

МЕТОДЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

УДК 004.93

О ВЛИЯНИИ КАЧЕСТВА БАЗ ДАННЫХ РЕНТГЕНОВСКИХ
СНИМКОВ БОЛЬНЫХ ТУБЕРКУЛЁЗОМ
НА ДИАГНОСТИКУ БОЛЕЗНИ© А. В. Карпов², В. И. Козик¹, Е. С. Нежевенко¹, Я. Ш. Шварц²

¹Институт автоматизации и электрометрии СО РАН,
630090, г. Новосибирск, просп. Академика Коптюга, 1

²Новосибирский научно-исследовательский институт туберкулёза
Министерства здравоохранения РФ,
630040, г. Новосибирск, ул. Охотская, 81А
E-mail: nedj@iae.nsk.su

Исследуется надёжность обучающих выборок, используемых для обучения свёрточных нейронных сетей, которые применяются для диагностики лёгочных заболеваний. На примере базы Kaggle показано, что выборка, в которой 3500 рентгенограмм здоровых пациентов и столько же рентгенограмм больных туберкулёзом, весьма неоднородна. При обучении на разных частях выборки и распознавании различных её частей получаются существенно различные результаты.

Ключевые слова: туберкулёз, рентгенограмма, свёрточные нейронные сети, диагноз, обучающие выборки.

DOI: 10.15372/AUT20220508

Введение. Ежегодно в мире туберкулёзом (ТБ) заболевают до 10 млн человек. В последние годы в РФ наблюдается снижение заболеваемости и смертности от ТБ, однако по-прежнему эти показатели остаются высокими, особенно в Сибирском и Дальневосточном федеральных округах. Большая площадь, занимаемая округами, низкая плотность населения и недостаточно развитая инфраструктура создают дополнительные трудности в организации оказания противотуберкулёзной помощи, косвенно влияя на сохранение благополучия в регионах. Сегодня есть основания предполагать, что пандемия COVID-19, снижая выявляемость ТБ и доступность противотуберкулёзной помощи, может нивелировать многие успехи последних лет в борьбе с ТБ.

Одна из ключевых задач в решении проблемы ТБ — ранняя диагностика заболевания. Во многом она может быть решена за счёт активного выявления ТБ с помощью регулярного флюорографического и рентгенологического обследования населения. Однако в связи с многообразием клинико-рентгенологических вариантов ТБ лёгких на этапе первичной медико-санитарной помощи почти в 40 % случаев ставится неверный диагноз. Одна из причин неверной диагностики связана с тем, что в последнее время туберкулёз лёгких всё чаще проявляется атипичной рентгенологической картиной. Основными причинами позднего выявления ТБ лёгких в учреждениях общей лечебной сети являются неправильная интерпретация клинических и рентгенологических проявлений заболевания при отсутствии бактериологического подтверждения, тяжёлые фоновые заболевания и сопутствующая патология.

Более того, дополнительную сложность в проблему ранней диагностики туберкулёза вносит латентная туберкулёзная инфекция (ЛТИ) — состояние, при котором микобакте-

рии туберкулёза присутствуют в организме человека, обуславливая положительные иммунологические реакции на *M. tuberculosis* (специфические антигены), но при этом клинические и рентгенологические признаки туберкулёза, как традиционно считается, отсутствуют. Тем не менее существующие иммунологические тесты неспособны отличить ЛТИ от активного туберкулёза. Классическая клиническая симптоматика при туберкулёзе: кашель, мокрота, повышение температуры и снижение массы тела, как и при ЛТИ, может отсутствовать, а рентгенологические признаки, в том числе апикальный фиброз, хиллярная аденопатия и гранулематозные изменения, могут наблюдаться и при туберкулёзе, и при ЛТИ или быть остаточными явлениями. Сегодня считается, что ЛТИ следует рассматривать как состояние динамического равновесия между иммунной системой организма и микобактериями туберкулёза, при котором только у 5–10 % лиц с ЛТИ в течение жизни инфекционный процесс перейдёт в стадию активного туберкулёза. Дифференцировать ЛТИ и активный туберкулёз принципиально важно, так как в первом случае в группах высокого риска назначается короткий профилактический курс антибиотикотерапии, а во втором — очень длительная и весьма токсичная полихимиотерапия. Иными словами, часто имеется необходимость проводить рентгенологическую/рентгенотомографическую дифференциальную диагностику ЛТИ и туберкулёза на основании изучения минорных признаков.

Наряду с ранней диагностикой рентгенологическое обследование больного туберкулёзом требуется для наблюдения за клиническим развитием, эффективностью лечения и оценкой степени излечения. Дифференциальная диагностика патологических процессов в лёгких может быть критически важной не только для своевременного лечения туберкулёза. В наши дни стали очень частыми случаи сочетания туберкулёза и рака лёгких, когда значительно возросла распространённость обоих заболеваний. Понятно, что качественная и быстрая диагностика крайне необходимы, поскольку отсрочка хирургического вмешательства может иметь роковые последствия для пациента.

С учётом затруднения при распознавании туберкулёзных процессов в описание каждой клинической формы включены основные дифференциально-диагностические признаки, которые сопоставляются с проявлениями других лёгочных заболеваний. Сюда входят, в частности, очаговые тени, ограниченные и распространённые инфильтраты, округлые и шаровидные тени, мелкоочаговая (милиарная) и крупноочаговая диссеминация, просветление, кольцевидные тени, изменение корней лёгких и средостения, локализация фокусов поражения и наличие в них обызвествлённых включений, очаговых и рубцовых изменений в прилежащих участках лёгкого.

Современная рентгенология прошла большой путь развития от простой аналоговой трубки до мощных цифровых технологий. Но обычно обследование органов грудной полости начинается с классического рентгенологического исследования. На этом этапе применяются самые доступные основные рентгенологические методы: рентгенография, рентгенофлюорография, рентгеноскопия. Особую ценность в диагностике торакальной патологии занимает рентгенофлюорография. Как правило, она применяется для профилактических, массовых исследований с целью раннего выявления, прежде всего, туберкулёза и рака лёгких. Основное достоинство методики состоит в экономичности: пропускная способность составляет до 150 человек в час. Благодаря возможности получения крупнокадрового изображения (100 × 100 и 110 × 110 мм) флюорографию в настоящее время стали применять в качестве диагностической методики. Преимуществом рентгенографии и флюорографии является высокая разрешающая способность, объективная документация выявленных изменений, что позволяет достоверно судить об их динамике, сравнивая с предыдущими или последующими снимками.

Таким образом, диагностика лёгочных заболеваний (в частности, туберкулёза лёгких)

путём исследования рентгенографических изображений лёгких — один из самых эффективных методов решения этой проблемы [1, 2].

В последнее время для этой цели активно применяются глубокие свёрточные нейронные сети как самые передовые средства машинного распознавания. Поскольку важнейшим этапом этого процесса является обучение сети, то огромную роль играет качество обучающей выборки. Это тем более важно, что практически оценить её визуально даже опытному врачу довольно сложно — для этого необходимо просмотреть тысячи изображений и оценить их не только с точки зрения патологий, но и качества изображений.

Если база данных надёжна, результаты классификации при обучении сети на различных её участках должны быть близки. Проверка этой надёжности является целью данной работы.

Методы исследования. Поскольку была поставлена задача исследовать базы данных, а не сеть, принято решение взять для исследования одну из первых успешных свёрточных сетей AlexNet [3], причём сеть была предобучена на базе данных ImageNet, содержащей миллионы изображений. Схема этой сети приведена в табл. 1.

Здесь показаны все слои (всего 25 слоёв), из них входной слой 227×227 , пять свёрточных слоёв, а три последних слоя заменены: в слое output (classificationLayer) заменено

Таблица 1

Схема сети AlexNet

1	data	Входное изображение в окне $227 \times 227 \times 3$ с нормализацией
2	conv1	Свёртка 96 ядер $11 \times 11 \times 3$ с шагом [4 4] и заполнением [0 0 0 0]
3	relu1	Функция активации
4	norm1	Межканальная нормализация с пятью каналами на элемент
5	pool1	Выбор максимального значения в окне 3×3 с шагом [2 2] и заполнением [0 0 0 0]
6	conv2	Две группы свёрток 128 ядер $5 \times 5 \times 48$ с шагом [1 1] и заполнением [2 2 2 2]
7	relu2	Функция активации
8	norm2	Межканальная нормализация с пятью каналами на элемент
9	pool2	Выбор максимального значения в окне 3×3 с шагом [2 2] и заполнением [0 0 0 0]
10	conv3	Свёртка 384 ядер $3 \times 3 \times 256$ с шагом [1 1] и заполнением [1 1 1 1]
11	relu3	Функция активации
12	conv4	Две группы свёрток 192 ядер $3 \times 3 \times 192$ с шагом [1 1] и заполнением [1 1 1 1]
13	relu4	Функция активации
14	conv5	Две группы свёрток 128 ядер $3 \times 3 \times 192$ с шагом [1 1] и заполнением [1 1 1 1]
15	relu5	Функция активации
16	pool5	Выбор максимального значения в окне 3×3 с шагом [2 2] и заполнением [0 0 0 0]
17	fc6	Полносвязный слой 4096
18	relu6	Функция активации
19	drop6	Игнорировать 50 % нейронов
20	fc7	Полносвязный слой 4096
21	relu7	Функция активации
22	drop7	Игнорировать 50 % нейронов
23	fc8	Полносвязный слой 1000
24	prob	Softmax
25	output	Классифицированные выходные данные



Рис. 1. Пример изображения из базы данных Kaggle

количество выходных нейронов по числу классов, слой `fc` (`fullyConnectedLayer`) — это единственный полносвязный слой, который обучается на медицинской базе данных, заменяется также слой `prob softmaxLayer`.

Выберем конкретную базу данных для исследования. Существуют различные базы данных для отработки методов классификации туберкулёза, например база данных Национальной медицинской библиотеки США [4], Белорусская база данных [5], порталы по борьбе с туберкулёзом [6]. Нами была принята в качестве обучающей тестовой выборки база данных Kaggle [7], ведь именно на её основе проводятся соревнования и оценивается качество нейронных сетей. В этой базе всего два класса — 3500 изображений здоровых лёгких (класс `Normal`) и 3500 изображений лёгких, поражённых туберкулёзом (класс `Tuberculosis`). Исходная размерность изображений 512×512 . Пример изображения из этой выборки показан на рис. 1.

Отметим, что единственное преобразование изображений, которое было произведено над ними перед вводом в нейронную сеть — это изменение размерности до значения, требуемого выбранной нейронной сетью — 227×227 пикселей. Не производились нормализация яркости, контрастности, гистограммы, какие-либо морфологические преобразования. Сокращение размерности изображений выполнялось в программе `IrfanView`, использован ресемплинг методом свёрток [8].

В первом эксперименте изображения каждого из классов случайным образом делились на обучающую (`train`) и валидационную (`validation`) подвыборки в пропорции $0,7 : 0,3$. Тестовая выборка не выделялась, поскольку число эпох фиксировано (6). В процессе обучения контролировались точности классификации обучающей и валидационной выборок и соответствующие ошибки. Процесс обучения показан на рис. 2. Обе кривые (точность и ошибка) выходят на асимптоту на второй эпохе обучения, правда, в течение последующих эпох случаются выбросы вплоть до окончания шестой эпохи. Кривые точности на обучающей и валидационной выборках несколько расходятся только в первой эпохе, далее они практически неразличимы, что говорит об отсутствии переобучения. На рисунке помимо кривых обучения справа можно видеть и параметры этого процесса.

Результаты исследования. Финальный результат распознавания в данной работе — точность валидационной выборки (`validation accuracy` 99,33 %). Это довольно высокая точность классификации. Так, группа исследователей [10], создавшая использованную нами базу данных, получила точность классификации 98,3 %. В этой работе они пишут:

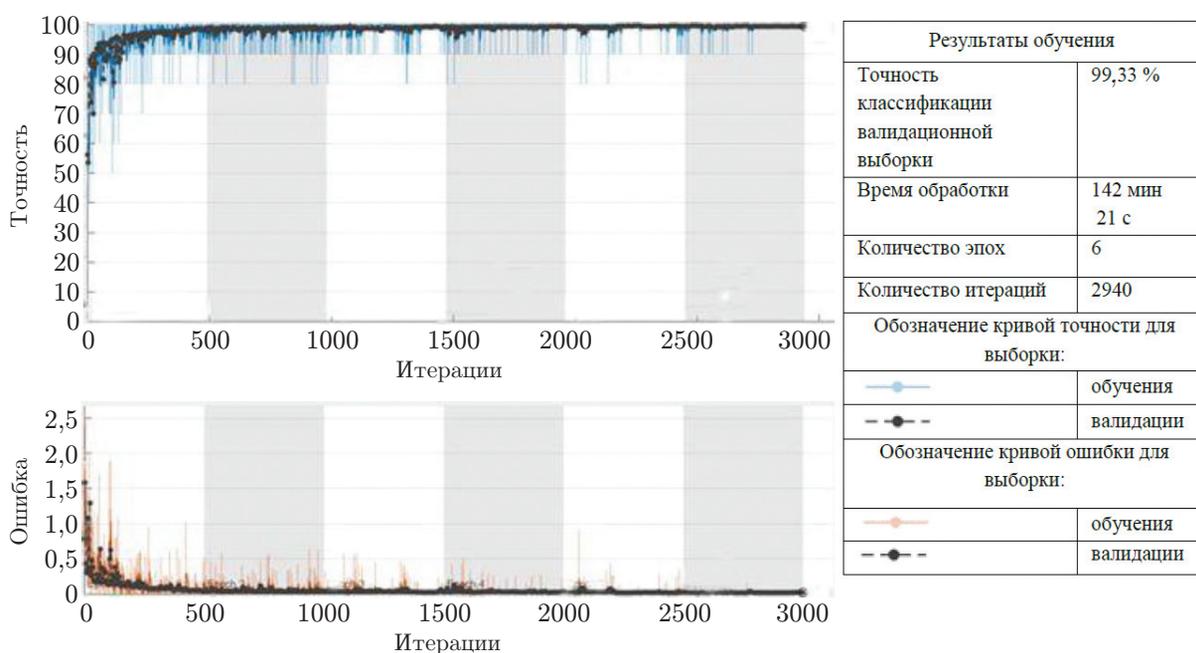


Рис. 2. Процесс обучения сети на выборке из 7000 изображений

«Убедитесь, что вы доверяете нам при использовании набора данных, кода и обученных моделей». В представленной работе был проведён эксперимент по проверке надёжности набора данных. Эксперимент заключался в том, что вся обучающая выборка была разбита на 7 подвыборок по 500 изображений в каждой, затем каждая из них была использована для обучения соответствующего количества нейронных сетей и через все сети пропущены имеющиеся подвыборки с применением их в качестве валидационных и определением каждый раз точности классификации.

Результаты этого исследования приведены в табл. 2. В первом столбце приведены номера рентгенограмм, используемых для обучения, во втором столбце — номера валидационных рентгенограмм, в третьем — точность классификации.

Видно, что результаты исследования неоднозначны: база данных неоднородна и точность классификации различных частей существенно зависит от того, какие её части используются для обучения. Для доказательства того, что эти различия не являются статистическим разбросом, сеансы обучения для каждой части выборки повторялись несколько раз, при этом результат распознавания (точность классификации) менялся на десятые доли процента, в то время как различие точности классификации при обучении на различных частях выборки достигало десятки процентов.

Ещё один эксперимент был проведён для проверки влияния обучения на разных частях базы данных на точность классификации сторонних рентгеновских снимков, не имеющих отношения к используемой базе данных, диагностика которых проводилась предварительно высококвалифицированными специалистами. Для этого ими были представлены рентгеновские снимки шести пациентов со следующими диагнозами:

- 1) правосторонний костально-базальный плеврит,
- 2) очаговый туберкулёз правого лёгкого в стадии уплотнения и петрификации,
- 3) очаговый туберкулёз лёгких в стадии инфильтрации,
- 4) диссеминированный туберкулёз правого лёгкого,
- 5) здоров,

Таблица 2

**Результаты распознавания различных
частей выборки при обучении
на разных подвыборках**

Номера снимков в обучающих подвыборках	Номера снимков в валидационных подвыборках	Точность
1–500	1–500	0,998
	501–1000	0,982
	1001–1500	0,984
	1501–2000	0,986
	2001–2500	0,928
	2501–3000	0,979
	3001–3500	0,604
1501–2000	1–500	0,985
	501–1000	0,982
	1001–1500	0,960
	1501–2000	0,999
	2001–2500	0,939
	2501–3000	0,983
	3001–3500	0,693
2001–2500	1–500	0,981
	501–1000	0,966
	1001–1500	0,964
	1501–2000	0,978
	2001–2500	0,997
	2501–3000	0,977
	3001–3500	0,963
2501–3000	1–500	0,991
	501–1000	0,988
	1001–1500	0,974
	1501–2000	0,987
	2001–2500	0,927
	2501–3000	1
	3001–3500	0,611
3001–3500	1–500	0,950
	501–1000	0,967
	1001–1500	0,939
	1501–2000	0,939
	2001–2500	0,943
	2501–3000	0,962
	3001–3500	1

Таблица 3

**Результаты распознавания
рентгеновских снимков пациентов
с достоверными диагнозами
при обучении на разных подвыборках**

Номера снимков в обучающих подвыборках	Диагнозы, поставленные сетью	Вероятность заболевания туберкулезом
1–500	Normal	0,996
	Normal	0,999
	Tub	0,029
	Tub	0,468
	Normal	0,999
	Normal	0,656
1501–2000	Normal	1
	Normal	1
	Tub	0,003
	Tub	0,767
	Normal	1
	Normal	0,047
2001–2500	Normal	0,999
	Normal	0,991
	Normal	0,940
	Tub	0,018
	Normal	1
	Tub	0,133
2501–3000	Normal	0,985
	Normal	0,985
	Tub	0,023
	Tub	0,113
	Normal	0,981
	Normal	0,714
3001–3500	Normal	0,970
	Tub	0,017
	Tub	0,043
	Tub	$4,79 \cdot 10^{-6}$
	Normal	0,992
	Normal	0,598
1–7000	Normal	1
	Normal	0,994
	Tub	0,023
	Tub	$5,94 \cdot 10^{-4}$
	Normal	1
	Normal	0,996

б) правосторонняя полисегментарная пневмония.

Из этого следует, что трое больны туберкулёзом, один здоров, а у двоих есть лёгочная патология, но не туберкулёз. Результат классификации снимков сетями, обученными на разных подвыборках, приведён в табл. 3. Первый столбец — номера снимков в обучающих подвыборках (последняя строка таблицы — это обучение сети на всей обучающей выборке), второй — диагнозы, поставленные сетью. Третий столбец требует особого пояснения. Это не точность классификации, как в табл. 2, которая определяется отношением числа правильных диагнозов к общему числу диагностируемых. Третий столбец — это уровень значимости (level of significance — уровень достоверности, надёжности, доверительный уровень). В медицинском смысле значение в этом столбце — степень уверенности врача в диагнозе. Видно, что только одна обучающая подвыборка из шести дала абсолютно правильный ответ: точно определила больных туберкулёзом, здорового человека, что же касается больных, но не туберкулёзом, то они были признаны здоровыми, причём уровень достоверности этого, очевидно, зависит от того, насколько болезнь этого пациента близка к туберкулёзу. Так, одна из пациенток с правосторонней полисегментарной пневмонией была признана здоровой (вернее, не больной туберкулёзом) с уровнем достоверности 59,9 %. Другими словами, при обучении на разных частях выборки диагнозы очень нестабильны. Правда, здоровый пациент (позиция 5) определяется правильно на всех обучающих подвыборках.

Заключение. В данной работе показано, что база данных Kaggle весьма неоднородна, поскольку обучение на разных её частях даёт неоднозначные результаты. Не очень понятно, как проводить верификацию такой базы, ясно лишь, что необходимо очень внимательно относиться к базам данных, представленным даже на авторитетных ресурсах. Понятно, что по такой малой валидационной выборке трудно судить о качестве диагностики, но сохраняется уверенность, что при правильном выборе базы данных можно обеспечить качественную диагностику, причём этот выбор можно проводить, предварительно проверив качество базы данных на своих собственных достоверных ресурсах, конечно, более объёмных по сравнению с представленными.

Финансирование. Работа выполнена при поддержке Министерства науки и высшего образования в рамках выполнения работ по государственному заданию № 121022000116-0 в ИАиЭ СО РАН.

СПИСОК ЛИТЕРАТУРЫ

1. **Silverman C.** An appraisal of the contribution of mass radiography in the discovery of pulmonary tuberculosis // Amer. Rev. Tuberculosis. 1949. **60**, N 4. P. 466–482
2. **Van't Hoog A. H., Meme H. K., Laserson K. F. et al.** Screening strategies for tuberculosis prevalence surveys: The value of chest radiography and symptoms // PLoS One. 2012. **7**, N 7. Published online.
3. **Krizhevsky A., Sutskever I., Hinton G. E.** ImageNet classification with deep convolutional neural networks // Advances in Neural Information Processing Systems. 2012. **25**, Iss. 2. DOI: 10.1145/3065386.
4. **Database Resources** of the National Center for Biotechnology Information. NCBI Educational Resources United States National Library of Medicine // Nucleic Acids Research. 2016. **44**, Database Iss. D7-D19. DOI: 10.1093/nar/gkv1290.
5. **Кириченко В. В., Спринджук М. В., Сергеев Р. С. и др.** Белорусский портал о туберкулёзе: уникальная база данных в открытом доступе. URL: <http://rep.bsmu.by/bitstream/handle/BSMU/8658/белорусский%20портал%20о%20туберкулезе.Image.Marked.pdf?sequence=1&isAllowed=y> (дата обращения: 20.07.2022).

6. **Rosenthal A., Gabrielian A., Engle E. W. et al.** The TB Portals: An Open-Access, Web-Based Platform for Global Drug-Resistant-Tuberculosis Data Sharing and Analysis // Journ. Clinical Microbiol. 2017. **55**, N 11. DOI: 10.1128/JCM.01013-17.
7. **Tuberculosis (TB) Chest X-ray Database.** URL: <https://www.kaggle.com/tawsifurrahman/tuberculosis-tb-chest-xray-dataset> (дата обращения: 20.07.2022).
8. **Resizing and Cropping Images Using IrfanView.** URL: <https://support.imodules.com/hc/en-us/articles/218261358-Resizing-and-Cropping-Images-Using-IrfanView> (дата обращения: 20.07.2022).
9. **Rahman T., Khandakar A., Kadir M. A. et al.** Reliable tuberculosis detection using chest X-Ray with deep learning, segmentation and visualization // IEEE Access. 2020. **8**. P. 191586–191601.

Поступила в редакцию 20.07.2022

После доработки 02.08.2022

Принята к публикации 29.08.2022
