему ставятся в соответствие 2 вещественных числа на отрезке $[0,1]$, которые выражают веса классов NORMAL и TB. Сумма весов для каждого изображения равна 1. Класс, чей вес больше, принимается за выходное значение алгоритма. Веса слоёв нейронной сети настраиваются в процессе обучения.

В качестве основ для нейронных сетей использовались EfficientNetV2-S и ResNet-18, последний слой которых заменялся на полносвязный с 2 выходами. Процедуры деления набора данных на подвыборки и обучения модели со взвешиванием классов для балансировки, а также начальные состояния весов и этапы предобработки были аналогичны таковым из метода в предыдущем разделе, однако эквализация гистограммы не проводилась. В роли функции потерь выступала кросс-энтропия, а в роли меры качества – сбалансированная точность.

При решении этой задачи из двух моделей лучшее качество классификации до прореживания показала модель на основе EfficientNetV2-S, и именно она участвует в сравнении. Вероятно, это обусловлено значительно возросшим объёмом и качеством выборки.

Результаты представлены в Табл. 5. Из неё видно, что изменение качества зависит от степени прореживания изображений, однако остаётся стабильно положительным. Сравнение показателей чувствительности и специфичности для класса ТВ приведено в Табл. 6.

Табл. 5. Сравнение качества классификации моделей, обученных на полном и прореженном наборе (сбалансированная точность)

	Удаление жёстких и мягких изображений			Удаление только жёстких изображений		
Доля удалённых	5%	10%	15%	5%	10%	15%
До	0.958	0.951	0.951	0.962	0.961	0.965
После	0.961	0.962	0.953	0.968	0.966	0.975

Табл. 6. Сравнение качества классификации моделей, обученных на полном и прореженном наборе (чувствительность / специфичность)

	Удаление жёстких и мягких изображений			Удаление только жёстких изображений		
Доля удалённых	5%	10%	15%	5%	10%	15%
До	0.923/ 0.994	0.909/ 0.994	0.908/ 0.995	0.930/ 0.994	0.927/ 0.995	0.934/ 0.996
После	0.933/ 0.990	0.933/ 0.991	0.915/ 0.990	0.943/ 0.994	0.941/ 0.992	0.958/ 0.993

Заключение

В данной работе была продемонстрирована возможность использования нейросетевых методов для определения уровня жёсткости рентгеновских снимков грудной клетки. Хотя из-за сложности задачи не удалось достичь высоких показателей качества, полученные результаты значительно превышают возможности случайного выбора ответа: сбалансированная точность составила 0.636, а коэффициенты корреляции Спирмена – 0.606 и 0.534. Однако качество работы полученного алгоритма заметно снижается в случае применения его к данным из других источников: для набора MC-SZ сбалансированная точность упала с 0.636 до 0.546, а значения меры качества ранжирования объектов снизились примерно вдвое. Для сохранения качества работы алгоритма на изображениях, отличающихся от таковых в обучающей выборке, необходимо применение разнообразных методик, таких как расширение обучающей выборки, контрастирование изображений с приведением их к единому стандарту, решение важной для медицины задачи cross-dataset generalization [24]. Вероятно, использование более строгих критериев

разметки изображений даст прирост стабильности работы метода, а использование в качестве дополнительных функций потерь *contrastive loss* и *triplet loss* повысит качество упорядочивания снимков относительно друг друга.

Однако даже такая несовершенная модель определения жёсткости позволяет увеличить качество работы алгоритма диагностики туберкулёза при условии предварительной фильтрации изображений перед обучением классификатора и получением предсказаний. Сокращение разброса жёсткости данных и увеличение их однородности заметно повысило точность обнаружения больных пациентов при сохранении или малом снижении специфичности: наибольший абсолютный и относительный прирост чувствительности алгоритма для класса ТВ наблюдался при удалении 10% самых жёстких и 10% самых мягких снимков (с 0.909 до 0.933) и при удалении 15% самых жёстких снимков (с 0.934 до 0.958). Во втором случае также было достигнуто наибольшее значение чувствительности: 0.958.

Благодарности

Работа выполнена при поддержке гранта РФФИ 19-5780014 БРИКС_Т (BRICS2019-394).

Литература

1. *Finlayson S.G. et al.* Adversarial attacks on medical machine learning // Science. – 2019. – Т. 363. – №. 6433. – С. 1287-1289.
2. *Chuiko G.A., Tsvetkov V.M.* Effects of X-ray hardness on fluorogram informativeness // Biomedical Engineering. – 1982. – Т. 16. – №. 4. – С. 117-119.
3. Тимофеева Л.А., Алешина Т.Н., Быкова А.В. Основные рентгенологические синдромы патологии легочной ткани: учеб. пособие // Чебоксары: Изд-во Чуваш. ун-та, 2013. – 104 с.
4. Сидоров А.И., Щербатых А.А., Покровская Л.Н. Методика анализа рентгенограмм: учебное пособие // Под ред. Синькова А.В. ГБОУ ВПО ИГМУ Минздрава. – Иркутск: ИГМУ, 2012. – 22 с.
5. *Nousiainen K. et al.* Automating chest radiograph imaging quality control // Physica Medica. – 2021. – Т. 83. – С. 138-145.
6. *von Berg J. et al.* Robust chest x-ray quality assessment using convolutional neural networks and atlas regularization // Medical Imaging 2020: Image Processing. – SPIE, 2020. – Т. 11313. – С. 391-398.
7. *Xiao-Qian J.I.A. et al.* Application value of convolutional neural network in quality control of direct digital chest X-ray images // Xi'an jiao tong da xue xue bao. Yi xue ban. – 2019. – №. 5. – С. 784.

8. *Sadre R. et al.* Validating deep learning inference during chest X-ray classification for COVID-19 screening // Scientific reports. – 2021. – T. 11. – №. 1. – C. 1-10.
9. *Dovganich A.A., Khvostikov A.V., Pchelintsev Y.A. et al.* Automatic out-of-distribution detection methods for improving the deep learning classification of pulmonary X-ray images // Journal of Image and Graphics. — 2022. — T. 10, №. 2.
10. *Oloko-Oba M., Viriri S.* A systematic review of deep learning techniques for tuberculosis detection from chest radiograph // Frontiers in Medicine. – 2022. – T. 9.
11. *Jaeger S. et al.* Automatic tuberculosis screening using chest radiographs // IEEE transactions on medical imaging. – 2013. – T. 33. – №. 2. – C. 233-245.
12. *Candemir S. et al.* Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration // IEEE transactions on medical imaging. – 2013. – T. 33. – №. 2. – C. 577-590.
13. *Liu Y. et al.* Rethinking computer-aided tuberculosis diagnosis // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. – 2020. – C. 2646-2655.
14. *He K. et al.* Deep residual learning for image recognition // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – C. 770-778.
15. *Huang G. et al.* Densely connected convolutional networks // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2017. – C. 4700-4708.
16. *Tan M., Le Q.* EfficientNetV2: Smaller models and faster training // International Conference on Machine Learning. – PMLR, 2021. – C. 10096-10106.
17. *Russakovsky O. et al.* ImageNet large scale visual recognition challenge // International journal of computer vision. – 2015. – T. 115. – №. 3. – C. 211-252.
18. *Rennie J.D.M., Srebro N.* Loss functions for preference levels: Regression with discrete ordered labels // Proceedings of the IJCAI multidisciplinary workshop on advances in preference handling. – AAAI Press, Menlo Park, CA, 2005. – T. 1.
19. *Pizer S.M. et al.* Adaptive histogram equalization and its variations // Computer vision, graphics, and image processing. – 1987. – T. 39. – №. 3. – C. 355-368.

20. *Baccianella S., Esuli A., Sebastiani F.* Evaluation measures for ordinal regression // 2009 Ninth international conference on intelligent systems design and applications. – IEEE, 2009. – C. 283-287.
21. *Brodersen K.H. et al.* The balanced accuracy and its posterior distribution // 2010 20th international conference on pattern recognition. – IEEE, 2010. – C. 3121-3124.
22. *Loshchilov I., Hutter F.* Decoupled weight decay regularization // arXiv preprint arXiv:1711.05101. – 2017.
23. *Zwillinger D., Kokoska S.* CRC standard probability and statistics tables and formulae. // New York: Chapman & Hall, 2000.
24. *Thambawita V. et al.* An extensive study on cross-dataset bias and evaluation metrics interpretation for machine learning applied to gastrointestinal tract abnormality classification // ACM Transactions on Computing for Healthcare. – 2020. – T. 1. – №. 3. – C. 1-29.