

<https://doi.org/10.21869/2223-1536-2026-16-2-122-149>



УДК 004.891.3, 004.932

Двухконтурная валидация нейросетевой сегментации пневмоторакса в условиях дисбаланса классов

А. В. Косинов¹✉, Р. А. Крупчатников¹, В. В. Серебровский¹,
М. А. Ефремов¹, Д. С. Кондрашов¹

¹ Юго-Западный государственный университет
ул. 50 лет Октября, д. 94, г. Курск 305040, Российская Федерация

✉ e-mail: skosinka@yandex.ru

Резюме

Целью исследования является разработка методики клинически ориентированного обучения, валидации и выбора нейросетевой модели сегментации пневмоторакса на рентгенограммах органов грудной клетки при выраженном дисбалансе классов и необходимости контролируемого ограничения ложных срабатываний в составе систем поддержки принятия решений.

Методы. В работе предложена двухконтурная схема оценивания качества сегментации, включающая технический контур, служащий для контроля сходимости обучения по значениям функции потерь и сырым показателям, а также клинический контур для оценки бинаризованных масок после фиксированной постобработки результатов. Для каждой эпохи обучения выполняется перебор порога бинаризации; рабочая точка считается допустимой при выполнении клинических ограничений, после чего итоговая эпоха выбирается как максимизирующая индекс Дайса на положительных случаях среди допустимых решений. Сегментатор реализован на основе архитектуры U-Net++ и обучается по двухэтапной схеме с повышением пространственного разрешения с использованием гибридной функции потерь.

Результаты. На клинически репрезентативном материале – наборе разработки SIIM-ACR – финальная конфигурация, выбранная по протоколу на 15-й эпохе, обеспечила индекс Дайса на положительных случаях = 0,753 (95% ДИ 0,734–0,771) при частоте ложноположительных срабатываний = 0,177 (95% ДИ 0,158–0,197) и чувствительности на уровне изображения = 0,948 и ощутимо превзошла базовую конфигурацию.

Заключение. Предложенная методика двухконтурной валидации и критериального выбора эпохи в допустимой области обеспечивает воспроизводимый контроль компромисса между качеством сегментации на патологических снимках и частотой ложных тревог на снимках без патологии, снижая риск выбора субоптимальной модели при ориентировании только на минимум функции потерь. Полученные результаты подтверждают целесообразность использования разработанного протокола при проектировании сегментационных модулей для систем поддержки принятия решений в задачах лучевой диагностики.

Ключевые слова: сегментация медицинских изображений; пневмоторакс; рентгенография органов грудной клетки; клинический контур; валидация; системы поддержки принятия решений.

Финансирование: Работа выполнена в рамках реализации программы развития ФГБОУ ВО «Юго-Западный государственный университет» проекта «Приоритет-2030».

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

Для цитирования: Косинов А. В., Крупчатников Р. А., Серебровский В. В., Ефремов М. А., Кондрашов Д. С. Двухконтурная валидация нейросетевой сегментации пневмоторакса в условиях дисбаланса классов // Известия Юго-Западного государственного университета. Серия: Управление, вычислительная техника, информатика. Медицинское приборостроение. 2026. Т. 16, № 2. С. 122–149. <https://doi.org/10.21869/2223-1536-2026-16-4-122-149>.

Поступила в редакцию 11.04.2026

Подписана в печать 10.05.2026

Опубликована 30.06.2026

© Косинов А. В., Крупчатников Р. А., Серебровский В. В., Ефремов М. А., Кондрашов Д. С., 2026

Известия Юго-Западного государственного университета. Серия: Управление, вычислительная техника, информатика. Медицинское приборостроение. 2026;16(2):122–149

Two-loop validation of neural network-based pneumothorax segmentation under class imbalance

Aleksandr V. Kosinov¹✉, Roman A. Krupchatnikov¹, Vadim V. Serebrovsky¹,
Mihail A. Efremov¹, Dmitry S. Kondrashov¹

¹ Southwest State University
50 Let Oktyabrya Str. 94, Kursk 305040, Russian Federation

✉ e-mail: skosinka@yandex.ru

Abstract

The purpose of the research is to develop a clinically oriented methodology for training, validating, and selecting a neural network model for pneumothorax segmentation on chest radiographs under severe class imbalance and the need for controlled limitation of false positive alerts in clinical decision support systems.

Methods. We propose a two-loop segmentation quality assessment framework comprising a technical loop for monitoring training convergence using the loss function values and raw metrics, and a clinical loop for evaluating binarized masks after a fixed post-processing pipeline. For each training epoch, a binarization threshold is swept; an operating point is deemed acceptable if predefined clinical constraints are satisfied. The final epoch is then selected as the one that maximizes Dice index on positive cases among all acceptable solutions. The segmenter is implemented using a U-Net++ architecture and trained via a two-stage scheme with increasing spatial resolution, using a hybrid loss function.

Results. On clinically representative data SIIM-ACR development set, configuration selected by the protocol achieved Dice index on positive cases = 0,753 (95% CI 0,734–0,771) with false positive rate = 0,177 (95% CI 0,158–0,197) and image-level sensitivity = 0,948, and substantially outperformed the baseline configuration.

Conclusion. The proposed two-loop validation methodology and constraint-based epoch selection within the admissible region provide reproducible control of the trade-off between segmentation quality on pathological images and the false alarm rate on non-pathological images, reducing the risk of selecting a suboptimal model when relying solely on minimum loss. The results support the use of the developed protocol in designing segmentation modules for decision support systems in radiological imaging tasks.

Keywords: medical image segmentation; pneumothorax; chest x-ray; clinical circuit; validation; decision support systems.

Funding: The work was carried out within the framework of the Priority-2030 project development program of the Federal State Budgetary Educational Institution of Higher Education "Southwest State University."

Conflict of interest: The Authors declare the absence of obvious and potential conflicts of interest related to the publication of this article.

For citation: Kosinov A.V., Krupchatnikov R.A., Serebrovsky V.V., Efremov M.A., Kondrashov D.S. Two-loop validation of neural network-based pneumothorax segmentation under class imbalance. *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta. Seriya: Upravlenie, vychislitel'naya tekhnika, informatika. Meditsinskoe priborostroenie = Proceedings of the Southwest State University. Series: Control, Computer Engineering, Information Science. Medical Instruments Engineering*. 2026;16(2):122–149. (In Russ.) <https://doi.org/10.21869/2223-1536-2026-16-2-122-149>.

Received 11.04.2026

Accepted 10.05.2026

Published 30.06.2026

Введение

Рентгенография органов грудной клетки остаётся одним из наиболее часто выполняемых исследований и обычно является первым методом визуализации при симптомах со стороны лёгких благодаря доступности, скоро-

сти и низкой лучевой нагрузке [1]. Пневмоторакс – потенциально жизнеугрожающее состояние, требующее в некоторых случаях неотложного вмешательства, однако выявление пневмоторакса на рентгеновских снимках может быть затруднено даже для опытных

врачей, особенно в сложных случаях и при наложении анатомических структур и наличии медицинских устройств, способных маскировать признаки [2].

Системы поддержки принятия решений (СППР) на базе глубинного обучения демонстрируют высокий потенциал в задачах сегментации [3]. Однако при переходе от исследовательских задач к клинической практике возникает ряд методологических проблем:

1. Нечувствительность стандартных метрик. Коэффициент Дайса для бинарной сегментации является математическим эквивалентом F1-меры, что определяет его специфическое поведение при оценке несбалансированных данных [4]. При этом данная метрика в явном виде не учитывает истинно-отрицательные результаты, создавая слепые зоны при валидации моделей скрининга [5]. Следовательно, правильное оценивание эффективности систем требует явной фиксации определения метрики при использовании конкретных схем агрегаций, которые вводят вклад отрицательных случаев или фоновых классов косвенно. На практике в задачах с выраженным преобладанием отрицательных изображений стандартные подходы часто приводят к скрытым искажениям оценок сегментации [6]. Как показывают современные консенсусные рекомендации, при редкой патологии агрегация метрик по всей выборке становится серьезным источником методологических артефактов [4]. В частности, детальный анализ валидационных пайплайнов выявляет как минимум два распространенных сценария неадекватной оценки качества моделей [5]. Во-первых, во многих реализациях

рассчитывается средний индекс Дайса по снимкам, где для отрицательных случаев с пустой истинной маской и пустым предсказанием метрике присваивается значение индекса Дайса = 1 [5]. При большом числе отрицательных снимков при обучении среднее значение такой метрики может быть завышено и перестает отражать качество сегментации на патологических примерах. Простейший контрпример: рассмотрим три снимка, из которых два – с отсутствующей патологией и один – с патологией. Если на обоих «здоровых» модели предсказывает пустую маску, а на снимке с патологией ошибается, то при соглашении индекс Дайса на здоровых = 1 среднее значение составит 0,67, хотя клинически значимое качество работы модели на патологии неудовлетворительно, при том, что это самый важный параметр. Во-вторых, при использовании усреднения по классам в постановках, где явно присутствует фон как отдельный класс, в итоговую величину может входить индекс Дайса фоновых классов, который может оказаться близким к 1 на большинстве изображений. Такое усреднение также формирует иллюзию высокого качества, хотя ключевой интерес представляет сегментация патологического региона. Для устранения этих эффектов и методических неоднозначностей в настоящем исследовании предлагается подход с разделением метрик по подмножествам негативных и позитивных случаев, а также вводится понятие клинического контура оценки, позволяющего интерпретируемо контролировать компромисс между чувствительностью и специфичностью.

2. Конфликт чувствительности и специфичности. Агрессивная оптимизация сегментатора для поиска мелких очагов должна сопровождаться контролем ложноположительных срабатываний, поскольку в скрининговом режиме цена ложноположительных срабатываний выражается не только в метриках, но и в росте числа дообследований, перегрузке клинического контура и снижении доверия к СППР [7].

3. Критерий остановки обучения. Стандартная практика выбора рабочей точки по минимуму функции потерь на валидационном наборе часто оказывается несостоятельной в медицинских задачах [5]. Это обусловлено тем, что для преодоления дисбаланса классов применяются специализированные целевые функции, чья динамика расходится с поведением прямых метрик качества [8]. Например, сложные агрессивные функции потерь, такие как Focal или Tversky loss, формируют собственные траектории сходимости, слабо коррелирующие с коэффициентом Дайса на поздних этапах обучения [9]. Это делает классический критерий «минимум валидационных потерь» недостаточным для выбора итоговой модели, требуя введения внешних клинических ограничений для определения оптимальной рабочей точки [10].

Следует отметить, что задачи анализа рентгенограмм грудной клетки имеют серьезную отечественную исследовательскую базу: в частности, успешно применялись алгоритмы обнаружения патологий на основе анализа локальных окон [11]. Дополнительно были разработаны методы параметрического моделирования мор-

фологических образований непосредственно на диагностических снимках [12]. Эти классические подходы послужили фундаментом для создания комплексных математических моделей, интегрируемых в интеллектуальные системы анализа флюорограмм [13]. Классические методы сегментации биомедицинских изображений на основе градиентных операторов также продолжают использоваться как базовый инструментальный аппарат и историческая основа для последующих нейросетевых решений [14]. Отдельный интерес представляют патентные решения по автоматической сегментации флюорограмм грудной клетки, подтверждающие прикладную востребованность направления¹. В смежных задачах лучевой диагностики описаны автоматизированные системы классификации рентгенограмм (например, молочной железы), что дополнительно подтверждает актуальность встраивания алгоритмов в диагностические рабочие процессы [15].

Помимо узких задач лучевой диагностики, нейросетевые алгоритмы активно внедряются в архитектуру комплексных медицинских классификаторов риска [16]. Значительный потенциал в этой области демонстрируют гибридные модели, адаптированные для анализа кардиологических патологий [17]. Подобные системы позволяют прогнозировать состояние сердечно-

¹ Способ автоматической сегментации флюорограмм грудной клетки больных пневмонией: пат. 2629629 Рос. Федерация. № 2016132680; заявл. 09.08.16; опубл. 30.08.17, Бюл. № 25.

сосудистой системы на основе автоматизированной обработки многомерных биосигналов [18]. Параллельно развиваются методы построения предиктивных моделей медицинского риска с использованием сверточных нейронных сетей и синергетических каналов [19]. Отдельным важным направлением выступает применение мультимодальных нейросетевых классификаторов для оценки функционального состояния дыхательной системы [20]. Для задач инженерного прототипирования и воспроизводимого проектирования нейросетевых моделей используются специализированные программные средства интерактивного построения моделей нейронных сетей¹.

Практически значимым направлением являются многозадачные постановки (одновременная детекция, сегментация и классификация), активно используемые в прикладных медицинских сценариях [21].

Несмотря на то, что задачи классической сегментации флюорограмм широко представлены как в отечественной, так и в зарубежной литературе, вопросы интерпретируемой валидации в условиях дисбаланса остаются нераскрытыми.

Целью исследования является разработка методики клинически ориентированного обучения, валидации и выбора нейросетевой модели сегментации пневмоторакса на рентгенограммах органов грудной клетки при выраженном дисбалансе классов и необходимости

контролируемого ограничения ложных срабатываний в составе систем поддержки принятия решений.

Материалы и методы

Набор данных и предварительная обработка

Экспериментальное исследование выполнено на открытом наборе данных SIIM-ACR Pneumothorax Segmentation². Набор включает в себя рентгенограммы грудной клетки человека в формате DICOM (медицинский стандарт для радиограмм) и маски патологий в формате RLE.

Формирование выборки: для воспроизводимости результатов и предотвращения утечки данных был разработан и реализован иерархический протокол разделения данных. Разбиение выполнялось по идентификатору PatientID. В текущей выгрузке набора SIIM-ACR данный идентификатор является уникальным на каждый снимок, поэтому разбиение эквивалентно разбиению по изображениям без пересечения идентификаторов между обучением, валидацией и отложенной выборкой. Для компенсации ограничения применяются бутстреп-доверительные интервалы и одинарная финальная оценка на отложенной выборке после заморозки протокола.

Итоговая статистика разбиения представлена ниже (табл. 1)

¹ Программа для интерактивного создания модели нейронной сети (NetModel): а.с. 2023685902 Рос. Федерация. № 2023683638; заявл. 08.11.23; опубл. 30.11.23.

² SIIM-ACR Pneumothorax Segmentation // Kaggle. URL: <https://www.kaggle.com/competitions/siim-acr-pneumothorax-segmentation> (дата обращения: 18.03.2026).

Таблица 1. Статистика разбиения набора данных**Table 1.** Dataset split statistics

Параметр / Metric	Значение / Value	Описание / Description
Всего снимков	10675	Общее количество уникальных рентген снимков грудной клетки человека
Позитивные случаи	2379	Снимки с подтвержденным пневмотораксом
Отрицательные случаи	8296	Снимки без признаков патологии
Выборка для разработки	9074	Обучающая и валидационная выборка, 85% от общего объема
Отложенная выборка	1601	Независимый тестовый набор, 15% от общего объема
Доля положительных случаев в выборках	~22,3%	Целевой баланс классов в выборках

Процесс разделения состоял из двух этапов с сохранением стратификации по целевому классу, доля снимков с подтвержденным пневмотораксом стремилась к ~22,3% в каждой выборке.

Шаг 1: выделение отложенной тестовой выборки. Из общего массива данных была выделена независимая слепая выборка, составляющая 15% от общего числа снимков (всего в выборке 1601 снимок, 357 позитивных случаев, 1244 отрицательных случая). Данная выборка была полностью изолирована от процесса обучения, подбора гиперпараметров и выбора архитектуры. Она использовалась однократно после фиксации протокола выбора рабочей эпохи на наборе разработки исключительно для финальной оценки качества итоговой модели.

Шаг 2: оставшиеся 85% данных (9074 снимка) составили набор для разработки системы и подбора гиперпараметров. Для организации перекрестной проверки на этом наборе данных был применен алгоритм Stratified Group K-Fold при $k = 5$ с группировкой по PatientID. Поскольку в текущей выгруз-

ке PatientID уникален на снимок, данный метод обеспечивает равномерное распределение классов во всех K-Fold подвыборках без пересечения идентификаторов.

Разбиение фиксировалось в виде метаданных и статических файлов. Данный подход обеспечивает полную детерминированность экспериментов. Обучение, выбор рабочей точки, подбор гиперпараметров делались исключительно на наборе для разработки, без доступа к отложенной выборке.

Обработка пропусков: снимки без аннотированной патологии (здоровые пациенты) рассматриваются как имеющие «нулевую маску» через представление в нулевом тензоре. Исключение таких снимков из валидации недопустимо, так как они формируют базу для оценки специфичности.

Официальный полный тестовый набор SIIM-ACR stage 2 (3 205 снимков) в исследовании намеренно не использовался, поскольку для него отсутствуют публичные эталонные маски, что исключает возможность проверяемой количественной оценки сегментации. Для

финальной проверки была использована собственная отложенная выборка.

При обучении к изображениям применялись аугментации. Каждая трансформация включалась независимо с заданной вероятностью для каждого изображения. Такой режим реализует регуляризацию через контролируемое разнообразие входных данных без изменения целевой разметки. Важно, что геометрические искажения, такие как повороты, сдвиги и масштабирование, были отключены, поскольку в медицинской практике рентгеновские снимки обычно не имеют заметных вращений относительно центра или сдвигов, благодаря этому пространственная структура снимка сохранялась, а также отсутствие вращений и сдвигов ограничивает неестественные деформации анатомии, что особенно важно для рентгенограмм грудной клетки человека.

Ключевая геометрическая операция – горизонтальное зеркалирование. Это означает, что примерно в половине обучающих примеров происходит отражение снимка по горизонтали, что снижает риск переобучения на шаблоны датасета и повышает инвариантность модели к ориентации пациента.

Фотометрическая часть аугментаций моделирует вариабельность реальной рентген-аппаратуры: $\text{brightness_contrast} = 0,2$, $\text{gamma} = 0,1$, $\text{noise} = 0,1$. Эти операции имитируют различия экспозиции, контрастности, нелинейного тонового отклика и шумовых характеристик между сканерами и протоколами. Дополнительно применялось CLANE преобразование, но только на train с параметрами: $p = 0,5$, $\text{clip_limit} = 2,0$, $\text{tile_grid} = 8 \times 8$, что усиливает локальный контраст и стабилизирует визуализацию тонких границ.

После аугментаций выполнялась нормализация в пространстве ImageNet, обеспечивающая согласованную статистику входа для энкодера с предобученными весами.

Архитектура модели и стратегия обучения

В качестве базовой архитектуры сегментатора выбрана модель U-Net++ с энкодером ResNet-34, с весами, предобученными на ImageNet. Чтобы понять выбор данной архитектуры, необходимо обратиться к принципам классической сети U-Net. U-Net состоит из двух симметричных частей: энкодер и декодер. Кодировщик работает как традиционная сверточная сеть: он последовательно сжимает изображение, выделяя из него смысловые признаки и отвечает на вопрос: «Что находится на изображении?». Однако при таком сжатии теряется информация о точном местоположении объектов. Для решения этой проблемы используется декодер, который пошагово увеличивает сжатое представление обратно до размеров исходного изображения и отвечает на вопрос: «Где находится объект?». Ключевой особенностью U-Net являются так называемые прямые связи, по которым детализированная информация из левой части передается в правую, помогая сети точно очертить границы патологии [22].

Несмотря на эффективность базовой U-Net, в ней прямые связи соединяют слои жестко и прямолинейно. Предложенная в настоящей работе модификация U-Net++ решает эту проблему за счет внедрения плотной, паутинообразной сети вложенных соединений. Вместо простого копирования

признаков U-Net++ пропускает их через промежуточные сверточные блоки. На практике это означает, что сеть автоматически сглаживает семантический разрыв между грубыми и тонкими признаками. Для задачи поиска пневмоторакса это критически важно: патология может проявляться как крошечной полоской, так и обширной областью. U-Net++ естественным образом агрегирует признаки разных масштабов, что вкупе с механизмом глубокого контроля обеспечивает высокую чувствительность и стабильность обучения на малых выборках [23].

Для достижения высокой чувствительности была предложена стратегия двухэтапного обучения:

- стадия 1 (прогрев, стадия предварительного обучения): обучение на разрешении 512 на 512 пикселей. Используется стандартная функция потерь Focal + индекс Дайса. Цель – быстрая сходимость весов энкодера и декодера и формирование грубых карт признаков;

- стадия 2 (дообучение, точная подгонка весов на высоком разрешении): дообучение лучших весов первой стадии на разрешении 1024 на 1024 пикселей. Данный этап критически важен, так как некоторые из признаков пневмоторакса могут являться текстурными патологиями высокого разрешения, детали которых могут быть потеряны при уменьшении разрешения до 512 на 512 пикселей [24].

Стабилизация: производится заморозка статистик Batch Normalization в энкодере для стабильности обучения на малом размере батча с накоплением градиента до эффективного. Вычислительные ресурсы были ограничены, поэтому на разрешении 1024 на 1024 пик-

селей использовался батч равным 4, а эффективный батч с аккумуляцией градиента был равен 64.

В дальнейших экспериментах базовая модель, обученная на разрешении 512 на 512, обозначается как конфигурация C12, а итоговая модель после второго этапа дообучения на 1024 на 1024 – как C108.

Функция потерь при обучении

Функция потерь: для шага 2 была сконструирована гибридная функция потерь с компонентой Tversky loss, бинарной кросс-энтропии (BCE) и focal loss, для борьбы с дисбалансом между фоном и объектом интереса сверточной нейронной сети, а также для минимизации пропусков патологии [8].

В работе используются параметры функции потерь Tversky: $\alpha = 0,3$, $\beta = 0,7$. Гиперпараметры метрики расчёта Tversky позволяют асимметрично штрафовать ложноположительные и ложноотрицательные ошибки. Выбор $\beta > \alpha$ обеспечивает более высокий штраф за пропуск пикселей с патологией, по сравнению с ложноположительным срабатыванием [25]. Это переводит модель в разряд моделей с повышенной чувствительностью, данный режим предпочтительный в задачах СППР [26].

Однако агрессивная оптимизация чувствительности (смещение штрафов через параметры α и β) неизбежно приводит к генерации избыточного числа ложноположительных срабатываний (шума) на снимках без патологий. Чтобы этот процесс не вышел из-под контроля и не привел к деградации специфичности, в настоящей работе вводится жесткий фильтр – клинический контур оценки с формализованным ограниче-

нием на уровень ложных тревог, который выступает в роли системного предохранителя.

Методика двухконтурной валидации (RAW/CLIN)

Для объективной оценки качества модели вводится разделение метрик на два контура:

– RAW-контур (технический): оценка вероятностных карт сети без постобработки. Используется для мониторинга сходимости и сырых метрик качества обучения сверточной нейронной сети;

– клинический контур (прикладной): оценка бинаризованных масок после применения порогового отсечения и фильтрации по минимальной площади.

Определение метрик: пусть для изображения i задана истинная бинарная маска патологии $M_i \in \{0,1\}^{H \times W}$, предсказание нейросети на этом изображении представлено вероятностной картой $P_i \in [0,1]^{H \times W}$. Бинаризация выполняется по порогу t :

$$B_i(t) = 1(P_i \geq t),$$

где $1(\cdot)$ – индикаторная функция, применяемая поэлементно. Для получения маски, соответствующей прикладному использованию в составе СППР, вводится клинический контур (CLIN) как фиксированный оператор постобработки:

$$\hat{M}_i(t) = \text{CLIN}(B_i(t)).$$

Оператор $\text{CLIN}(\cdot)$ является неизменяемым во всех экспериментах настоящего исследования.

Как было показано ранее, агрегированные метрики на дисбалансных выборках искажают реальную картину качества. Для устранения этих эффектов

оценка проводится отдельно на подмножествах позитивных и негативных изображений. Позитивным считается изображение с непустой истинной маской: $|M_i| > 0$, негативным – изображение с пустой истинной маской $|M_i| = 0$, включая случаи отсутствия разметки RLE (run-length encoding), трактуемые как нулевая маска.

Ключевые метрики клинического контура: показатели семейства Dice/F1 широко применяются в современных работах по медицинской сегментации как базовые метрики качества, что требует аккуратной фиксации правил их расчета и агрегации [27].

Dice_{pos} – основная метрика качества сегментации в клиническом контуре – среднее значение коэффициента Дайса, рассчитываемое только на позитивных случаях, для защиты от завышения метрик через истинно отрицательные срабатывания.

Индекс Дайса на одном изображении определяется как

$$\text{Dice}(i, t) = \frac{2 \cdot |M_i \cap \hat{M}_i(t)|}{|M_i| + |\hat{M}_i(t)|}.$$

Dice_{pos} , усредняемый только по позитивным изображениям:

$$\text{Dice}_{\text{pos}}(t) = \frac{1}{|I_{\text{pos}}|} \cdot \sum_{i \in I_{\text{pos}}} \text{Dice}(i, t),$$

где $I_{\text{pos}} = \{i : |M_i| > 0\}$.

FP_Rate – частота ложноположительных случаев – риск ложных тревог на негативных изображениях, для которых модель предсказала наличие маски патологии. Снимки без патологии не исключаются из оценки, а используют

ся строго для контроля риска ложных тревог на негативном подмножестве:

$$FP_{Rate}(t) = \frac{1}{|I_{neg}|} \cdot \sum_{i \in I_{neg}} 1 \left(|\hat{M}_i(t)| > 0 \right),$$

где $I_{neg} = \{i : |M_i| = 0\}$.

Pos_any — доля положительных снимков, на которых предсказанная маска в клиническом контуре не пуста. Метрика используется как индикатор покрытия / обнаружения и трактуется как чувствительность модели на уровне изображения:

$$pos_{any}(t) = \frac{1}{|I_{pos}|} \cdot \sum_{i \in I_{pos}} 1 \left(|\hat{M}_i(t)| > 0 \right).$$

Pos_any можно трактовать как чувствительность (sensitivity) модели.

Таким образом, $Dice_pos$ характеризует качество локализации / сегментации на патологических снимках, FP_Rate — частоту ложных тревог на снимках без патологии, а Pos_any — устойчивость обнаружения патологии на уровне изображения. Эти метрики совместно задают интерпретируемый контроль компромисса чувствительность — ложные тревоги при дисбалансе классов.

Поскольку в задачах сегментации на дисбалансных данных минимум функции потерь на валидации может не совпадать с клинически предпочтительным качеством (эффект расхождения функции потерь и прикладных метрик), в настоящей работе применяется формализованный протокол выбора рабочей точки и эпохи обучения, основанный исключительно на клиническом контуре.

Предрегистрация сетки порогов. Для исключения подбора порога бинаризации после начала эксперимента ис-

пользуется заранее фиксированная сетка порогов клинического контура:

$$BASE_{thr} = \{0,001; 0,002; 0,005; 0,01; 0,02; 0,03; 0,05; 0,10; 0,15; 0,20; 0,30; 0,50\}.$$

В ходе экспериментов допускается оценивание модели только в точках $t \in base_{thr}$.

Клинические ограничения: для каждой эпохи обучения e на наборе разработки рассматриваются значения метрик клинического контура при переборе порога $t \in base_{thr}$. Рабочая точка считается допустимой если одновременно выполняются ограничения:

$$FP_{Rate}(e, t) \leq fp_{target}, \text{ и } pos_{any}(e, t) \geq min_{pos_any},$$

где $FP_{target} = 0,30$ и $Min_{pos_any} = 0,20$.

В качестве клинических метрик также использовались значения $pr_min_area = 300$ пикселей. Клинический контур также пересчитывался после изменения разрешения изображений в случаях, где применялось изменение разрешения.

Выполняется запрет подгонки на отложенной выборке. Отложенная выборка используется после выбора рабочих точек и заморозки всех параметров, основываясь на данных, полученных во время работы с валидационной выборкой.

Процедура выбора эпохи

Учитывая эффект расхождения строгой математической функции потерь и реальных метрик качества модели, вместо традиционного выбора модели по минимуму потерь при обучении предложен следующий алгоритм ограниченного выбора: задаются клинические ограничения и на каждой эпохе ва-

лидации производится поиск порога бинаризации, удовлетворяющего ограничениям.

Пусть E – множество эпох обучения, а e – конкретная эпоха. Для каждой эпохи $e \in E$ на валидационной выборке определяются метрики. Вычисляется область допустимых решений, определяемая клинической политикой: финальная модель выбирается путем максимизации качества сегментации внутри допустимой области. Эпоха считается

$$\mathcal{F}(e) = \{t \in \text{base}_{\text{thr}} : \text{FP}_{\text{Rate}}(e, t) \leq \text{fp}_{\text{target}} \wedge \text{pos}_{\text{any}}(e, t) \geq \min_{\text{pos}_{\text{any}}}\},$$

где $\text{FP}_{\text{target}} = 0,30$, $\text{Min}_{\text{pos}_{\text{any}}} = 0,20$, FP_{Rate} определяется строго на подмножестве отрицательных снимков, а Pos_{any} – как доля положительных снимков с непустой предсказанной маской после клинического контура.

Клинические ограничения, используемые в клиническом контуре, задаются не как окончательная эксплуатационная спецификация для скрининговой системы, а как внутренний критерий выполнимости для этапа разработки и методической валидации сегментационного модуля. В частности, порог $\text{FP}_{\text{target}} = 0,30$ применяется как верхняя граница допустимого уровня ложных тревог на отрицательных снимках в контуре подбора для предотвращения деградации специфичности при целенаправленной оптимизации чувствительности сегментатора. Такая постановка отражает целевую роль сегментатора в данной статье: получить высокочувствительный модуль локализации, устойчиво сохраняющий обнаружение патологии при контролируемом росте ложноположительных срабатываний, но без претензии на достижение конечного уровня ложноположительных срабаты-

ся допустимой, если такое решение существует.

В соответствии с предрегистрацией протокола оценивания выбор итоговой модели выполняется исключительно по клиническому контуру и определяется правилом первичной конечной точки. Для каждой эпохи e и каждого порога бинаризации рассчитываются метрики клинического контура, после чего формируется множество допустимых решений:

ваний, требуемого в реальном клиническом рабочем процессе. Выбор относительно мягкого ограничения на этапе разработки мотивирован архитектурной декомпозицией задачи на несколько исследований. В данном исследовании фокусируется методика обучения и клинически ориентированного выбора эпохи / рабочей точки для сегментатора при сильном дисбалансе классов и эффекте расхождения метрик. При этом жесткое снижение уровня ложноположительных срабатываний до уровня, сопоставимого с эксплуатационными требованиями к скринингу, рассматривается как задача системной архитектуры и решается в последующих исследованиях. Таким образом, ограничения $\text{FP}_{\text{target}} = 0,30$, $\text{Min}_{\text{pos}_{\text{any}}} = 0,20$ следует трактовать как операционное ограничение разработки.

Рабочая точка для итоговой эпохи определяется по правилу первичной конечной точки:

$$e^* = \arg \max_{\text{pos}} \text{Dice}_{\text{pos}}(e, t^*(e)).$$

Таким образом, в рамках выбора лучшей рабочей точки показатель FP_{Rate} используется только как огра-

ничество выполнимости, а не как оптимизируемая цель. Любые дополнительные характеристики, например динамика FP_Rate по эпохам или анализ локальных минимумов FP_Rate, рассматриваются как вторичные результаты и не изменяют предопределённого правила выбора лучших рабочих точек.

Данный подход демонстрирует, что выбранная модель удовлетворяет требованиям безопасности по уровню ложноположительных срабатываний и эффективности, игнорируя локальные минимумы функции потерь, имеющие низкую клиническую ценность.

При финальной проверке на отложенной выборке порог бинаризации не подбирался повторно и не пересканировался. Использовалось значение порога, выбранное на наборе разработки по предопределённому правилу, а на отложенной выборке выполнялась только однократная оценка метрик в фиксированной рабочей точке и для выбранной эпохи. Это исключает использование отложенной выборки для подбора параметров и сохраняет корректность схемы однократной проверки.

Статистический анализ

Для количественной оценки неопределённости метрик клинического контура использовались 95% доверительные интервалы, полученные методом бутстреп-ресэмплинга. Единицей наблюдения в статистическом анализе являлось изображение (случай), а не отдельные пиксели. Для каждой сравниваемой конфигурации модели формировался массив наблюдений размера N, включающий число изображений в соответствующей подвыборке. Далее выполнялось 1000 бутстреп-итераций. На каждой итерации строился бутстреп-

сэмпл путем выборки N строк из таблицы статистики эпохи. Ресэмплирование проводилось совместно по всем строкам. С учётом того что метрики определены на различных подмножествах данных (Dice_pos и Pos_any – на позитивных случаях, FP_Rate – на негативных), бутстреп выполнялся отдельно без стратификации по классу. На каждой итерации формировалась выборка с возвращением того же размера, что и исходное соответствующее подмножество, после чего метрика пересчитывалась по формальным определениям клинического контура. Границы 95% доверительного интервала определялись по 2,5- и 97,5-перцентилям эмпирического распределения бутстреп-оценок, без явной стратификации по позитивным и негативным снимкам. Это корректно при больших объёмах данных и устойчивой представленности классов, однако при малых выборках может приводить к повышенной вариативности позитивных метрик.

Результаты и их обсуждение

Сравнительный анализ эффективности базового и предложенного подходов

Для оценки эффективности предложенной методики выполнено сравнение клинических метрик сегментации на наборе разработки для двух конфигураций: базовой модели C12 (обучение 512 на 512 пикселей) и предложенной модели C108 (1024 на 1024 пикселей – второй этап с последующей тонкой настройкой). Метрики рассчитывались в клиническом контуре и в рабочей точке, выбранной на наборе разработки согласно предопределённому протоколу. Рассмотрим сводные результаты (табл. 2).

Таблица 2. Клинические метрики на наборе разработки для сравниваемых конфигураций**Table 2.** Clinical metrics on the development set for the compared configurations

Модель / Конфигурация / Model / Configuration	Dice_pos, среднее [95% ДИ] / Dice_pos, mean (95% CI)	Pos_any, среднее [95% ДИ] / Pos_any, mean (95% CI)	FP_Rate, среднее [95% ДИ] / FP_Rate, mean (95% CI)
C12	0,488 [0,466–0,509]	0,876 [0,845–0,908]	0,464 [0,440–0,491]
C108, эпоха 8, минимум функции потерь	0,569 [0,547–0,590]	0,983 [0,969–0,995]	0,424 [0,399–0,452]
C108, эпоха 12	0,699 [0,679–0,718]	0,958 [0,937–0,978]	0,253 [0,233–0,277]
C108, эпоха 17	0,721 [0,700–0,740]	0,962 [0,944–0,981]	0,224 [0,202–0,247]
C108, эпоха 15, вы- бранная согласно про- токолу	0,753 [0,734–0,771]	0,948 [0,905–0,956]	0,177 [0,158–0,197]

Для показателей в таблице 2 приведены 95% ДИ, полученные методом бутстреп-перевыборки.

Сравнение выполнялось строго в терминах клинического контура и в соответствии с предопределённым правилом первичной конечной точки: для каждой эпохи на наборе разработки выбиралась рабочая точка, удовлетворяющая клиническим ограничениям, после чего итоговая эпоха определялась как максимизирующая Dice_pos среди допустимых решений. В рамках этого протокола для конфигурации C108 наилучшее значение Dice_pos достигается на 15-й эпохе – 0,753 [0,734–0,771] при FP_Rate 0,177 [0,158–0,197] и Pos_any 0,948, поэтому именно эта эпоха рассматривается как выбранная модель на уровне разработки.

Базовая конфигурация C12 демонстрирует Dice_pos 0,488 [0,466–0,509] при FP_Rate 0,464 [0,440–0,491] и Pos_any 0,876. Высокое значение FP_Rate означает, что значимая доля негативных снимков сопровождается ложной тревогой в клиническом контуре, что делает конфигурацию C12

нежелательной в составе СППР без дополнительного механизма подавления ложноположительных срабатываний на уровне системы. Кроме того, никакая из эпох конфигурации C12 не смогла пройти порог клинического уровня.

Промежуточные точки обучения C108 (эпохи 12 и 17) приведены как вторичный анализ динамики компромисса между качеством сегментации и ложными тревогами. Так, на 17-й эпохе наблюдается более высокое покрытие Pos_any (0,962) при Dice_pos 0,721 [0,700–0,740] и FP_Rate 0,224 [0,202–0,247], тогда как 15-я эпоха обеспечивает максимальный Dice_pos при несколько меньшем Pos_any и более низком FP_Rate. Эти различия иллюстрируют, что выбор лучшей точки зависит от заранее зафиксированного критерия, а не постфактум предпочтением одной из метрик.

По сравнению с базовой конфигурацией C12 предложенная конфигурация C108 (выбранная по протоколу на 15-й эпохе) демонстрирует следующие изменения на наборе для разработки:

1) рост качества сегментации на позитивных случаях: среднее значение коэффициента Dice_pos увеличивается с 0,488 до 0,753, абсолютная разница составляет $\Delta = 0,265$. Относительный прирост составил 54,3% относительно C12;

2) высокое покрытие обнаружения на уровне снимка: метрика Pos_anu возрастает с 0,876 до 0,948 по сравнению с C12 $\Delta = 0,072$ или +8,2%. Модель детектирует 94,8% случаев патологии (против 87,7% у C12), что указывает на сохранение высокого уровня обнаружения патологии при выбранной рабочей точке;

3) снижение ложных тревог на отрицательных снимках: доля ложнопозитивных результатов снизилась более чем в 2,62 раза – с 0,464 до 0,177, на 61,9%, абсолютное снижение составляет $\Delta = 0,287$.

Абляционное сравнение критериев выбора модели: ключевым элементом анализа является доказательство эффективности предложенного алгоритма выбора эпохи. Для этого в таблицу 2 включена конфигурация «C108, эпоха 8» –

это та же самая нейронная сеть, но зафиксированная по стандартному критерию ранней остановки в точке абсолютного минимума валидационной функции потерь. Сравнение показывает, что выбор эпохи по val_loss (эпоха 8) проигрывает выбору по клиническому алгоритму (эпоха 15) в качестве локализации и, что более критично, демонстрирует недопустимо высокий уровень ложных тревог. Эта абляция внутри одной архитектуры показывает: существенный прирост качества обусловлен не только переходом на высокое разрешение, но и применением разработанного двухконтурного протокола валидации, который уберег модель от фиксации в клинически субоптимальной точке.

Для наглядной демонстрации того, как алгоритм определяет рабочую точку внутри выбранной эпохи, на рисунке 1 представлено поведение метрик клинического контура для итоговой конфигурации C108 (эпоха 15) при переборе порогов бинаризации.

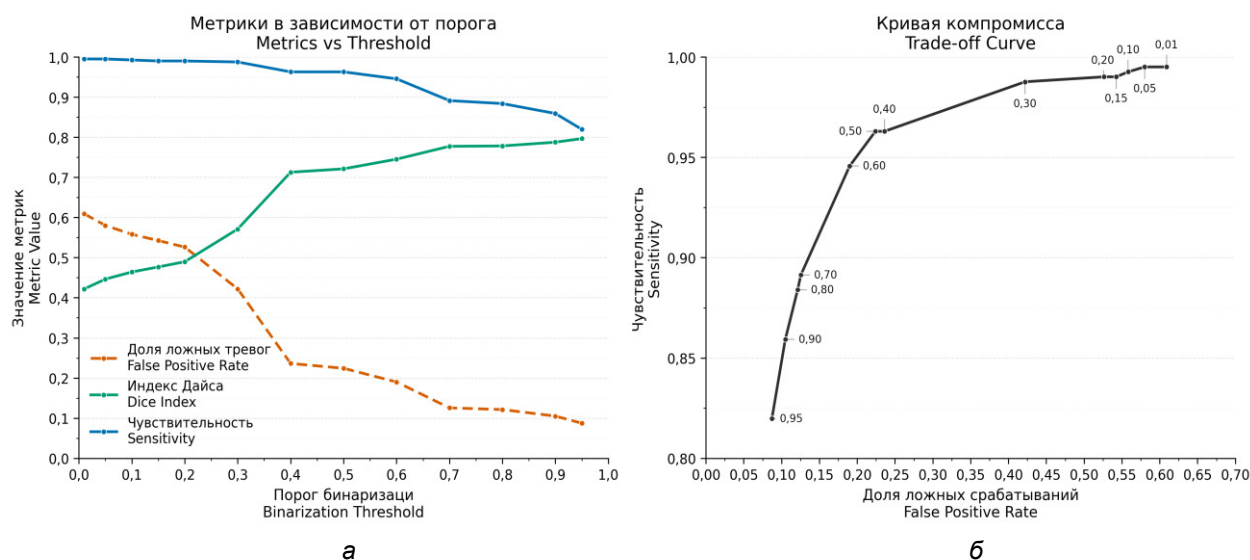


Рис. 1. Анализ метрик клинического контура для конфигурации C108 (эпоха 15): а – динамика метрик в зависимости от порога бинаризации; б – кривая компромисса «чувствительность – доля ложных срабатываний» с указанием значений порогов на точках

Fig. 1. Analysis of clinical contour metrics for the C108 configuration (epoch 15): а – metrics dynamics depending on the binarization threshold; б – "sensitivity – false positive rate" trade-off curve

Левый график иллюстрирует разнонаправленную динамику: по мере увеличения порога доля ложных срабатываний (оранжевая пунктирная линия) стремительно падает, индекс Дайса (зеленая линия) растет и стабилизируется, в то время как чувствительность (синяя линия) демонстрирует пологую деградацию.

Правый график параметрической кривой компромисса наглядно объясняет результат работы алгоритма: рабочая точка для порога $t = 0,50$ располагается на выраженном изгибе кривой. В этой точке достигается оптимальный баланс: модель выдерживает клиническое ограничение, сохраняя при этом высокую чувствительность обнаружения. Выбор более низких порогов дает минимальный прирост к чувствительности, но приводит к резкому, непропорциональному росту числа ложных тревог.

В совокупности результаты на наборе разработки показывают, что предложенная методика обучения и клинически ориентированного выбора

рабочей точки обеспечивает более благоприятный баланс Dice_pos и FP_Rate по сравнению с базовым протоколом. Интерпретация наблюдаемого улучшения как эффекта увеличения разрешения и изменения функции потерь соответствует исходной инженерной гипотезе, однако причинный вклад отдельных компонентов (разрешение, состав функции потерь, правила клинического контура) требует отдельного абляционного исследования и рассматривается как предмет последующего анализа.

Проверка на отложенной выборке и K-Fold

Для оценки стабильности предложенной архитектуры и устойчивости алгоритма выбора рабочей точки была проведена процедура перекрестной проверки (Stratified Group K-Fold при $K = 5$) на наборе для разработки. Детализированные метрики клинического контура по каждому из разбиений представлены ниже (табл. 3).

Таблица 3. Результаты 5-кратной перекрестной проверки (Stratified Group K-Fold) для предложенной модели C108

Table 3. Results of 5-fold cross-validation (Stratified Group K-Fold) for the proposed C108 model

Подвыборка / Subset	Объем разбиения / Split size	Dice_pos, среднее [95% ДИ] / Dice_pos, mean (95% CI)	FP_Rate среднее [95% ДИ] / FP_Rate, mean (95% CI)	Pos_any среднее [95% ДИ] / Pos_any, mean (95% CI)
Фолд 0	405 / 1410	0,753 [0,734–0,771]	0,177 [0,158–0,197]	0,931 [0,905–0,956]
Фолд 1	405 / 1410	0,801 [0,785–0,817]	0,157 [0,139–0,177]	0,975 [0,960–0,988]
Фолд 2	404 / 1411	0,815 [0,799–0,830]	0,150 [0,133–0,169]	0,978 [0,964–0,991]
Фолд 3	404 / 1411	0,792 [0,776–0,808]	0,174 [0,154–0,194]	0,975 [0,960–0,989]
Фолд 4	404 / 1410	0,804 [0,786–0,820]	0,165 [0,145–0,184]	0,975 [0,960–0,990]
В среднем (стандартное отклонение, SD)	404 / 1410	0,793 $\pm$ 0,023	0,165 $\pm$ 0,011	0,967 $\pm$ 0,020

Результаты экспериментов подтвердили высокую устойчивость модели к изменению обучающей и валидационной подвыборок. Усредненные значения метрик по всем пяти фолдам составили: $\text{Dice_pos} = 0,793 \pm 0,023$, $\text{FP_Rate} = 0,165 \pm 0,011$ и $\text{Pos_any} = 0,967 \pm 0,020$. Во всех без исключения сплитах выбранные алгоритмом рабочие точки уверенно выполнили заданные клинические ограничения. Примечательно, что показатели нулевого фолда, где $\text{Dice_pos} = 0,753$, подробно проанализированные в сравнительном анализе, являются наиболее консервативной оценкой. На остальных четырех подвыборках качество локализации патологии варьировалось в диапазоне 0,792-0,815 при снижении доли ложных

срабатываний вплоть до 0,150. Низкая дисперсия метрик между фолдами показывает, что предложенный двухконтурный протокол валидации обеспечивает надежный и воспроизводимый результат, не зависящий от случайного разбиения данных. В таблице 4 показаны данные по тестированию на отложенной выборке.

Финальная проверка выбранной по протоколу конфигурации C108 выполнена на независимой отложенной выборке, которая не использовалась для обучения, подбора гиперпараметров, выбора эпохи или выбора рабочей точки. На отложенной выборке применялись те же параметры клинического контура и та же рабочая точка, которые были зафиксированы на наборе разработки.

Таблица 4. Результаты тестирования на отложенной выборке

Table 4. Results on the holdout test set

Показатель / Metric	Среднее значение / Mean	95% ДИ / 95% CI
Dice_pos	0,737	[0,718–0,755]
FP_Rate	0,203	[0,183–0,224]
Pos_any	0,924	[0,895–0,951]

Дополнительно для интерпретации метрик на уровне изображений: при N позитивных = 357 и N негативных = 1244 получено истинно положительных = 330 и ложно негативных = 27, $\text{Pos_any} = 0,924$, ложноположительных = 252 и истинно отрицательных = 992 – метрика $\text{FP_Rate} = 0,203$. Таким образом, при переносе на отложенную выборку модель сохраняет высокий уровень обнаружения патологии на уровне снимка и контролируруемую частоту ложных тревог, оставаясь в допустимой области, заданной ограничениями протокола.

Наблюдаемое отличие показателей, по сравнению с набором разработки, ожидаемо для независимой проверки и

не меняет вывод о воспроизводимости протокола выбора модели.

Анализ динамики обучения и феномен расхождения метрик

На рисунке 2 представлена динамика функции потерь и ключевых метрик модели для конфигурации C108 в процессе дообучения на разрешении 1024 на 1024 пикселя. График иллюстрирует выявленный в ходе эксперимента эффект расхождения значений функции потерь и целевых клинических метрик. Минимум значения функции потерь на проверочной выборке не совпадает с эпохой, обеспечивающей наилучшее клинически интерпретируемое качество по заранее заданному правилу выбора.

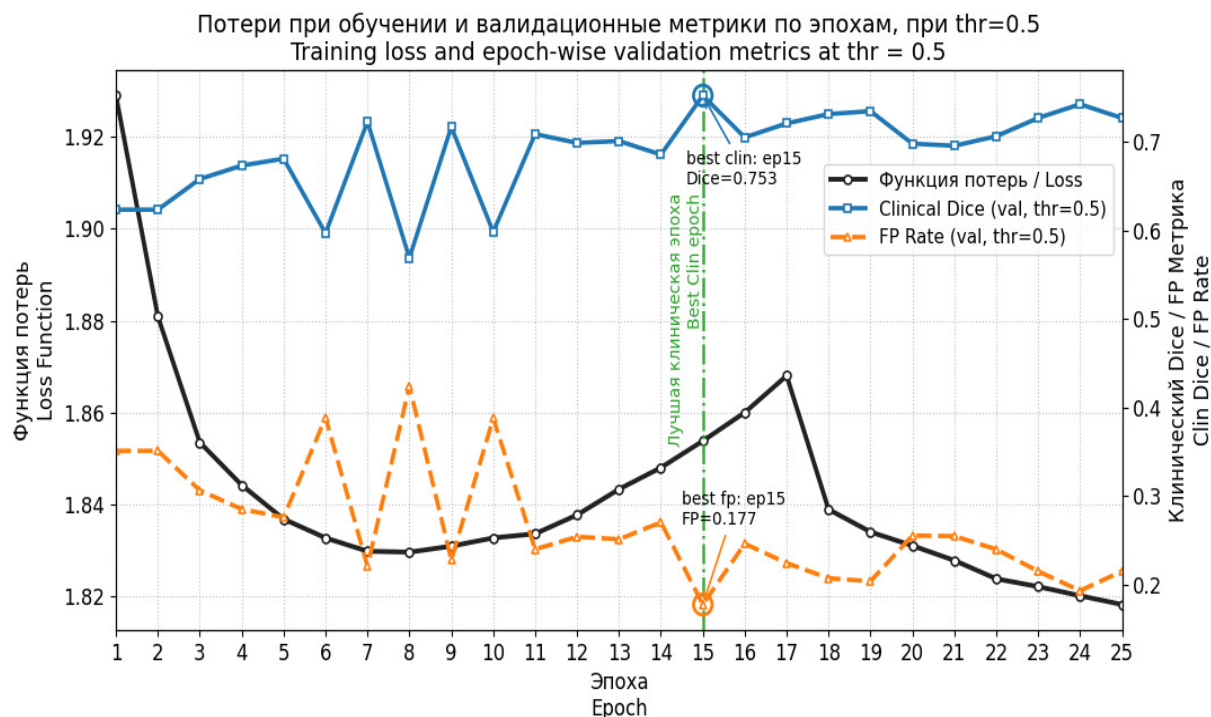


Рис. 2. Динамика функции потерь и клинических метрик для C108 на наборе разработки в ходе дообучения

Fig. 2. Dynamics of the loss function and clinical metrics for the C108 configuration on the development set during fine-tuning

Динамика (см. рис. 2) условно разделяется на две фазы. В первой фазе (ранние эпохи) уменьшение функции потерь сопровождается согласованным улучшением клинических метрик. Во второй фазе наблюдается расхождение: при отсутствии дальнейшего улучшения (или даже при росте) значения функции потерь клинические метрики продолжают улучшаться или стабилизируются на более благоприятном уровне, что делает выбор сохранённой модели по минимуму функции потерь потенциально субоптимальным для клинического контура. Графики построены в рабочей точке по порогу $\text{thr} = 0.5$, выбранному по предрегистрации:

1. Фаза согласованной оптимизации (эпохи 1–8): наблюдается совместная динамика — уменьшение значения функции потерь сопровождается ростом Dice_{pos} и снижением FP_Rate . Ло-

кальный минимум функции потерь достигается примерно на 8-й эпохе.

2. Фаза расхождения (эпохи 9–25): после достижения минимума функция потерь на проверочной выборке не улучшается, однако клинические метрики демонстрируют сохранение или улучшение. В частности, Dice_{pos} достигает максимума на 15-й эпохе, (см. рис. 2). Это указывает, что минимум функции потерь не является надёжным критерием выбора клинически предпочтительной рабочей точки при заданной постановке задачи и выбранных компонентах функции потерь.

Практическое следствие показанного расхождения состоит в том, что стандартная ранняя остановка обучения, ориентированная на минимум функции потерь на проверочной выборке, выбрала бы модель вблизи эпохи локального ми-

нимума потерь и тем самым могла бы зафиксировать модель, уступающую выбранной по клиническому правилу. В настоящей работе лучшая эпоха определяется предрегистрацией: выбирается эпоха, максимизирующая $Dice_pos$ в допустимой области, заданной ограничениями на FP_Rate и Pos_any . Следовательно, траектории функции потерь и вторичных метрик интерпретируются как описательные и не изменяют правила выбора модели. Наблюдения (см. рис. 2) поддерживают тезис о том, что при выраженном дисбалансе и гибридных функциях потерь минимизация математической ошибки не гарантирует опти-

муму по клинически интерпретируемым метрикам.

Качественный анализ результатов сегментации

Для интерпретации количественных различий метрик выполнен визуальный анализ предсказаний на репрезентативной выборке случаев из набора разработки, включающей как позитивные снимки с различной выраженностью пневмоторакса, так и негативные снимки с типичными источниками ложных срабатываний (костные структуры, трубки, артефакты). Рассмотрим примеры сравнений для C12 и C108 (рис. 3–5).

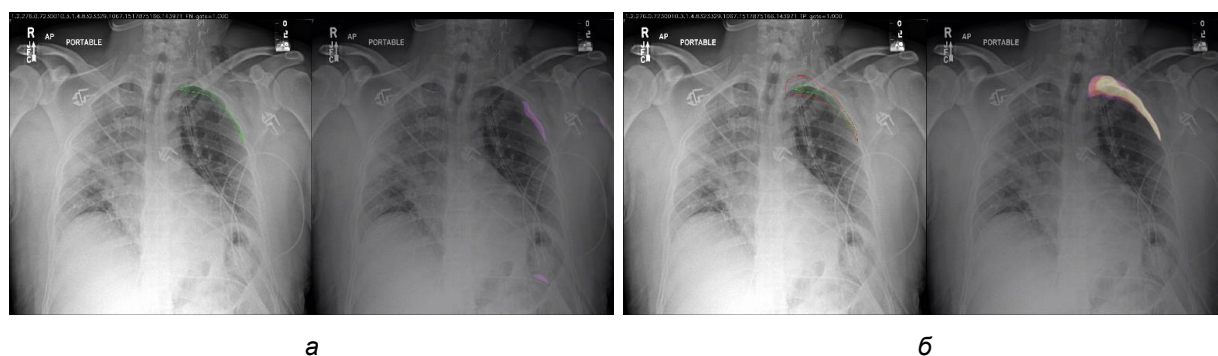


Рис. 3. Пример позитивного случая с малым объёмом пневмоторакса: эталонные данные и предсказания C12 (а) и C108 (б) в клиническом контуре

Fig. 3. Example of a positive case with a small pneumothorax volume: ground truth and clinical-loop predictions of C12 (a) and C108 (b)

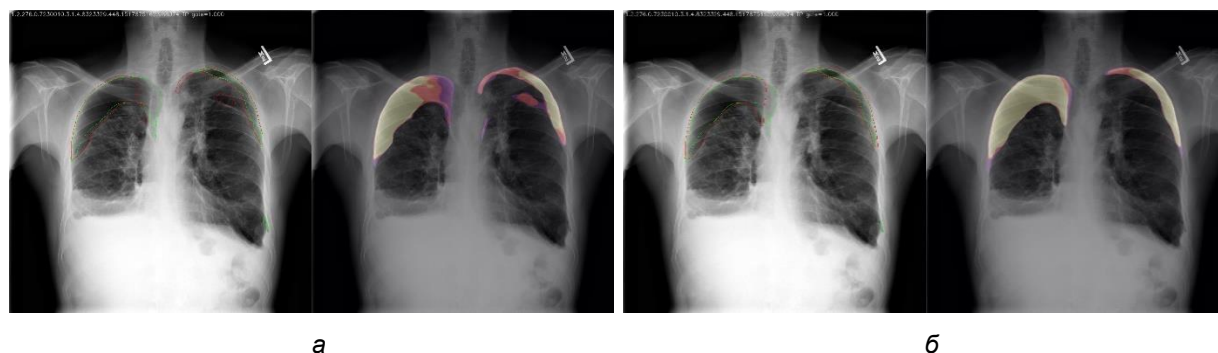


Рис. 4. Пример позитивного случая с распространённым пневмотораксом: сравнение топологии масок, фрагментация предсказания и связанная область выражены лучше: а – C12; б – C108

Fig. 4. Example of a positive case with extensive pneumothorax: comparison of mask topology; prediction fragmentation is reduced and the connected component is represented more clearly: а – C12; б – C108

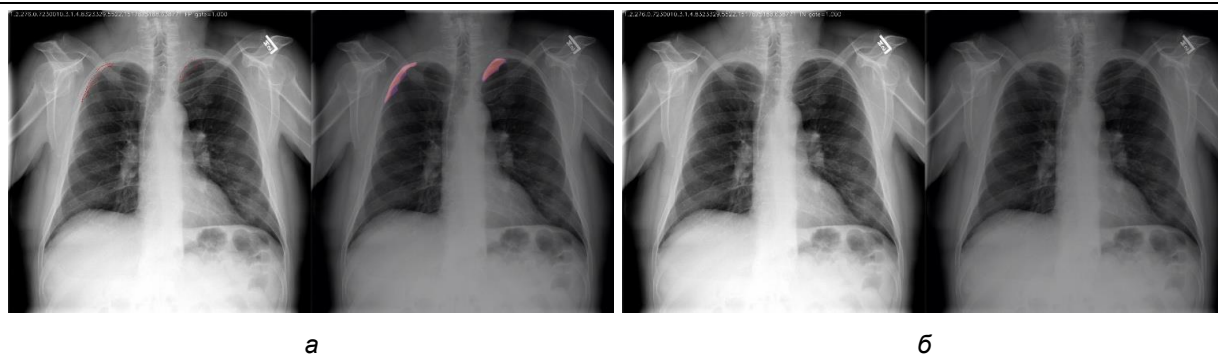


Рис. 5. Пример отрицательного случая: ложные срабатывания C12 и подавление ложной маски в C108: а – C12; б – C108

Fig. 5. Example of a negative case: false positives in C12 and suppression of the spurious mask in C108: а – C12; б – C108

Визуальный анализ выявляет три типовых различия между базовой и предложенной конфигурациями, согласующиеся с наблюдаемыми изменениями Dice_pos и FP_Rate в таблице 2:

1) слабый локальный сигнал на положительных снимках (см. рис. 3). В ряде случаев пневмоторакс проявляется как тонкая плевральная линия или локальное ослабление лёгочного рисунка. В таких примерах C12 чаще демонстрирует неполное покрытие или пропуск области, тогда как C108 формирует более устойчивую локализацию и более связную маску в зоне патологии;

2) топологическая целостность маски при выраженной патологии (см. рис. 4). На некоторых положительных снимках C12 склонна к фрагментации предсказания на отдельные «островки», что затрудняет интерпретацию локализации. В C108 в аналогичных примерах чаще наблюдается формирование связной области предсказания, согласующейся с контуром поражения;

3) структурно обусловленные ложные срабатывания на отрицательных снимках (см. рис. 5). На отрицательных случаях базовая модель чаще реагирует на контрастные костные структуры и

артефакты, формируя мелкие ложные области. В предложенной конфигурации ложные области, как правило, имеют меньшую выраженность / площадь и реже проходят клинический контур, что согласуется со снижением FP_Rate (см. табл. 2).

Полученные результаты показывают, что клинически ориентированный протокол валидации и выбора модели позволяет улучшить качество сегментации на положительных случаях при одновременном контроле ложных тревог на отрицательных снимках. В соответствии с заранее зафиксированным правилом первичной конечной точки выбранная конфигурация C108 (15-я эпоха) достигает Dice_pos 0,753 при FP_Rate 0,177 и Pos_any 0,948 на наборе разработки, тогда как базовая конфигурация C12 демонстрирует Dice_pos 0,488 при FP_Rate 0,464 и Pos_any 0,876. Важным является не только само улучшение показателей, но и то, что оно получено в рамках фиксированных клинических ограничений, что снижает риск неконтролируемого смещения модели в сторону высокой чувствительности ценой взрыва ложноположительных срабатываний.

С методической точки зрения результаты поддерживают два вывода. Во-первых, разнесение метрик по подмножествам позитивных и негативных изображений даёт интерпретируемую картину качества и устраняет неоднозначности, возникающие при использовании агрегированных показателей на дисбалансных данных. Во-вторых, демонстрируемое расхождение между динамикой функции потерь и клинических метрик подтверждает необходимость выбора эпохи не по минимуму функции потерь, а по клиническому правилу внутри допустимой области.

Предложенная двухконтурная схема оценки позволяет разделить две задачи, которые на практике часто смешиваются. Технический контур используется для контроля сходимости и мониторинга обучения, тогда как клинический контур задаёт прикладной смысл метрик – оценку бинаризованных масок после процедур, которые соответствуют будущему применению модуля в составе СППР.

Ключевой особенностью является то, что «лучшая модель» определяется не по локальным экстремумам функции потерь, а по первичной конечной точке: максимум Dice_pos в допустимой области, заданной ограничениями на FP_Rate и Pos_any. Такой подход принципиально повышает воспроизводимость процесса выбора модели: правило выбора фиксировано и не зависит от постфактум предпочтений исследователя, а значит, снижает риск непреднамеренной подгонки результатов под ожидаемый эффект.

Следует подчеркнуть, что введённое ограничение FP_Rate в данной работе трактуется как ограничение вы-

полноты на этапе разработки, а не как эксплуатационная спецификация конечной клинической системы. Роль сегментатора в текущем исследовании – сформировать высокочувствительный модуль локализации, сохраняющий обнаружение патологии при контролируемом уровне ложных тревог. В этом режиме допустимо наличие остаточных ложноположительных срабатываний на отрицательных снимках, поскольку дальнейшее снижение FP рассматривается как задача системного согласования модулей на уровне СППР.

Тем самым ограничение FP_Rate используется как предохранитель от деградации специфичности при целенаправленной оптимизации чувствительности, но не как цель оптимизации. Такой дизайн согласуется с архитектурной декомпозицией: в последующих исследованиях предполагается введение внешнего механизма подавления ложных тревог, который позволит приблизить уровень ложных тревог к требованиям практического применения без разрушения чувствительности сегментатора.

Оценка выполнена на одном открытом наборе данных. Кроме того, в используемой выгрузке идентификатор PatientID является уникальным для одного снимка, из-за чего разбиение по PatientID эквивалентно разбиению на уровне изображений. Это ограничивает строгость вывода о переносимости на новых пациентов и повышает требования к дополнительным проверкам устойчивости. Кроме того, вычислительные ресурсы были ограничены, что не позволило обучать нейронную сеть с размером батч больше 4 на разрешении 1024 на 1024 пикселя, несмотря на ис-

пользование аккумуляции градиента между батчами, итоговый градиент на более мощных системах может отличаться.

В рамках текущей версии исследования не проводилась внешняя валидация, могут потребоваться проверки переносимости.

Несмотря на то, что улучшение качества согласуется с гипотезой о пользе повышения разрешения и модификации функции потерь, причинный вклад отдельных компонентов (разрешение, состав функции потерь, процедуры клинического контура) в данной версии не изолирован строгими абляциями. Следовательно, интерпретация носит инженерный характер и должна быть подтверждена отдельными сравнительными экспериментами.

Открытые наборы данных потенциально содержат шум разметки и неполноту аннотаций, включая случаи пропущенной патологии. Это может влиять на оценку, поскольку часть ложных срабатываний может соответствовать неразмеченным истинным проявлениям. В дальнейшем целесообразно формализовать и задокументировать процедуру аудита разметки и включить её результаты как ограничение валидности выводов.

Дальнейшая работа естественным образом следует из архитектурной постановки задачи. Во-первых, требуется системный механизм подавления ложных тревог на уровне СППР, позволяющий снизить до эксплуатационно приемлемого уровня при сохранении покрытия обнаружения. Во-вторых, необходимы абляционные эксперименты, изолирующие вклад повышения разрешения, состава функции потерь и пра-

вил клинического контура. В-третьих, требуется проверка переносимости на внешнем домене и анализ устойчивости метрик при изменении распределения данных.

Выводы

В настоящем исследовании решена актуальная методическая проблема объективной оценки и выбора нейросетевых моделей сегментации медицинских изображений в условиях выраженного дисбаланса классов. Предложенная двухконтурная схема валидации, разделяющая технический мониторинг обучения (RAW-контур) и оценку прикладной эффективности (CLIN-контур), позволяет преодолеть фундаментальное ограничение стандартных метрик – иллюзию высокого качества при агрегировании результатов по всей выборке.

В ходе работы эмпирически доказано наличие феномена расхождения между математическим оптимумом функции потерь и клинически значимым качеством модели. Установлено, что традиционная практика ранней остановки обучения по локальному минимуму валидационных потерь в задачах скрининга может приводить к выбору субоптимальных решений.

Применение формализованного протокола выбора рабочей эпохи на основе строгих априорных ограничений позволило извлечь максимальную пользу из гибридной функции потерь, ориентированной на высокую чувствительность. Выбранная таким образом модель успешно подтвердила свою эффективность на независимой отложенной выборке: обеспечена высокая точность локализации патологии. $Dice_pos = 0,737$, и надежное покрытие обнаружения

Pos_any = 0,924, при этом уровень ложных тревог на снимках здоровых пациентов удержан в безопасных рамках клинических ограничений: FP_Rate = 0,203.

Ключевой научно-практический вклад работы заключается в переходе от ретроспективного подбора порогов к фиксированному правилу первичной конечной точки. Данный подход принципиально повышает воспроизводимость экспериментов, исключает риск непреднамеренной подгонки результатов под ожидаемый эффект и демонстрирует, что сегментационный модуль

отвечает требованиям безопасности по частоте ложных срабатываний. Разработанная методика формирует надежный инженерный базис для проектирования алгоритмов лучевой диагностики и их последующей интеграции в клинические системы поддержки принятия решений. Дальнейшие исследования будут направлены на проведение строгих абляционных экспериментов, валидацию протокола на внешних клинических базах данных и разработку системных механизмов каскадного подавления остаточных ложных тревог, а также создание СППР.

Список литературы

1. Lung imaging methods: indications, strengths and limitations / D. L. Tárnoki, K. Karlinger, C. A. Ridge, F. J. Kiss, T. Györke, E. M. Grabczak, Á. D. Tárnoki // *Breathe* (Sheff). 2024. Vol. 20, N 3. P. 230127. <https://doi.org/10.1183/20734735.0127-2023>.
2. Multidisciplinary Evaluation of an AI-Based Pneumothorax Detection Model: Clinical Comparison with Physicians in Edge and Cloud Environments / I. Dal, H. B. Kaya // *J. Multidiscip. Healthc.* 2025. Vol. 18. P. 4099–4111. <https://doi.org/10.2147/JMDH.S535405>.
3. Clinical adoption of deep learning target auto-segmentation for radiation therapy: challenges, clinical risks, and mitigation strategies / A. De Biase, N. M. Sijtsema, T. Janssen, C. Hurkmans, C. Brouwer, P. van Ooijen // *BJR Artificial Intelligence*. 2024. Vol. 1, N 1. P. ubae015. <https://doi.org/10.1093/bjrai/ubae015>.
4. Metrics reloaded: recommendations for image analysis validation / L. Maier-Hein, A. Reinke, P. Godau [et al.] // *Nat. Methods*. 2024. Vol. 21, N 2. P. 195–212. <https://doi.org/10.1038/s41592-023-02151-z>.
5. Understanding metric-related pitfalls in image analysis validation / A. Reinke, M. D. Tizabi, M. Baumgartner [et al.] // *Nat. Methods*. 2024. Vol. 21, N 2. P. 182–194. <https://doi.org/10.1038/s41592-023-02150-0>.
6. Mohammadi M., Mollazade K., Behrooz-Khazaei N. Under- and over-segmentation: new metrics for image segmentation accuracy measurement // *Array*. 2025. Vol. 28. P. 100624. <https://doi.org/10.1016/j.array.2025.100624>.
7. Association Between False-Positive Results and Return to Screening Mammography in the Breast Cancer Surveillance Consortium Cohort / D. L. Miglioretti, L. Abraham, B. L. Sprague [et al.] // *Ann. Intern. Med.* 2024. Vol. 177, N 10. P. 1297–1307. <https://doi.org/10.7326/M24-0123>.

8. Conquering class imbalances in deep learning-based segmentation of dental radiographs with different loss functions / M. Büttner, L. Schneider, A. Krasowski [et al.] // J. Dent. 2024. Vol. 148. P. 105063. <https://doi.org/10.1016/j.jdent.2024.105063>.
9. Cabezas M., Diez Y. An Analysis of Loss Functions for Heavily Imbalanced Lesion Segmentation // Sensors. 2024. Vol. 24, N 6. P. 1981. <https://doi.org/10.3390/s24061981>.
10. Reinke A., Tizabi M. D., Maier-Hein L. Advancing standards in biomedical image analysis validation: A perspective on Metrics Reloaded // Clin. Transl. Med. 2025. Vol. 15, N 9. P. e70237. <https://doi.org/10.1002/ctm2.70237>.
11. Малютин И. А., Кузьмин А. А., Шаталова О. В. Методы и алгоритмы анализа рентгенограмм грудной клетки, использующие локальные окна в задачах обнаружения патологий // Прикаспийский журнал: управление и высокие технологии. 2017. № 3 (39). С. 131–138.
12. Моделирование морфологических образований на рентгенограммах грудной клетки в интеллектуальных диагностических системах медицинского назначения / П. С. Кудрявцев, А. А. Кузьмин, Д. Ю. Савинов, С. А. Филист, О. В. Шаталова // Прикаспийский журнал: управление и высокие технологии. 2017. № 3 (39). С. 109–120.
13. Математические модели для интеллектуальных систем классификации рентгенограмм грудной клетки / М. В. Дюдин, П. С. Кудрявцев, К. В. Подмастерьев, С. А. Филист, О. В. Шаталова // Известия Юго-Западного государственного университета. Серия: Управление, вычислительная техника, информатика. Медицинское приборостроение. 2016. № 2 (19). С. 94–107.
14. Анализ биомедицинских изображений различными методами сегментации, основанными на операторах вычисления градиента / С. А. Филист, Е. А. Шашкова, О. В. Шаталова, Р. А. Томакова // Перспективы развития информационных технологий. 2011. № 3-1. С. 146–150.
15. An Automated System for Classification of Radiographs of the Breast / A. R. Dabagov, V. A. Gorbunov, S. A. Filist [et al.] // Biomedical Engineering. 2020. Vol. 53, N 6. P. 425–428. <https://doi.org/10.1007/s10527-020-09957-7>.
16. Hybrid neural networks with virtual flows in medical risk classifiers / K. Khatatneh, S. Filist, O. Shatalova [et al.] // Journal of Intelligent and Fuzzy Systems. 2022. Vol. 43, N 1. P. 1621–1632. <https://doi.org/10.3233/JIFS-212617>.
17. Коровин Е. Н., Шаталова О. В., Жилин В. В. Применение гибридной нейронной сети с макрослоями для классификации сердечно-сосудистых заболеваний // Биомедицинская радиоэлектроника. 2014. № 9. С. 32–37.
18. Biotechnical neural network system for predicting cardiovascular health state using processing of bio-signals / S. Filist, R. T. Al-Kasasbeh, O. V. Shatalova [et al.] // International Journal of Medical Engineering and Informatics. 2024. Vol. 16, N 4. P. 324–349. <https://doi.org/10.1504/IJMEI.2024.139884>.

19. Сверточная нейронная сеть для моделей классификаторов медицинского риска с синергетическими каналами / Р. А. Крупчатников, Д. А. Медников, З. У. Протасова [и др.] // Известия Юго-Западного государственного университета. Серия: Управление, вычислительная техника, информатика. Медицинское приборостроение. 2021. Т. 11, № 2. С. 25–50.
20. Multimodal neural network classifier of the functional state of the respiratory system / S. A. Filist, M. B. Myasnyankin, R. I. Safronov, A. A Kuzmin // Journal of Physics: Conference Series. 2021. Vol. 2094. P. 032064. <https://doi.org/10.1088/1742-6596/2094/3/032064>.
21. A multitask approach for automated detection and segmentation of thyroid nodules in ultrasound images / A. Radhachandran, A. Kinzel, J. Chen, V. Sant, M. Patel, R. Masamed, C. W. Arnold, W. Speier // Comput. Biol. Med. 2024. Vol. 170. P. 107974. <https://doi.org/10.1016/j.compbio.2024.107974>.
22. UNet++: A Nested U-Net Architecture for Medical Image Segmentation / Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, J. Liang // Deep Learn Med Image Anal Multimodal Learn Clin Decis Support. 2018. Vol. 11045. P. 3–11. https://doi.org/10.1007/978-3-030-00889-5_1.
23. Deeply-Supervised Nets / C. Y. Lee, S. Xie, P. Gallagher, Z. Zhang, Z. Tu // arXiv. URL: <https://arxiv.org/abs/1409.5185> (дата обращения: 16.03.2026).
24. Effect of image resolution on automated classification of chest X-rays / M. I. U. Haque, A. K. Dubey, I. Danciu, A. C. Justice, O. S. Ovchinnikova, J. D. Hinkle // J. Med. Imaging (Bellingham). 2023. Vol. 10, N 4. P. 044503. <https://doi.org/10.1117/1.JMI.10.4.044503>.
25. CACSNet for automatic robust classification and segmentation of carotid artery calcification on panoramic radiographs using a cascaded deep learning network / S. W. Yoo, S. Yang, J. E. Kim [et al.] // Sci. Rep. 2024. Vol. 14. P. 13894. <https://doi.org/10.1038/s41598-024-64265-4>.
26. How threshold customisation affects the performance of a multiclass X-ray AI model for primary care triage: a retrospective study / J. Z. T. Sim, J. Lin, Q. W. Fong, A. Y. Q. Soon, L. W. Khin, S. Balakrishnan, T. Lin, S. Wong, C. H. Tan // BMJ Open. 2026. Vol. 16, N 2. P. e111127. <https://doi.org/10.1136/bmjopen-2025-111127>.
27. Tabassum M., Di Ieva A., Liu S. Meta transfer learning for brain tumor segmentation using nnUNet in meningioma and metastasis cases // Sci. Rep. 2025. Vol. 15. P. 37599. <https://doi.org/10.1038/s41598-025-20957-z>.

References

1. Tárnoki D.L., Karlinger K., Ridge C.A., Kiss F.J., Györke T., Grabczak E.M., Tárnoki Á.D. Lung imaging methods: indications, strengths and limitations. *Breathe (Sheff)*. 2024;20(3):230127. <https://doi.org/10.1183/20734735.0127-2023>.

2. Dal I., Kaya H.B. Multidisciplinary Evaluation of an AI-Based Pneumothorax Detection Model: Clinical Comparison with Physicians in Edge and Cloud Environments. *J. Multidiscip. Healthc.* 2025;18:4099-4111. <https://doi.org/10.2147/JMDH.S535405>.
3. De Biase A., Sijtsema N.M., Janssen T., Hurkmans C., Brouwer C., van Ooijen P. Clinical adoption of deep learning target auto-segmentation for radiation therapy: challenges, clinical risks, and mitigation strategies. *BJR Artificial Intelligence.* 2024;1(1):ubae015. <https://doi.org/10.1093/bjrai/ubae015>.
4. Maier-Hein L., Reinke A., Godau P., et al. Metrics reloaded: recommendations for image analysis validation. *Nat. Methods.* 2024;21(2):195-212. <https://doi.org/10.1038/s41592-023-02151-z>.
5. Reinke A., Tizabi M.D., Baumgartner M., et al. Understanding metric-related pitfalls in image analysis validation. *Nat. Methods.* 2024;21(2):182-194. <https://doi.org/10.1038/s41592-023-02150-0>.
6. Mohammadi M., Mollazade K., Behroozi-Khazaei N. Under- and over-segmentation: new metrics for image segmentation accuracy measurement. *Array.* 2025;28:100624. <https://doi.org/10.1016/j.array.2025.100624>.
7. Miglioretti D.L., Abraham L., Sprague B.L., et al. Association Between False-Positive Results and Return to Screening Mammography in the Breast Cancer Surveillance Consortium Cohort. *Ann. Intern. Med.* 2024;177(10):1297-1307. <https://doi.org/10.7326/M24-0123>.
8. Büttner M., Schneider L., Krasowski A., et al. Conquering class imbalances in deep learning-based segmentation of dental radiographs with different loss functions. *J. Dent.* 2024;148:105063. <https://doi.org/10.1016/j.jdent.2024.105063>.
9. Cabezas M., Diez Y. An Analysis of Loss Functions for Heavily Imbalanced Lesion Segmentation. *Sensors.* 2024;24(6):1981. <https://doi.org/10.3390/s24061981>.
10. Reinke A., Tizabi M. D., Maier-Hein L. Advancing standards in biomedical image analysis validation: A perspective on Metrics Reloaded. *Clin. Transl. Med.* 2025;15(9):e70237. <https://doi.org/10.1002/ctm2.70237>.
11. Malyutina I.A., Kuz'min A.A., Shatalova O.V. Methods and algorithms for analyzing chest radiographs using local windows in pathology detection tasks. *Prikspiiskii zhurnal: upravlenie i vysokie tekhnologii = Caspian Journal: Management and High Technologies.* 2017;(3):131-138. (In Russ.)
12. Kudryavtsev P.S., Kuz'min A.A., Savinov D.Yu., Filist S.A., Shatalova O.V. Modeling of morphological formations on chest radiographs in intelligent medical diagnostic systems. *Prikspiiskii zhurnal: upravlenie i vysokie tekhnologii = Caspian Journal: Management and High Technologies.* 2017;(3):109-120. (In Russ.)
13. Dyudin M.V., Kudryavtsev P.S., Podmaster'ev K.V., Filist S.A., Shatalova O.V. Mathematical models for intelligent systems of chest radiograph classification. *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta. Seriya: Upravlenie, vychislitel'naja tekhnika,*

informatika. Meditsinskoe priborostroenie = Proceedings of the Southwest State University. Series: Control, Computer Engineering, Information Science. Medical Instruments Engineering. 2016;2(19):94-107. (In Russ.)

14. Filist S.A., Shashkova E.A., Shatalova O.V., Tomakova R.A. Analysis of biomedical images by various segmentation methods based on gradient operators. *Perspektivy razvitiya informatsionnykh tekhnologii = Prospects for the Development of Information Technology.* 2011;(3-1):146-150. (In Russ.)

15. Dabagov A.R., Gorbunov V.A., Filist S.A., et al. An Automated System for Classification of Radiographs of the Breast. *Biomedical Engineering.* 2020;53(6):425-428. <https://doi.org/10.1007/s10527-020-09957-7>.

16. Khatatneh K., Filist S., Shatalova O., et al. Hybrid neural networks with virtual flows in medical risk classifiers. *Journal of Intelligent and Fuzzy Systems.* 2022;43(1):1621–1632. <https://doi.org/10.3233/JIFS-212617>.

17. Korovin E.N., Shatalova O.V., Zhilin V.V. Application of a hybrid neural network with macrolayers for the classification of cardiovascular diseases. *Biomeditsinskaya radioelektronika = Biomedical Radioelectronics.* 2014;(9):32-37. (In Russ.)

18. Filist S., Al-Kasasbeh R.T., Shatalova O.V., et al. Biotechnical neural network system for predicting cardiovascular health state using processing of bio-signals. *International Journal of Medical Engineering and Informatics.* 2024;16(4):324-349. <https://doi.org/10.1504/IJMEI.2024.139884>.

19. Krupchatnikov R.A., Mednikov D.A., Protasova Z.U., et al. Convolutional neural network for medical risk classifier models with synergetic channels. *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta. Seriya: Upravlenie, vychislitel'naja tekhnika, informatika. Meditsinskoe priborostroenie = Proceedings of the Southwest State University. Series: Control, Computer Engineering, Information Science. Medical Instruments Engineering.* 2021;11(2):25–50. (In Russ.)

20. Filist S.A., Myasnyankin M.B., Safronov R.I., Kuzmin A.A. Multimodal neural network classifier of the functional state of the respiratory system. *Journal of Physics: Conference Series.* 2021;2094:032064. <https://doi.org/10.1088/1742-6596/2094/3/032064>.

21. Radhachandran A., Kinzel A., Chen J., Sant V., Patel M., Masamed R., Arnold C.W., Speier W. A multitask approach for automated detection and segmentation of thyroid nodules in ultrasound images. *Comput. Biol. Med.* 2024;170:107974. <https://doi.org/10.1016/j.compbimed.2024.107974>.

22. Zhou Z., Siddiquee M.M.R., Tajbakhsh N., Liang J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. *Deep Learn Med Image Anal Multimodal Learn Clin Decis Support.* 2018;11045:3–11. https://doi.org/10.1007/978-3-030-00889-5_1.

23. Lee C.Y., Xie S., Gallagher P., Zhang Z., Tu Z. Deeply-Supervised Nets. Available at: <https://arxiv.org/abs/1409.5185> (accessed 16.03.2026).

24. Haque M.I.U., Dubey A.K., Danciu I., Justice A.C., Ovchinnikova O.S., Hinkle J.D. Effect of image resolution on automated classification of chest X-rays. *J. Med. Imaging (Bellingham)*. 2023;10(4):044503. <https://doi.org/10.1117/1.JMI.10.4.044503>.
25. Yoo S.W., Yang S., Kim J.E., et al. CACSNet for automatic robust classification and segmentation of carotid artery calcification on panoramic radiographs using a cascaded deep learning network. *Sci. Rep.* 2024;14:13894. <https://doi.org/10.1038/s41598-024-64265-4>.
26. Sim J.Z.T., Lin J., Fong Q.W., Soon A.Y.Q., Khin L.W., Balakrishnan S., Lin T., Wong S., Tan C.H. How threshold customisation affects the performance of a multiclass X-ray AI model for primary care triage: a retrospective study. *BMJ Open*. 2026;16(2):e111127. <https://doi.org/10.1136/bmjopen-2025-111127>.
27. Tabassum M., Di Ieva A., Liu S. Meta transfer learning for brain tumor segmentation using nnUNet in meningioma and metastasis cases. *Sci. Rep.* 2025;15:37599. <https://doi.org/10.1038/s41598-025-20957-z>.

Информация об авторах / Information about the Authors

Косинов Александр Владимирович, аспирант кафедры биомедицинской инженерии, Юго-Западный государственный университет, г. Курск, Российская Федерация, e-mail: skosinka@yandex.ru, ORCID: 0009-0008-0426-5045

Aleksandr V. Kosinov, Postgraduate at the Department of Biomedical Engineering, Southwest State University, Kursk, Russian Federation, e-mail: skosinka@yandex.ru, ORCID: 0009-0008-0426-5045

Крупчатников Роман Анатольевич, доктор технических наук, профессор, профессор кафедры биомедицинской инженерии, Юго-Западный государственный университет, г. Курск, Российская Федерация, e-mail: roman0406@yandex.ru, ORCID: 0000-0002-4951-8607, Scopus ID: 24471766700

Roman A. Krupchatnikov, Dr. Sci. (Engineering), Professor, Professor at the Department of Biomedical Engineering, Southwest State University, Kursk, Russian Federation, e-mail: roman0406@yandex.ru, ORCID: 0000-0002-4951-8607, Scopus ID: 24471766700

Серебровский Вадим Владимирович, доктор технических наук, профессор, профессор кафедры программной инженерии, Юго-Западный государственный университет, г. Курск, Российская Федерация, e-mail: SV1111@mail.ru, Researcher ID: O-8221-2015, ORCID: 0000-0002-1923-3862

Vadim V. Serebrovsky, Dr. of Sci. (Engineering), Professor, Professor of the Department of Software Engineering, Southwest State University, Kursk, Russian Federation, e-mail: SV1111@mail.ru, Researcher ID: O-8221-2015, ORCID: 0000-0002-1923-3862

Ефремов Михаил Александрович, кандидат технических наук, доцент кафедры информационной безопасности, Юго-Западный государственный университет, г. Курск, Российская Федерация, e-mail: Efremov-ma@mail.ru, ORCID: 0000-0002-2828-6261, Scopus ID: 58712948800

Mihail A. Efremov, Cand. Sci. (Engineering), Associate Professor at the Department of Information Security, Southwest State University, Kursk, Russia Federation, e-mail: Efremov-ma@mail.ru, ORCID: 0000-0002-2828-6261, Scopus ID: 58712948800

Кондрашов Дмитрий Сергеевич, кандидат технических наук, преподаватель кафедры биомедицинской инженерии, Юго-Западный государственный университет, г. Курск, Российская Федерация, e-mail: kondrashov012@mail.ru, Researcher ID: KFT-0791-2024, ORCID: 0000-0002-3925-4601

Dmitry S. Kondrashov, Cand. Sci. (Engineering), Lecturer at the Department of Biomedical Engineering, Southwest State University, Kursk, Russia Federation, e-mail: kondrashov012@mail.ru, Researcher ID: KFT-0791-2024, ORCID: 0000-0002-3925-4601