

# Применение нейросетевого подхода для анализа снимков компьютерного томографа при заболеваниях надпочечников

А.А. Приваленко, К.С. Зайцев, С.А. Бурякина, Н.В. Тарбаева, С.М. Китаев, К.В. Кухтиков

**Аннотация.** Целью настоящей работы является исследование возможностей моделей глубокого обучения в диагностике заболеваний надпочечников по изображениям компьютерного томографа (КТ). Для поддержки принятия врачебных решений и автоматизированной классификации найденных новообразований по КТ брюшной полости разработан двухэтапный нейросетевой подход, сочетающий сегментацию и классификацию патологий. Он позволяет визуализировать найденные образования и работать с несколькими образованиями, находящимися на одном КТ-снимке. В исследовании использован набор данных, предоставленный НИИЦ эндокринологии им. академика И.И. Дедова, на момент написания статьи состоящий из 228 КТ-снимков. Для оптимизации времени обработки снимки были конвертированы в видеоформат MP4 (54 кадра в видео), что сократило объем данных без существенной потери диагностической ценности. Изображения прошли предварительную обработку, для борьбы с дисбалансом классов была использована аугментация. Каждый КТ-снимок имеет по три метки, каждая из которых соответствует наличию новообразования соответствующего вида, а именно: злокачественный, доброкачественный или неопределенный фенотип. Для инстанса сегментации новообразований применена модель YOLOv11-seg с предобучением на датасете COCO. Для классификации использована модель 3DResNet-50, обученная на выделенных областях. Предлагаемый комбинированный двухступенчатый подход реализуется в программном комплексе «Ассистент врача-эндокринолога».

**Ключевые слова** — нейронные сети, сегментация, классификация, заболевание надпочечников, КТ-изображения

## 1. ВВЕДЕНИЕ

Медицинская сфера является одной из ключевых и социально значимых отраслей, где внедрение современных методов компьютерного зрения, основанных на искусственном интеллекте, позволяет усовершенствовать методы сохранения здоровья людей. Необходимость перехода к персонализированной, предиктивной и профилактической медицине, а также развитие

высокотехнологичного здравоохранения отметил в своем Указе от 28.02.2024 г. № 145 «О Стратегии научно-технологического развития Российской Федерации» Президент Российской Федерации.

Одним из наиболее информативных методов диагностики внутренних органов является компьютерная томография (КТ). Она позволяет получать детализированные срезы тканей с высокой пространственной точностью, в том числе и для выявления новообразований надпочечников, в качестве первичной диагностики.

Надпочечники играют критически важную роль в регуляции множества физиологических процессов за счёт синтеза гормонов, таких как кортизол, адреналин и альдостерон. Эти гормоны участвуют в управлении стрессовыми реакциями, обменом веществ, артериальным давлением и водно-солевым балансом организма. Опухоли надпочечников, особенно злокачественные, представляют собой серьёзную угрозу из-за скрытой клинической картины и высокого риска поздней диагностики, что делает задачу их раннего выявления особенно актуальной. Эти патологии часто маскируются под другие заболевания, что затрудняет их раннее выявление, поэтому использование искусственного интеллекта для их раннего выявления является актуальной задачей.

Несмотря на актуальность этой эндокринологической задачи, ей в настоящее время уделено не так много внимания в публикациях, как, например, задачам пульмонологии или дерматологии. Важность задачи подчеркивает статья [1] указывающая, что только 30% случаев рака надпочечников ограничены самим органом на момент постановки диагноза. Это указывает на то, что большинство опухолей обнаруживаются на поздних стадиях, когда они уже распространились за пределы надпочечников. На точность оценки функции надпочечников влияет множество факторов, и при компьютерной томографии случайные образования обнаруживаются с частотой около 4-5% [2]. Поэтому крайне важно разработать

неинвазивный и упрощенный метод предоперационной диагностики.

Термин «инциденталома надпочечников» (ИН) означает новообразование надпочечников, обнаруженные случайно при визуализации других заболеваний брюшной полости, выполняемых по причинам, не связанным с заболеваниями надпочечников. Несмотря на их частое выявление, по-прежнему сложно распознать и лечить небольшой процент ИН, которые представляют значительный риск либо из-за своей гормональной активности, либо из-за злокачественной природы [3].

Ранняя постановка диагноза имеет решающее значение для оперативного и эффективного лечения.

Традиционно обнаружение поражений надпочечников при компьютерной томографии зависит от ручного анализа снимков радиологами, который может быть трудоемким и несистематическим.

Чтобы повысить уровень обнаружения патологий надпочечников по КТ-снимкам предлагается использовать двухэтапный нейросетевой подход, сочетающий сегментацию и классификацию патологий надпочечников.

## 2. СОПУТСТВУЮЩИЕ РАБОТЫ

Глубокое обучение стало революционным инструментом в анализе биомедицинских изображений, обеспечивая высокую точность в задачах сегментации и классификации [4, 5, 6, 7].

Однако, в большинстве существующих подходов при классификации патологий надпочечников по-прежнему преобладают ручные методы сегментации и традиционные алгоритмы машинного обучения [8, 9, 10, 11], опирающиеся на ограниченное количество признаков и часто не учитывающие пространственный контекст. Это объясняется как ограниченным количеством доступных и обширных размеченных датасетов для обучения моделей, так и сложностями детекции и диагностики мелких анатомических структур. На рисунке 1 представлены метрики качества модели сегментации органов брюшной полости [12]. Из рисунка видно, что точность для правого и левого надпочечников (adrenal glands) снижена на 15–20% по сравнению с другими более крупными органами. Сложность сегментации надпочечников из-за их размера также отмечается в исследовании [13].

При проведении литературного анализа мы столкнулись с проблемой ограниченного количества материалов и доступных данных для обучения моделей. Удалось найти лишь небольшое количество публичных датасетов, и все они содержат менее 1000 размеченных снимков, тогда как для обучения устойчивых моделей требуется несколько тысяч примеров. В исследованных статьях медианное значение числа

пациентов составляет - 125 [14]. Это подчеркивает необходимость использования приемов, эффективных для обучения моделей на небольших датасетах. К тому же, в публикациях обычно решается задача бинарной классификации, а сегментация при этом делается вручную.

Organ	Mean Score
Spleen	0.9326725602149963
Right Kidney	0.9482816457748413
Left Kidney	0.9359169006347656
Gallbladder	0.8265718221664429
Esophagus	0.7950014472007751
Liver	0.9665778279304504
Stomach	0.8874692916870117
Aorta	0.9340504440788269
Inferior Vena Cava	0.8848044872283936
Pancreas	0.8297500610351562
Right Adrenal Gland	0.7409824132919312
left adrenal gland	0.7365583181381226
Duodenum	0.774817705154419
Bladder	0.7977799773216248
Prostate/Uterus	0.7930908203125

Рис.1 Качество сегментации органов брюшной полости.

Исследования в области анализа медицинских изображений показывают, что 2D-модели уступают 3D-моделям в тестах 3D-компьютерного зрения, поскольку они изначально не используют поперечную или временную информацию [15]. В приведенном исследовании предварительное обучение на видеоданных продемонстрировало лучшие результаты в задачах обнаружения узлов в легких и оценке эмболий.

В статье [15] также отмечается преимущество предварительного обучения на видео в широком спектре задач. Перед тонкой настройкой модели для решения медицинских задач, было проведено предварительное обучение на наборе данных для распознавания действий человека на видео - Kinetics. В нашем исследовании был применен похожий подход для решения задачи сегментации надпочечников с использованием модели yolol1-seg, прошедшая предварительное обучение на датасете СОСО. Это дало положительный результат. Однако стоит отметить, что не все предобученные модели могут подойти в решении той или иной задачи.

Ниже приведены два списка нейросетевых моделей для решения задач сегментации и классификации.

Основные модели сегментации, применяемые для анализа опухолей различных органов (печени/легких/надпочечников) на КТ-изображениях.

1. Семейство моделей *EfficientNet* продемонстрировали высокие показатели IoU\_Scores для сегментации фона, почек и опухоли, при этом средние значения IoU\_Scores варьировались от 0,976 для B0 до 0,980 для B4 [16]. B7 получил наивысший балл IoU\_Score за сегментацию почек, в то время как B4 получил наивысший балл за сегментацию опухолей. В исследовании использовался набор данных KiTS19 для КТ-изображений с контрастным усилением. Используя семантическую сегментацию для моделей U-Net семейства *EfficientNet*, метод оказался еще более надежным и помог врачам в точном обнаружении опухоли и классификации изображений для ранней диагностики.

2. В исследовании [17] была предложена архитектура *3D U-Net*, объединённая с глубокой супервизией и экспоненциальной логарифмической функцией потерь для повышения эффективности обучения. Модель достигла коэффициента Dice 0,969 для почек и 0,805 для опухолей на наборе данных KiTS19, что свидетельствует о высокой точности.

3. В исследовании, опубликованном в журнале *Applied Sciences* [18], была предложена модель, объединяющая архитектуру *Feature Pyramid Network (FPN)* с *SE-ResNet* для точной сегментации и классификации почек и опухолей на КТ-изображениях. Модель достигла средних значений IoU от 0,981 до 0,988 и коэффициента Dice до 0,993. Эксперименты показали эффективность семейства *ResNet*.

4. В этом исследовании [19] была разработана модель на основе *YOLOv3* и *InceptionResNetV2*. Модель *YOLOv3* достигла средней точности (mAP) 0,99 при локализации опухолей печени, а общая точность сегментации превысила 95%.

5. Авторы статьи [20] предложили архитектуру *U-Net* на основе "Reset U-Net" для сегментации медицинских изображений 3D, решающую проблему исчезающего градиента и повышающую производительность модели.

6. Модель *UniverSeg*, представленная в статье [21], является революционной моделью в области медицинской сегментации изображений и одной из разновидностей U-net нейросетей, в которой реализован принцип поиска результата без необходимости предварительного обучения на целевых данных.

Основные нейросетевые модели классификации, применяемые для анализа опухолей на КТ-изображениях надпочечников.

1. В исследовании [22] используется последовательная обработка небольших участков целого изображения для извлечения локальных характеристик. Эта процедура повторяется несколько раз, после чего выделенные признаки

отправляются в классификатор. Точность классификации - запредельная (99,87%), особенно учитывая использование классических методов *ML*. А так как в точности повторить эксперимент не представляется возможным из-за отсутствия точной информации по выбранной модели и способу ее обучения, такие высокие результаты кажутся следствием утечки данных.

2. В работе [8] по диагностике аденом надпочечников присутствует ручной выбор срезов, содержащих опухоль, формируя "кубы". Из каждого куба поэтапно извлекаются признаки, пока не остается десять наиболее информативных, после чего применяются классические методы *ML*.

3. Использование предварительно обученных глубоких сверточных нейронных сетей для бинарной классификации КТ-снимков показало эффективность трансферного обучения на основе *VGG* и *ResNet* в исследованиях [23, 24]

4. Исследователи в рамках статьи [25] разработали 2.5D модель глубокого обучения для выполнения бинарной классификации по 3D КТ-сканам пациентов. Она определяет наличие поражения у надпочечника, то есть решает задачу бинарной классификации. Авторы верно подмечают тенденцию применения стандартных трехмерных моделей глубокого обучения для классификации медицинских КТ-снимков, использование которых дает некоторые многообещающие результаты [26, 27, 28]. Однако приобретение достаточного количества видеоданных для обучения является сложной задачей. К сложностям можно отнести высокие эксплуатационные расходы, ограниченное количество доступных компьютерных томографов и облучение пациентов. В большинстве случаев производительность трехмерных моделей глубокого обучения ограничена небольшим и недиверсифицированным набором данных, отмечают авторы. Их подход интересен тем, что берется небольшой срез (три кадра в кинопетле), что способствует использованию небольшой вычислительной мощности для обучения модели, однако теряется автоматизация, поскольку надо выбрать серию кадров, где находится надпочечник. Для обучения модели и тестирования производительности в качестве архитектуры модели была выбрана *3D ResNet34* (с AUC ROC = 0,79), хотя также велись эксперименты с семейством *EfficientNet*.

5. В исследовании [13] разработанный подход способен различать нормальные надпочечники и те, в которых есть новообразования. Отмечается, что классификатор имеет большую ошибку, чем модель сегментации.

Таблица 1. Метрики моделей сегментации

Модель	Precision	Recall	F1-score	Размер выборки
3D U-Net	0.7	0.64	0.67	255

3D U-Net	0.69	0.76	0.73	38
Yolo_3	0.95	0.92	0.93	1250
ViT	0.84	0.6	0.63	113
3D V-Net	0.9	0.9	0.9	870

В таблице 1 представлены метрики моделей сегментации, выбранные из литературных источников. Стоит отметить, что Yolo\_3 обучена на самом большом наборе данных и показывает лучшее качество.

В результате исследования сопутствующих работ можно сделать следующие выводы:

- 1) имеются ограничения в части дисбаланса классов для «неопределенных» случаев. Их, как показывает практика, мало (5-10%), разрабатывается чаще всего бинарный классификатор, отделяющий доброкачественные образования от злокачественных, и сегментация проводится вручную;
- 2) исследования, в которых в роли классификатора использовались модели на основе ResNet, встречались чаще других;
- 3) 2D-модели уступают 3D-моделям в тестах 3D-компьютерного зрения;
- 4) современные архитектуры глубоких нейронных сетей демонстрируют более высокую точность и обобщающую способность по сравнению с классическими методами машинного обучения и ранними сверточными решениями;
- 5) отмечается преимущество предварительного обучения на видео в широком спектре задач;
- 6) для анализа КТ-изображений печени/легких/опухолей чаще других используются семейства моделей EfficientNet, ResNet, Yolo, U-Net, VGG, ViT.

### 3. ПОДГОТОВКА ДАННЫХ

Для проведения исследований был использован набор данных, предоставленный НИИЦ эндокринологии им. академика И.И. Дедова, который на данный момент состоит из 228 КТ-снимков (122 с патологиями). Стоит отметить, что это превышает типичный объем, представленный в литературе, и потенциально повышает обобщающую способность модели. Это заслуга врачей, с которыми мы сотрудничаем.

Для оптимизации времени обработки снимки были конвертированы в видеоформат MP4 (54 кадра в видео), что сократило объем данных и способствовало увеличению скорости обучения моделей без ощутимой потери качества.

Изображения прошли предварительную обработку с целью стандартизации входных данных. Обработка включала извлечение определённого количества (эмпирически было выбрано 54) равномерно распределённых по

длительности кадров из видео, полученных при конвертации томографических серий. Каждый кадр был масштабирован до фиксированного разрешения 224×224 пикселей, конвертирован в оттенки серого или RGB-формат в зависимости от архитектуры модели и нормализован по интенсивности пикселей до диапазона [0, 1]. Для случаев, когда длина видеофрагмента не достигает требуемого количества кадров, предусмотрен padding-слой для единообразия входной размерности. Также были удалены дубликаты и проверено, нет ли утечки данных из тренировочной выборки в тестовую. Обработка выполнялась с использованием OpenCV и PyTorch transforms, включая преобразование в тензоры и нормализацию.

Для борьбы с дисбалансом классов (рис.2) и увеличения объёма обучающей выборки до 300 примеров на класс применялись методы аугментации, включающие пространственные трансформации (горизонтальные повороты, масштабирование, случайные обрезки), а также цветовые искажения (яркость, контраст). Это позволило повысить обобщающую способность модели.

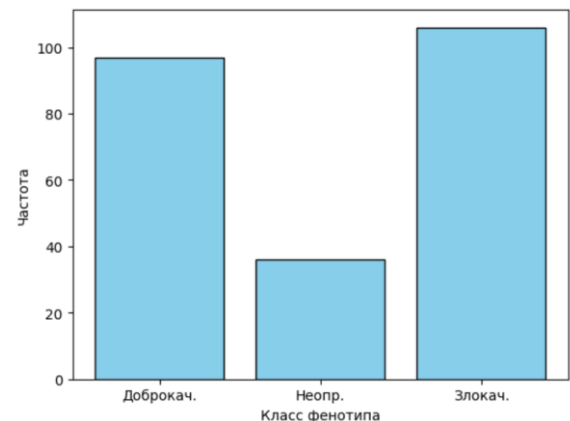


Рис. 2 Дисбаланс классов исходной выборки.

В силу небольшого объема данных, в качестве метки предсказания вместо детализированных диагнозов, которых в нашем случае порядка 10, было принято решение использовать вид новообразования

(злокачественный/доброкачественный/неопределенный). Выбранный подход имеет как преимущества, так и ограничения. Во-первых, сокращение до 3 классов увеличивает количество примеров на класс, снижая риск недообучения и повышая обобщающую способность модели. Во-вторых, это предотвращает недостаток примеров на каждый класс. В-третьих, в практике врача первичный вопрос является: «Есть ли онкологическая угроза?», а не точный гистологический тип. Мы придерживаемся цели получения быстрого диагноза об отсутствии патологий. Для уточнения диагноза могут потребоваться дополнительные исследования или мнения других врачей, особенно для неопределённых новообразований, которые,

несомненно, нуждаются в дальнейшем обследовании, что может включать более сложные и дорогостоящие рентгенологические исследования или инвазивные процедуры, такие как ПЭТ КТ, биопсия, резекция образования.

В случае работы с моделями, приспособленными к 2D-данным, была проведена дополнительная предобработка данных. Заранее из всего объёма данных случайным образом отбиралась тестовая выборка. Из последовательности кадров удалялись первые и последние 10 кадров, так как они, как правило, содержали либо вход в зону интереса, либо выход

из неё и не несли значимой диагностической нагрузки. Далее оставшиеся последовательности были преобразованы в отдельные изображения с целевым размером 224×224 пикселя формата .png. А для упрощения работы модели маски были переведены в бинарный вид. Тренировочный датасет состоял из 7400 примеров, на 3400 из которых имелось образование (для баланса выборки). После этого были произведены контроль на наличие дублирующих кадров и перераспределение датасета путём случайного перемешивания. Пример КТ-срезов и масок к ним представлен на рис 3.

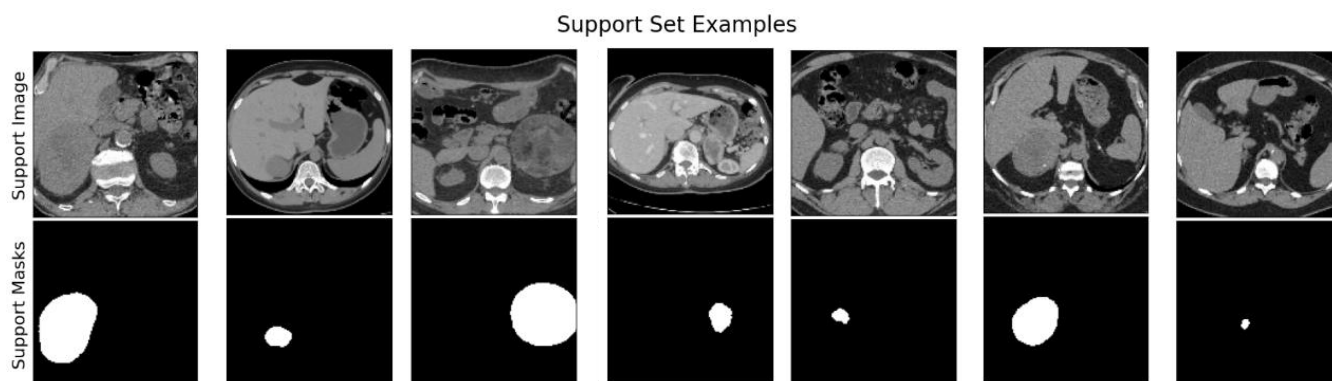


Рис. 3 Пример КТ-срезов с бинарными масками, определяющими опухоль

Поскольку исходный датасет содержал значительное число негативных примеров (кадров без видимых патологий), была применена стратегия балансировки классов. Для этого все изображения были разделены на две группы — с наличием и без наличия образования — на основании содержимого бинарных масок. Количество изображений без патологий было урезано до числа положительных примеров, что позволило сбалансировать объём обучающей выборки и улучшить обобщающую способность модели.

Подготовленные изображения и маски были сохранены в структурированную файловую систему и конвертированы в формат COCO JSON (для работы с моделью Yolo), включающий описание аннотаций и контуров патологических образований. Для извлечения контуров использовался алгоритм OpenCV с пороговой сегментацией (Otsu thresholding) и функцией cv2.findContours. Каждая аннотация включала в себя координаты ограничивающего прямоугольника (bounding box), площадь области, а также координаты сегмента (polygon segmentation), соответствующие найденному контуру.

#### 4. МОДЕЛИ И МЕТОДЫ

С развитием современных архитектур компьютерного зрения — таких как U-Net, YOLO, Vision Transformers (ViT) — появилась возможность интегрировать пространственные

зависимости между срезами, учитывать структуру органов и агрегировать информацию на уровне всего обследования, а не отдельных изображений. Современные архитектуры демонстрируют превосходство в распознавании пространственных и контекстных признаков, но их высокая вычислительная нагрузка и зависимость от больших размеченных датасетов пока остаются проблемами.

**YOLO (You Only Look Once)** — это комплекс моделей, для детекции, сегментации и классификации объектов на изображениях с высокой скоростью и точностью [29].

**ResNet (Residual Network)** — архитектура, предложенная для борьбы с проблемой исчезающего градиента при обучении глубоких нейросетей. Введённые остаточные блоки (skip connections) позволяют эффективно обучать очень глубокие модели, улучшая точность и стабильность сходимости. Она широко применяется в задачах классификации изображений и служит основой для построения других моделей. **U-Net** — одна из ключевых архитектур для сегментации медицинских изображений, разработанная с целью эффективной работы с ограниченным количеством данных [30]. В частности, архитектура Resnet U-Net решает проблему исчезающего градиента и улучшает качество сегментации трёхмерных данных. Вариант U-Net, основанный на EfficientNet, сочетает в себе меньшую сложность модели и высокую точность, обеспечивая лучшее соотношение между

производительностью и вычислительными затратами.

**EfficientNet** — семейство сверточных нейросетей, разработанных с использованием метода масштабирования модели по глубине, ширине и разрешению входных изображений одновременно. Это позволяет добиться высокой точности при меньшем числе параметров и вычислительных затратах, что делает EfficientNet эффективным выбором для задач классификации и сегментации.

**VGG** — классическая сверточная архитектура, отличающаяся простотой и однородностью структуры. Несмотря на значительный объем параметров, VGG остаётся популярной базовой моделью благодаря своей предсказуемости и эффективности в задачах классификации.

**ViPTT** (Vision Transformer with Progressive Tokenization and Training) — современный подход, основанный на трансформерах, который применяет токенизацию изображений и учится моделировать пространственные зависимости с помощью механизмов внимания. ViPTT улучшает обработку больших изображений и позволяет учитывать глобальные контекстные связи, что особенно важно для сложных задач анализа медицинских данных [31].

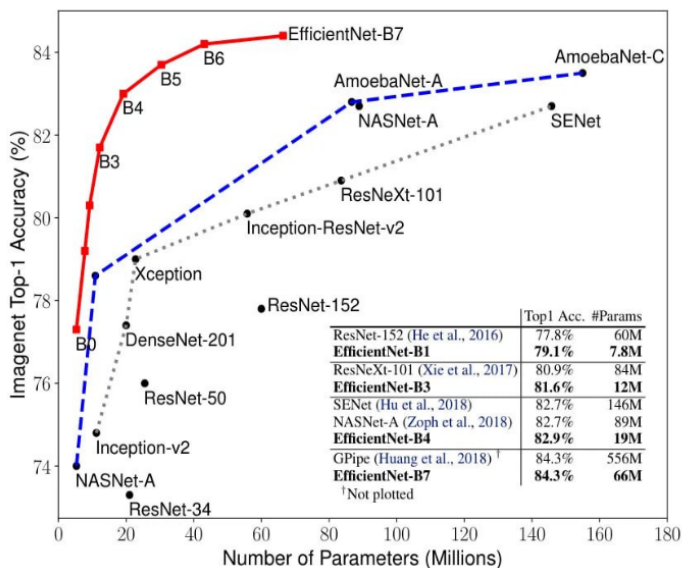


Рис. 4 График зависимости точности моделей от количества параметров

Стоит отметить, что порядок разработки рассматриваемых моделей следует из их новизны, а именно: UniverSeg, ViPTT, YOLO, EfficientNet, ResNet, U-Net, VGG

Рекомендуется анализировать и применять модели в данном порядке, так как более современные архитектуры, как правило, демонстрируют улучшенные базовые показатели качества, и обладают более эффективными

механизмами извлечения признаков [32], что можно заметить по рисунку 4.

В следующем разделе представлены результаты экспериментов на используемых данных. С целью улучшения обобщающей способности моделей были проведены экспериментальные исследования, включающие в себя применение различных техник (regularization, finetuning). Dropout - регуляризационный метод, применяемый в нейронных сетях, который случайным образом отключает заданную долю нейронов на каждой итерации обучения. Файнтьюнинг (Finetuning) – процесс поиска оптимальных значений гиперпараметров модели (например, learning rate, batch size, количество слоев и т.д.) с целью найти комбинацию, обеспечивающую наилучшую производительность модели на валидационном наборе данных.

Модель классификации обучалась на видеоданных формата mp-4 после умножения исходного КТ-снимка на бинарную маску с новообразованием. Таким образом, на вход классификатору подавались уже выделенные области интереса (ROI), что позволяет сосредоточить обучение модели на фенотипических особенностях самих образований. Для оценки качества классификации использовались стандартные метрики, основанные на матрице ошибок (confusion matrix), включая accuracy, recall (полнота), precision (точность) и F1-score. Эти метрики позволяют оценить как общее качество модели, так и её способность улавливать редкие, но клинически значимые классы. Приоритетом в данной медицинской задаче является полнота.

Модель сегментации обучалась как на видеоданных, используя предобученные для данной задачи модели, так и на 2D-срезах. В частности, был опробован подход, представленный в [31], в котором трёхмерные снимки КТ преобразуются в последовательность кадров, которые поэтапно обрабатываются моделью инстанс сегментации, после чего отправляются в классификатор. Для оценки качества сегментации были использованы следующие метрики:

Precision и Recall — для оценки точности предсказаний в сравнении с истинными масками.

mAP50 (mean Average Precision при IoU  $\geq 0.5$ ) — отражает долю верных предсказаний с достаточным перекрытием с истинными объектами.

mAP50-95 — более строгая метрика, усреднённая по порогам IoU от 0.5 до 0.95 с шагом 0.05, позволяющая оценить устойчивость модели к изменению требований по точности сегментации.

## 5. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

### 5.1 Задача сегментации

В силу того, что не было найдено в открытых источниках предобученных 3D-моделей сегментации вместе с весовыми коэффициентами, начали этап экспериментов с моделями, обученными сегментации медицинских

образований, а именно – с UniverSeg. Модель на вход принимает (images, support\_images, support\_labels) и выдает маски. В документации подход работы объясняется следующим образом (рис. 5):

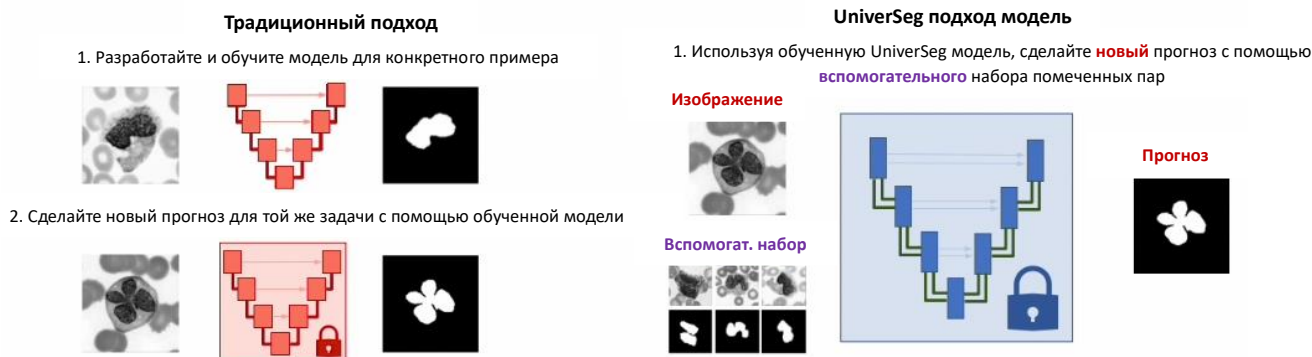


Рис.5. Принцип обучения модели UniverSeg (из документации)

Для обучения и тестирования был подготовлен специальный dataloader с кадрами, предобработка которых приведена в разделе “Подготовка данных”, причем протестировали два подхода: датасет, где половина кадров с образованием, а вторая – без него; датасет только с бинарными масками образований.

К сожалению, на наших данных оба случая не дали результатов, модель не смогла построить ни одной маски. Варьирование параметров, увеличение специального параметра “пары поддержки” (support\_set) до максимального возможного - тоже не помогли.

Также некоторые, обученные на КТ-снимках, модели сегментации не подошли по аналогичной причине, поэтому следующая модель была выбрана из опыта исследователей в использовании крупных моделей с дообучением для решения медицинской задачи, а также ввиду большого набора литературы, а именно – YOLO. Была использована последняя доступная версия на момент написания статьи - 11. Важно отметить, что предварительная обработка изображений оставалась неизменной, однако каждая модель из семейства YOLO требует специфического формата аннотаций и структуры датасета, соответствующих её версии и реализации.

Всего существует 5 видов моделей инстанс сегментации YOLO, различающихся количеством параметров, это nano, small, medium, large и extra large. YOLO определяет точность двух видов объектов – bounding box (обводка образования, показывающая насколько точно модель определила местоположение образования), и mask – насколько точно внутри bounding box были выделены пиксели. В таблице 2 представлены метрики качества модели каждого размера, на рисунке 6 – их столбчатые диаграммы.

По результатам данного исследования было обнаружено, что модели small, medium, large и xlarge имеют улучшение метрик в пределах погрешности. Следовательно, остановились на модели small, как на самой оптимальной по соотношению качество/время вычислений. С этой моделью продолжились эксперименты, а именно выбор batch-size (таблица 3), а также различные манипуляции (dropout, использование кадров с пустыми масками), результаты которых можно видеть в таблице 4.

Графики метрик при обучении лучшей модели представлены на рис 7 и 8. Пример результата применения модели приведен на рис 9.

Таблица 2. Метрики качества модели YOLO разных размеров

	Box				Masks				time	GFLOPs
	Precision	Recall	mAP-50	mAP50-95	Precision	Recall	mAP-50	mAP50-95		
nano	0,952	0,869	0,919	0,69	0,906	0,794	0,875	0,517	1,994	10,2
small	0,951	0,906	0,945	0,736	0,906	0,842	0,9	0,555	2,04	30

medium	0,953	0,883	0,936	0,729	0,921	0,829	0,9	0,558	2,6	123
large	0,957	0,904	0,947	0,751	0,927	0,809	0,896	0,551	3,083	141,9
xlarge	0,954	0,904	0,954	0,752	0,921	0,818	0,904	0,551	4,057	318,5

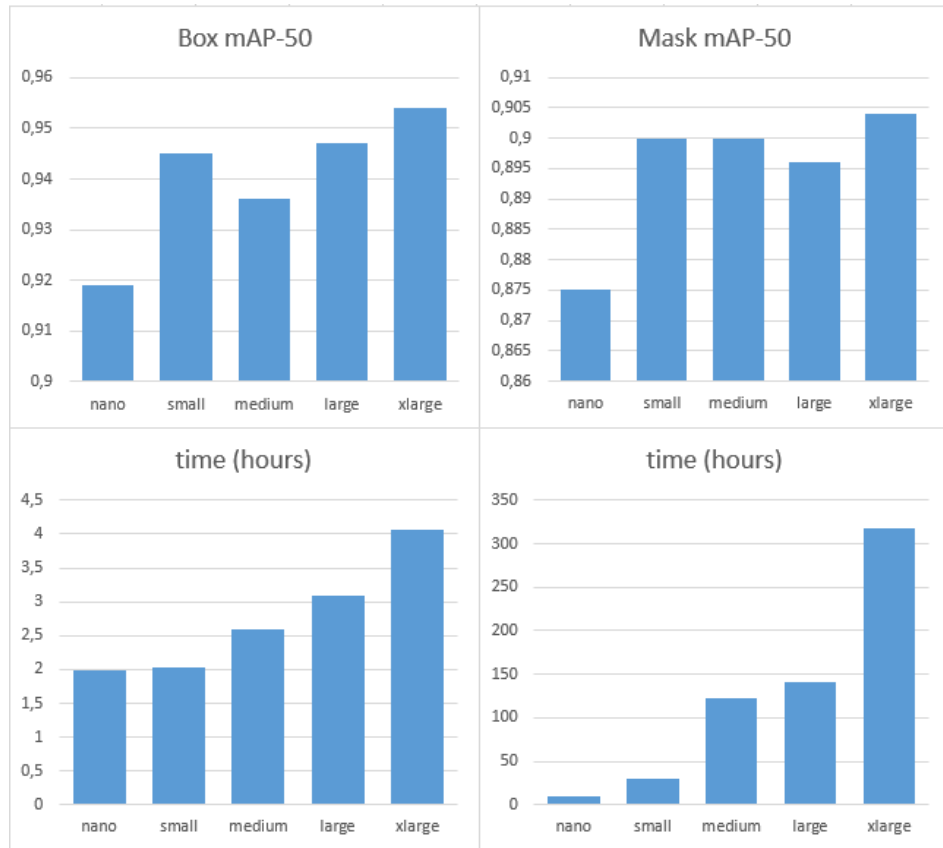


Рис. 6 Сравнение метрик модели YOLO разных размеров

Таблица 3. Выбор batch-size при обучении модели YOLOv11 small

batch_size	время (час)	mAP50 (detection)	mAP50 (segmentation)
4	1,073	0,888	0,848
8	0,677	0,876	0,839
16	0,459	0,893	0,832
32	0,307	0,886	0,842
64	0,274	0,889	0,867

Таблица 4. Сравнение результатов работы YOLOv11 small

	Box				Masks			
	Precision	Recall	mAP-50	mAP50-95	Precision	Recall	mAP-50	mAP50-95
small	0,951	0,906	0,945	0,736	0,906	0,842	0,9	0,555
small + dropout	0,989	0,891	0,949	0,751	0,941	0,841	0,908	0,561
small + empty masks	0,969	0,882	0,934	0,71	0,929	0,8	0,879	0,523



small + dropout + empty masks	0,978	0,834	0,908	0,685	0,945	0,771	0,876	0,52
----------------------------------	-------	-------	-------	-------	-------	-------	-------	------

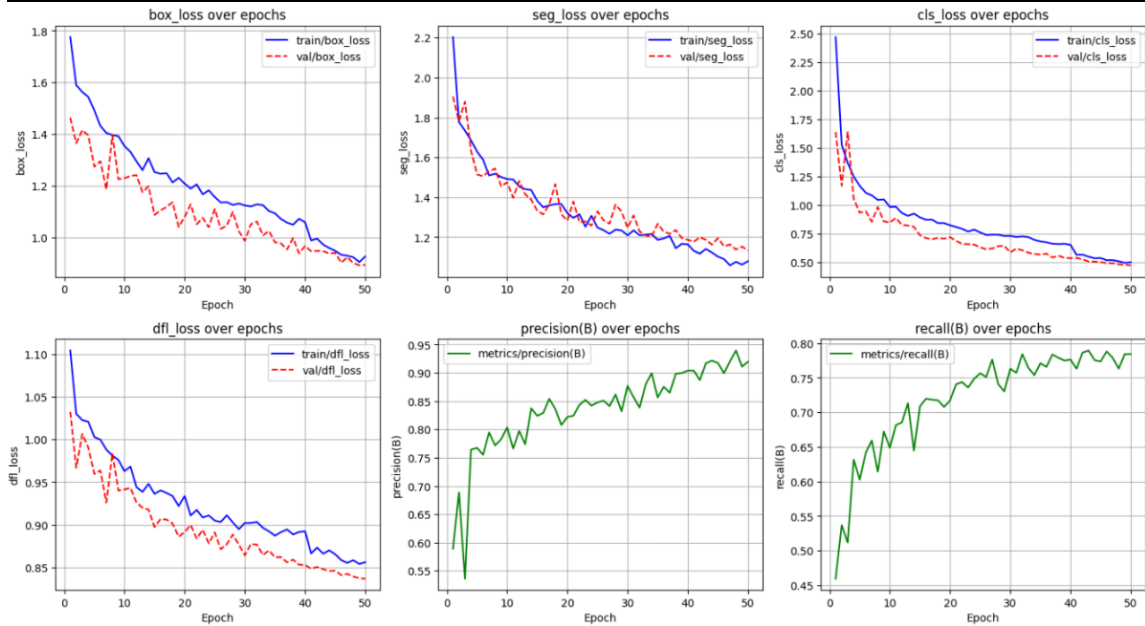


Рис.7. Графики метрик при обучении YOLOv11 small

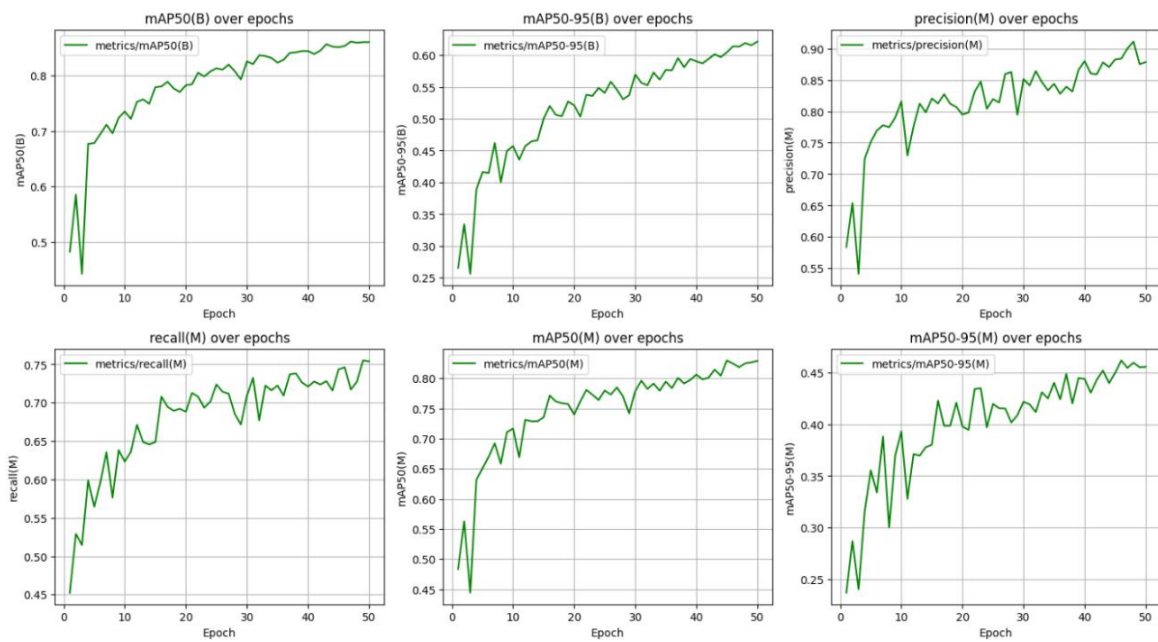


Рис.8 Графики метрик при обучении YOLOv11 small

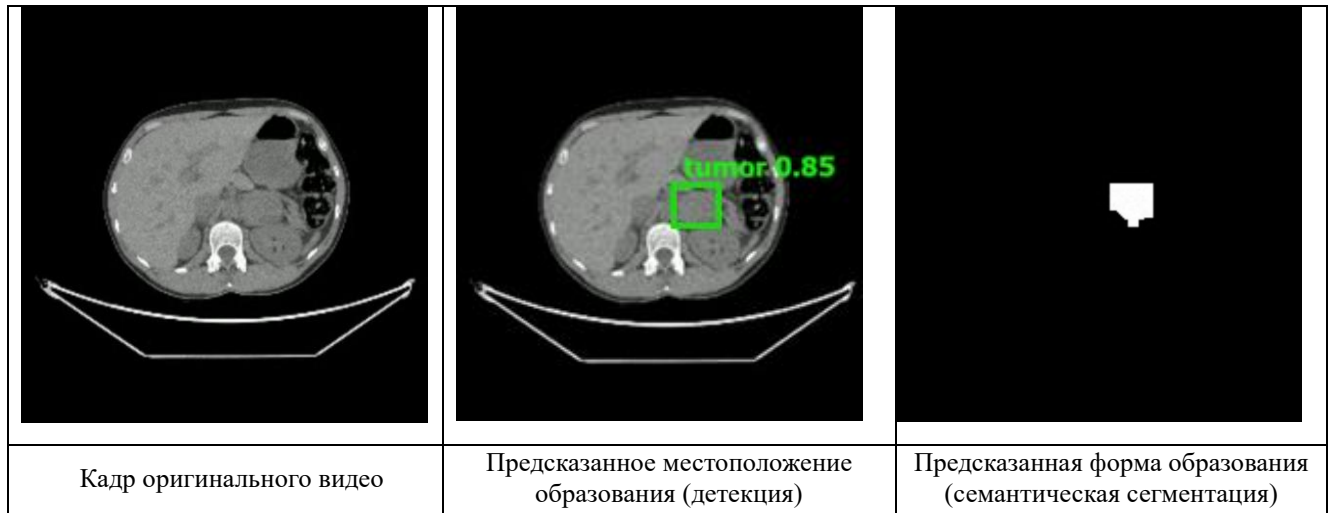


Рис.9 Пример результата применения модели YOLOv11 small

По проведенным исследованиям можно сделать следующие выводы: вариант “small+dropout” значительно увеличил precision, оставив изменения других метрик в пределах погрешности, это при определении опухоли дает меньше ложноположительных результатов. Итоговые метрики демонстрируют высокие результаты. Для сравнения, самая тяжеловесная по размеру модель YOLO имеет mAP50-95=0.547 в задаче сегментации на эталонном датасете, что заметно меньше, чем у полученной модели и может судить об успешности эксперимента по применению модели YOLOv11 для инстанс сегментации изображений с образованиями надпочечников.

## 5.2 Задача классификации

Для выбора модели классификации были обучены и протестированы семейства моделей, упомянутые в разделе 2 “сопутствующие работы”.

Для контроля обучения и отслеживания переобучения строились графики ошибки и качества по метрике f-score для валидационной и тренировочной выборок. Также по тестовой выборке строилась матрица ошибок (confusion matrix) и выводились метрики как по каждому классу, так и в микро/макро режимах.

В ходе экспериментов было установлено, что модель EfficientNet-B0 демонстрирует признаки переобучения уже на ранних этапах обучения и показывает крайне низкую способность к распознаванию второго и третьего классов. Модель ViViT также отличается высокой вычислительной затратностью и схожими ограничениями в точности классификации. В отличие от них, модель ResNet50 продемонстрировала устойчивую динамику снижения ошибки на валидационной выборке, а также лучшие значения ключевых метрик на тестовой выборке, что отражено в таблице 5.

Графики обучения представлен на рис. 10, а матрица ошибок - на рис. 11.

Таблица 5. Метрики качества моделей для задачи классификации

Модель	Precision	Recall	F1-Score
Vivit	0.55	0.69	0.58
ResNet18	0.51	0.51	0.51
ResNet34	0.70	0.78	0.74
ResNet50	0.86	0.94	0.90
ResNet101	0.80	0.88	0.84
EfficientNetB0	0.40	0.55	0.46
EfficientNetB3	0.65	0.72	0.68
EfficientNetB4	0.75	0.80	0.77

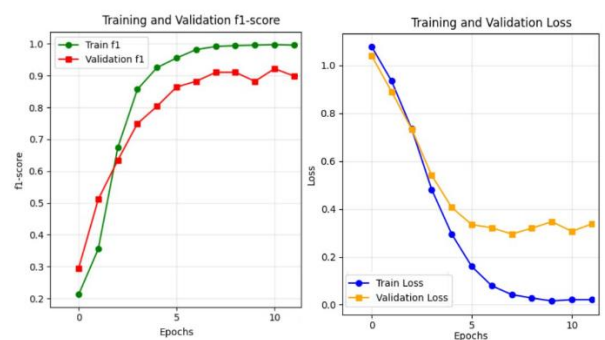


Рис. 10 Графики обучения модели ResNet50

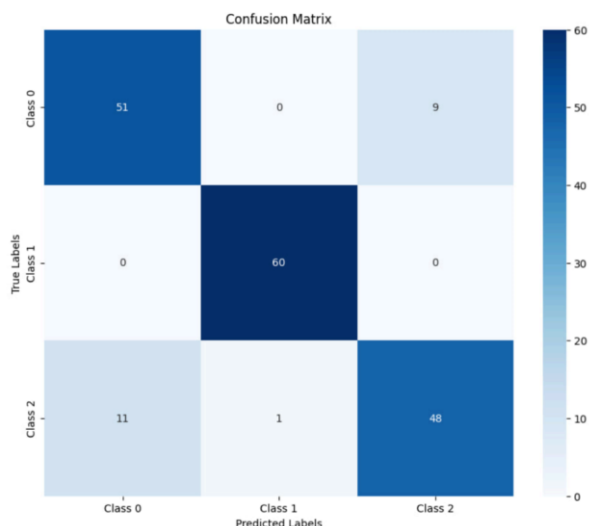


Рис. 11 Матрица ошибок модели ResNet50

### 5.3 Пайплайн

Мы предлагаем конвейер обработки данных (пайплайн), сочетающий задачи сегментации и классификации опухолевых образований на КТ-данных в видеоформате:

1) На первом этапе применяется модель инстанс сегментации, задачей которой является локализация и выделение новообразований на каждом кадре видео. Сегментация осуществляется на уровне отдельных изображений, извлечённых из видео, после чего результаты, относящиеся к одному

новообразованию, формируют единую маску. Полученные бинарные маски умножаются на исходные изображения, в результате чего на следующий этап — классификацию — поступают изображения, содержащие только выявленные образования. Это позволяет сфокусировать обучение классификатора на патологических зонах.

2) Классификатор, представленный архитектурой ResNet-50, обрабатывает видеоролики, соответствующие одному случаю опухоли, после чего делает предсказание класса фенотипа новообразования.

В случае, если на одном КТ-исследовании обнаружено несколько опухолей, каждая из них анализируется отдельно, что позволяет обеспечить точность идентификации и избежать конфликта между различными фенотипами.

КТ-снимки, не содержащие патологических изменений, исключаются на этапе сегментации, тем самым устраняется влияние неинформативных данных на этап классификации. Это решение позволяет повысить точность и устойчивость классификатора за счёт фокусировки на клинически значимых признаках.

Таким образом, предложенный пайплайн, представленный на рис. 12, обеспечивает модульное разделение задач локализации и классификации опухолей, позволяя использовать специализированные архитектуры нейросетей на каждом из этапов.

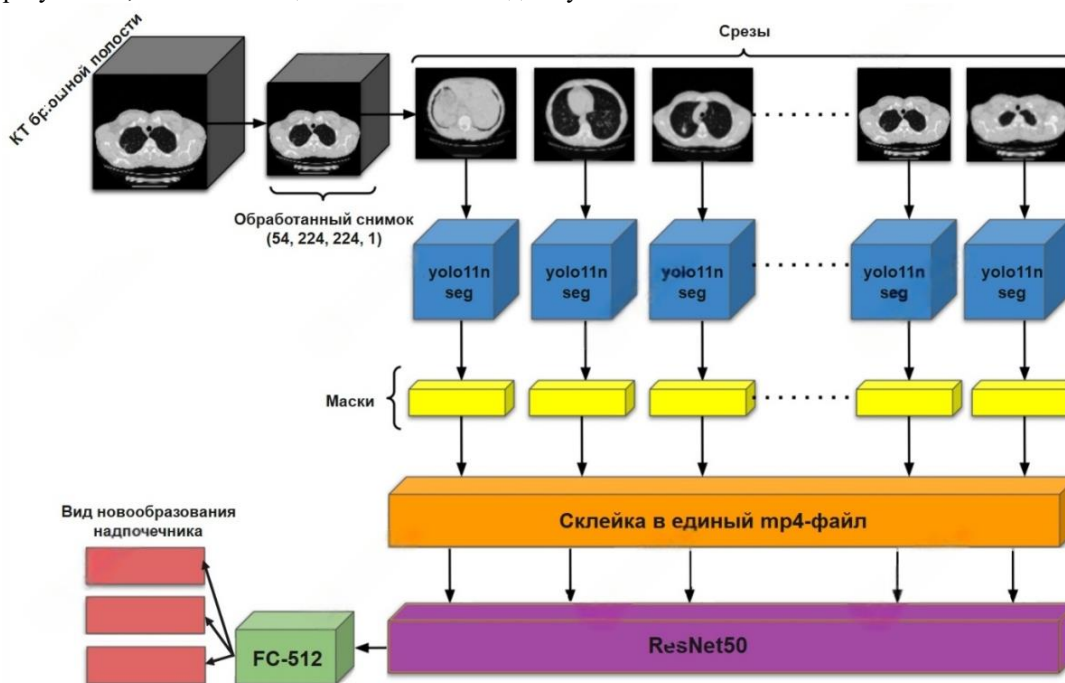


Рис. 12 Предлагаемый пайплайн анализа КТ-снимков

## 6. ЗАКЛЮЧЕНИЕ

В настоящем исследовании был предложен универсальный и воспроизводимый подход к

автоматизированному анализу опухолей надпочечников на КТ-данных с использованием современных архитектур глубокого обучения. Мы преодолели ограничение, связанное с

ограниченным объёмом доступных размеченных данных, предлагая использовать нейросетевые модели, предобученные на наборах данных достаточного масштаба и способных выходить за рамки их области применения.

Ключевой особенностью предложенного решения является двухэтапный подход, включающий предварительную сегментацию изображений с помощью модели YOLOv11-seg и последующую классификацию с применением 3D ResNet-50. Такой подход позволяет не только точно локализовать патологию, но и учитывать пространственные зависимости между срезами, что принципиально важно при работе с 3D-объектами. В отличие от традиционных методов, требующих ручной сегментации или анализа отдельных фрагментов изображения, мы реализовали полностью автоматизированный процесс, позволяющий сократить трудозатраты со стороны специалистов.

Предложенный подход обладает рядом принципиальных преимуществ:

- замена традиционных алгоритмов на современные нейросетевые решения с возможностью визуализации результатов;
- отказ от анализа фиксированных патчей в пользу обработки всей последовательности КТ-срезов;
- анализ нескольких новообразований в рамках одного КТ-исследования;
- универсальность модели сегментации;
- простота внедрения в медицинские информационные системы (МИС).

Таким образом, данная работа демонстрирует перспективность применения новых нейросетевых архитектур в медицинской диагностике и может послужить основой для дальнейших исследований по автоматизации анализа КТ-снимков с редкими или сложно классифицируемыми патологиями даже в условиях ограниченности данных.

## 7. БИБЛИОГРАФИЯ

- [1] Tobias Carling «How is adrenal cancer diagnosed?» [Electronic resource]. Available: <https://www.adrenal.com/blog/how-is-adrenal-cancer-diagnosed>.
- [2] Андреева А.В. Подход врача общей практики к ведению пациентов с инцидентомами надпочечников. *Терапевт*, 2019, № 12 с. 70-78.
- [3] Adrenal incidentalomas: imaging challenges—role of MDCT scan versus MRI in evaluating adrenal incidentalomas. [Electronic resource]. Available: <https://ejnm.springeropen.com/articles/10.1186/s43055-021-00437-w>.
- [4] U-Net: Convolutional Networks for Biomedical Image Segmentation. [Electronic resource]. Available: [https://link.springer.com/chapter/10.1007/978-3-319-24574-4\\_28](https://link.springer.com/chapter/10.1007/978-3-319-24574-4_28).
- [5] Y-Net: Joint Segmentation and Classification for Diagnosis of Breast Biopsy Images [Electronic resource]. Available: [https://link.springer.com/chapter/10.1007/978-3-030-00934-2\\_99](https://link.springer.com/chapter/10.1007/978-3-030-00934-2_99).
- [6] Hybrid Deep Learning Framework for Classification of Kidney CT. [Electronic resource]. Available: [arXiv:2502.04367v1](https://arxiv.org/abs/2502.04367v1).
- [7] Accurate classification of lung nodules on CT images using the TransUnet. [Electronic resource]. Available: <https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2022.1060798/full>.
- [8] Using CT radiomic features based on machine learning models to subtype adrenal adenoma. [Electronic resource]. Available: <https://link.springer.com/article/10.1186/s12885-023-10562-6>.
- [9] Computed Tomography-Based Machine Learning Differentiates Adrenal Pheochromocytoma From Lipid-Poor Adenoma. [Electronic resource]. Available: [https://www.frontiersin.org/journals/endocrinology/articles/10.3389/fendo.2022.833413/full?utm\\_source=chatgpt.com](https://www.frontiersin.org/journals/endocrinology/articles/10.3389/fendo.2022.833413/full?utm_source=chatgpt.com).
- [10] Machine learning-based quantitative texture analysis of CT images of small renal masses: Differentiation of angiomyolipoma without visible fat from renal cell carcinoma. [Electronic resource]. Available: <https://link.springer.com/article/10.1007/s00330-017-5118-z?fromPaywallRec=true>.
- [11] Machine learning-based texture analysis for differentiation of radiologically indeterminate small adrenal tumors on adrenal protocol CT scans. [Electronic resource]. Available: <https://link.springer.com/article/10.1007/s00261-021-03136-2>.
- [12] GitHub - rafiibnsultan/3D-CT-Image-Segmentation-Using-MONAI. [Electronic resource]. Available: <https://github.com/rafiibnsultan/3D-CT-Image-Segmentation-Using-MONAI>.
- [13] Machine Learning for Adrenal Gland Segmentation and Classification of Normal and Adrenal Masses at CT. [В Интернете]. Available: <https://pubs.rsna.org/doi/full/10.1148/radiol.220101>.
- [14] Diagnostic Accuracy of CT Texture Analysis in Adrenal Masses: A Systematic Review. [Electronic resource]. Available: <https://www.mdpi.com/1422-0067/23/2/637>.
- [15] Video Pretraining Advances 3D Deep Learning on Chest CT Tasks. [Electronic resource]. Available: <https://arxiv.org/html/2304.00546>.
- [16] EfficientNet family U-Net models for deep learning semantic segmentation of kidney tumors on CT images. [Electronic resource]. Available: <https://www.frontiersin.org/articles/10.3389/fcomp.2023.1235622/full>.
- [17] Multi-Scale Supervised 3D U-Net for Kidneys and Kidney Tumor Segmentation. [Electronic resource]. Available: <https://arxiv.org/abs/2004.08108>.
- [18] FPN-SE-ResNet Model for Accurate Diagnosis of Kidney Tumors Using CT Images. [Electronic resource]. Available: <https://www.mdpi.com/2076-3417/13/17/9802>.
- [19] Liver Tumor Localization Based on YOLOv3 and 3D-Semantic Segmentation Using Deep Neural Networks. [Electronic resource]. Available: <https://pubmed.ncbi.nlm.nih.gov/35453870/>.
- [20] An attempt at beating the 3D U-Ne. [Electronic resource]. Available: <https://arxiv.org/abs/1908.02182>.

- [21] UniverSeg: Universal Medical Image Segmentation. [Electronic resource]. Available: <https://vitalab.github.io/article/2024/05/24/universeg.html>.
- [22] Automated Adrenal Gland Disease Classes Using Patch-Based Center Symmetric Local Binary Pattern Technique with CT Images. [Electronic resource]. Available: <https://link.springer.com/article/10.1007/s10278-022-00759-9?fromPaywallRec=true>.
- [23] CT-based deep learning model for the prediction of DNA mismatch repair deficient colorectal cancer: a diagnostic study. [Electronic resource]. Available: <https://pubmed.ncbi.nlm.nih.gov/36949511/>.
- [24] A transfer learning based deep learning model to diagnose covid-19 CT scan images. [Electronic resource]. Available: [https://pmc.ncbi.nlm.nih.gov/articles/PMC9177227/?utm\\_source=chatgpt.com](https://pmc.ncbi.nlm.nih.gov/articles/PMC9177227/?utm_source=chatgpt.com).
- [25] Detecting adrenal lesions on 3D CT scans using a 2.5D deep learning model. [Electronic resource]. Available: <https://www.medrxiv.org/content/10.1101/2023.02.22.23286184v1>.
- [26] 3D Deep Learning from CT Scans Predicts Tumor Invasiveness of Subcentimeter Pulmonary Adenocarcinomas. [Electronic resource]. Available: <https://aacrjournals.org/cancerres/article/78/24/6881/633067/3D-Deep-Learning-from-CT-Scans-Predicts-Tumor>.
- [27] Boosting Learning by Representing 3D Medical Imaging to 2D Features for Small Data. [Electronic resource]. Available: <https://arxiv.org/html/2002.04251v3?>.
- [28] 3D Deep Learning on Medical Images: A Review. [Electronic resource]. Available: <https://arxiv.org/abs/2004.00218>.
- [29] Ultralytics YOLO11,» [Electronic resource]. Available: <https://docs.ultralytics.com/models/yolo11/>.
- [30] 3D radiotherapy dose prediction on head and neck cancer patients with a hierarchically densely connected U-net deep learning architecture. [Electronic resource]. Available: <https://iopscience.iop.org/article/10.1088/1361-6560/ab039b/meta>.
- [31] ViPTT-Net: Video pretraining of spatio-temporal model for tuberculosis type classification from chest CT scans. [Electronic resource]. Available: <https://arxiv.org/abs/2105.12810>.
- [32] Vision Transformers: State of the Art and Research Challenges. [Electronic resource]. Available: <https://arxiv.org/abs/2207.03041>.

Статья получена 30 июля 2025.

Приваленко Антон Александрович, Национальный Исследовательский Ядерный Университет МИФИ, магистрант, [privalenkoan@gmail.com](mailto:privalenkoan@gmail.com)

Зайцев Константин Сергеевич, Национальный Исследовательский Ядерный Университет МИФИ, профессор [kszajtsev@mephi.ru](mailto:kszajtsev@mephi.ru)

Бурякина Светлана Алексеевна, ФГБУ «НМИЦ эндокринологии» Минздрава России, врач-рентгенолог референс-центра лучевых методов диагностики [sburyakina@yandex.ru](mailto:sburyakina@yandex.ru)

Тарабаева Наталья Викторовна, ФГБУ «НМИЦ эндокринологии» Минздрава России, руководитель референс-центра лучевых методов диагностики, кандидат медицинских наук, [Ntarbaeva@inbox.ru](mailto:Ntarbaeva@inbox.ru)

Китаев Степан Максимович, Национальный Исследовательский Ядерный Университет МИФИ, магистрант, [stepan.kitaevv@yandex.ru](mailto:stepan.kitaevv@yandex.ru)

Кухтиков Кирилл Владиславович, Национальный Исследовательский Ядерный Университет МИФИ, бакалавр, [kkv013@campus.mephi.ru](mailto:kkv013@campus.mephi.ru)

# A Neural Network Approach for the Analysis of Computed Tomography Images in Adrenal Gland Diseases

A.A. Privalenko, K.S. Zaytsev, S.A. Buryakina, N.V. Tarbaeva, S.M. Kitaev, K.V. Kukhtikov

**Abstract - This study investigates the potential of deep learning models for the diagnosis of adrenal gland diseases using computed tomography (CT) images. To support clinical decision-making and automate the classification of identified neoplasms in abdominal CT scans, a two-stage neural network approach, combining segmentation and classification, was developed. This approach allows for the visualization of detected lesions and accommodates multiple lesions within a single CT image. The study utilized a dataset provided by the Academician I.I. Dedov National Medical Research Center of Endocrinology, comprising 228 CT scans at the time of writing. To optimize processing time, images were converted to MP4 video format (54 frames per video), reducing data volume without significantly compromising diagnostic value. Images underwent preprocessing, and data augmentation was employed to address class imbalance. Each CT scan was annotated with three labels, corresponding to the presence of a neoplasm with a malignant, benign, or indeterminate phenotype. For lesion instance segmentation, a YOLOv11-seg model pre-trained on the COCO dataset was implemented. A 3DResNet-50 model, trained on the segmented regions, was used for classification. The proposed combined two-stage approach is implemented in a software suite designated “Assistant Endocrinologist”.**

**Keywords:** Neural Networks, Segmentation, Classification, Adrenal Gland Disease, CT Images

## REFERENCES

- [1] Tobias Carling «How is adrenal cancer diagnosed?» [Electronic resource]. Available: <https://www.adrenal.com/blog/how-is-adrenal-cancer-diagnosed>.
- [2] Andreeva A.V. Podhod vracha obshhej praktiki k vedeniju pacientov s incidentalomami nadpochechnikov. *Terapevt*, 2019, # 12 s. 70-78.
- [3] Adrenal incidentalomas: imaging challenges—role of MDCT scan versus MRI in evaluating adrenal incidentalomas. [Electronic resource]. Available: <https://ejrn.springeropen.com/articles/10.1186/s43055-021-00437-w>.
- [4] U-Net: Convolutional Networks for Biomedical Image Segmentation. [Electronic resource]. Available: [https://link.springer.com/chapter/10.1007/978-3-319-24574-4\\_28](https://link.springer.com/chapter/10.1007/978-3-319-24574-4_28).
- [5] Y-Net: Joint Segmentation and Classification for Diagnosis of Breast Biopsy Images [Electronic resource]. Available: [https://link.springer.com/chapter/10.1007/978-3-030-00934-2\\_99](https://link.springer.com/chapter/10.1007/978-3-030-00934-2_99).
- [6] Hybrid Deep Learning Framework for Classification of Kidney CT. [Electronic resource]. Available: [arXiv:2502.04367v1](https://arxiv.org/abs/2502.04367v1).
- [7] Accurate classification of lung nodules on CT images using the TransUnet. [Electronic resource]. Available: <https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2022.1060798/full>.
- [8] Using CT radiomic features based on machine learning models to subtype adrenal adenoma. [Electronic resource]. Available: <https://link.springer.com/article/10.1186/s12885-023-10562-6>.
- [9] Computed Tomography-Based Machine Learning Differentiates Adrenal Pheochromocytoma From Lipid-Poor Adenoma. [Electronic resource]. Available: [https://www.frontiersin.org/journals/endocrinology/articles/10.3389/fendo.2022.833413/full?utm\\_source=chatgpt.com](https://www.frontiersin.org/journals/endocrinology/articles/10.3389/fendo.2022.833413/full?utm_source=chatgpt.com).
- [10] Machine learning-based quantitative texture analysis of CT images of small renal masses: Differentiation of angiomyolipoma without visible fat from renal cell carcinoma. [Electronic resource]. Available: <https://link.springer.com/article/10.1007/s00330-017-5118-z?fromPaywallRec=true>.
- [11] Machine learning-based texture analysis for differentiation of radiologically indeterminate small adrenal tumors on adrenal protocol CT scans. [Electronic resource]. Available: <https://link.springer.com/article/10.1007/s00261-021-03136-2>.
- [12] GitHub - rafiibnsultan/3D-CT-Image-Segmentation-Using-MONAI. [Electronic resource]. Available: <https://github.com/rafiibnsultan/3D-CT-Image-Segmentation-Using-MONAI>.
- [13] Machine Learning for Adrenal Gland Segmentation and Classification of Normal and Adrenal Masses at CT. [В Интернете]. Available: <https://pubs.rsna.org/doi/full/10.1148/radiol.220101>.
- [14] Diagnostic Accuracy of CT Texture Analysis in Adrenal Masses: A Systematic Review. [Electronic resource]. Available: <https://www.mdpi.com/1422-0067/23/2/637>.
- [15] Video Pretraining Advances 3D Deep Learning on Chest CT Tasks. [Electronic resource]. Available: <https://arxiv.org/html/2304.00546>.
- [16] EfficientNet family U-Net models for deep learning semantic segmentation of kidney tumors on CT images. [Electronic resource]. Available: <https://arxiv.org/html/2304.00546>.

- <https://www.frontiersin.org/articles/10.3389/fcomp.2023.1235622/full>.
- [17] Multi-Scale Supervised 3D U-Net for Kidneys and Kidney Tumor Segmentation. [Electronic resource]. Available: <https://arxiv.org/abs/2004.08108>.
- [18] FPN-SE-ResNet Model for Accurate Diagnosis of Kidney Tumors Using CT Images. [Electronic resource]. Available: <https://www.mdpi.com/2076-3417/13/17/9802>.
- [19] Liver Tumor Localization Based on YOLOv3 and 3D-Semantic Segmentation Using Deep Neural Networks. [Electronic resource]. Available: <https://pubmed.ncbi.nlm.nih.gov/35453870/>.
- [20] An attempt at beating the 3D U-Ne. [Electronic resource]. Available: <https://arxiv.org/abs/1908.02182>.
- [21] UniverSeg: Universal Medical Image Segmentation. [Electronic resource]. Available: <https://vitalab.github.io/article/2024/05/24/universeg.html>.
- [22] Automated Adrenal Gland Disease Classes Using Patch-Based Center Symmetric Local Binary Pattern Technique with CT Images. [Electronic resource]. Available: <https://link.springer.com/article/10.1007/s10278-022-00759-9?fromPaywallRec=true>.
- [23] CT-based deep learning model for the prediction of DNA mismatch repair deficient colorectal cancer: a diagnostic study. [Electronic resource]. Available: <https://pubmed.ncbi.nlm.nih.gov/36949511/>.
- [24] A transfer learning based deep learning model to diagnose covid-19 CT scan images. [Electronic resource]. Available: [https://pmc.ncbi.nlm.nih.gov/articles/PMC9177227/?utm\\_source=chatgpt.com](https://pmc.ncbi.nlm.nih.gov/articles/PMC9177227/?utm_source=chatgpt.com).
- [25] Detecting adrenal lesions on 3D CT scans using a 2.5D deep learning model. [Electronic resource]. Available: <https://www.medrxiv.org/content/10.1101/2023.02.22.23286184v1>.
- [26] 3D Deep Learning from CT Scans Predicts Tumor Invasiveness of Subcentimeter Pulmonary Adenocarcinomas. [Electronic resource]. Available: <https://aacrjournals.org/cancerres/article/78/24/6881/633067/3D-Deep-Learning-from-CT-Scans-Predicts-Tumor>.
- [27] Boosting Learning by Representing 3D Medical Imaging to 2D Features for Small Data. [Electronic resource]. Available: <https://arxiv.org/html/2002.04251v3?>.
- [28] 3D Deep Learning on Medical Images: A Review. [Electronic resource]. Available: <https://arxiv.org/abs/2004.00218>.
- [29] Ultralytics YOLO11. [Electronic resource]. Available: <https://docs.ultralytics.com/models/yolo11/>.
- [30] 3D radiotherapy dose prediction on head and neck cancer patients with a hierarchically densely connected U-net deep learning architecture. [Electronic resource]. Available: <https://iopscience.iop.org/article/10.1088/1361-6560/ab039b/meta>.
- [31] ViPTT-Net: Video pretraining of spatio-temporal model for tuberculosis type classification from chest CT scans. [Electronic resource]. Available: <https://arxiv.org/abs/2105.12810>.
- [32] Vision Transformers: State of the Art and Research Challenges. [Electronic resource]. Available: <https://arxiv.org/abs/2207.03041>.