

# Модель нейронной сети для диагностики заболеваний желудочно-кишечного тракта по эндоскопическим изображениям

М. А. Моураов

Санкт-Петербургский  
государственный  
электротехнический  
университет «ЛЭТИ»  
им. В.И. Ульянова (Ленина)  
mmouraov2@gmail.com

Е. Ю. Белова

Санкт-Петербургский  
государственный  
электротехнический  
университет «ЛЭТИ»  
им. В.И. Ульянова (Ленина)  
eyshukeylo@gmail.com

В. Ч. Тедтоева

Северо-Западный  
государственный  
медицинский университет  
им. И.И. Мечникова  
tedtoevaviktoria@gmail.com

**Аннотация.** Статья посвящена созданию нейросетевой модели для автоматизированной диагностики патологий пищеварительной части желудочно-кишечного тракта на основе глубокого обучения. Описана архитектура модели, адаптированной для анализа изображений капсульной эндоскопии, и этапы предобработки данных датасета Kvasir-Capsule. Приведены результаты валидации модели: достигнуты показатели  $mAP@0,5 = 0,905$  и  $mAP@0,5:0,95 = 0,688$  при скорости обработки 244 кадра/с.

**Ключевые слова:** глубокое обучение; эндоскопия; диагностика заболеваний ЖКТ; нейронные сети; компьютерное зрение; Kvasir-Capsule

## I. ВВЕДЕНИЕ

Заболевания пищеварительной части желудочно-кишечного тракта (ЖКТ) остаются одной из ключевых проблем здравоохранения. По данным Всемирной организации здравоохранения, ежегодно рак желудка диагностируется более чем у миллиона человек по всему миру, причем около 760 тысяч случаев заканчиваются летальным исходом [4]. В России, согласно статистике Министерства здравоохранения за 2023 год, зафиксировано порядка 50 тысяч новых случаев заболеваний желудка и кишечника, включая злокачественные опухоли, полипы и хронические воспаления [5].

Несмотря на активное применение методов компьютерного зрения, существующие решения обладают рядом ограничений: снижение точности при наличии артефактов капсульной эндоскопии, недостаточная эффективность детекции мелких патологий (<5 мм) и низкая обобщающая способность моделей при применении к данным из других клинических центров. Кроме того, большинство публикаций фокусируются на задачах классификации или сегментации, тогда как задача точной пространственной детекции с учётом ориентации вытянутых структур (ангиоэктазии, линейные кровоизлияния) остаётся недостаточно проработанной. В связи с этим актуальна разработка специализированной детекторной модели, устойчивой к клиническим артефактам и обеспечивающей высокую точность локализации.

Для выявления патологий ЖКТ применяют несколько подходов. Наиболее распространены эндоскопические методы, такие как гастроскопия и колоноскопия, которые позволяют врачу осмотреть внутреннюю поверхность желудка и кишечника [6]. Однако такие исследования требуют высокой квалификации специалиста, по-

скольку анализ изображений проводится вручную, что иногда приводит к ошибкам. В последние годы активно применяются методы искусственного интеллекта, в частности сверточные нейронные сети, которые позволяют автоматизировать анализ эндоскопических изображений, увеличить точность и скорость выявления патологии [7].

Анализ научных публикаций за 2020–2025 гг. показывает, что для диагностики заболеваний ЖКТ наиболее широко применяются три класса задач компьютерного зрения: классификация, детекция и семантическая сегментация [8].

Для классификации эндоскопических изображений эффективно используются архитектуры ResNet, EfficientNet и DenseNet. В работе [9] модель ResNet-152 на датасете Kvasir достигла точности 90% для бинарной классификации «норма/патология». Однако такие модели чувствительны к качеству изображений: при наличии артефактов (размытие, неравномерное освещение) точность снижается до 75–80% [10].

Задачи детекции патологий (локализация полипов, язв, ангиоэктазий) решаются с помощью одностадийных детекторов семейства YOLO и двухстадийных архитектур типа Faster R-CNN. В исследовании [11] модель YOLOv5 показала  $mAP@0.5 = 0,89$  на наборе CVC-ClinicDB при скорости 30 кадров/с, что делает её перспективной для анализа видеопотока в реальном времени. Тем не менее, детекция мелких объектов (<5 мм) остаётся проблемной из-за ограниченного разрешения изображений.

Для сегментации границ патологических участков стандартом де-факто является архитектура U-Net и её модификации. В работе [12] базовая U-Net достигла IoU = 0,82 на датасете Kvasir-SEG, тогда как улучшенная версия UNet++ показала IoU = 0,87 за счёт вложенных плотных соединений [13].

Цель работы — разработка и валидация нейросетевой модели для комплексной диагностики патологий ЖКТ по эндоскопическим изображениям капсульной эндоскопии.

Задачи исследования:

1. Проанализировать существующие подходы к автоматизированной диагностике заболеваний ЖКТ.

2. Адаптировать архитектуру YOLOv8 OBB для детекции патологий с учётом ориентации объектов.

3. Провести предобработку и балансировку данных датасета Kvasir-Capsule.

4. Обучить и валидировать модель, оценить полученные результаты по заданным метрикам качества.

## II. МАТЕРИАЛЫ И МЕТОДЫ

### A. Набор данных

а) Исследование выполнено на основе набора данных Kvasir-Capsule [15], содержащего 118 видеозаписей капсульной эндоскопии (около 4,8 млн кадров), из которых размечено 44 228 изображений. Разметка включает 13 классов патологий и анатомических ориентиров: *Normal Mucosa, Pylorus, Ileo-Cecal Valve, Blood, Angiectasia, Ulcer, Polyp, Erosion, Hematin, Lymphoid Hyperplasia, Chyle, Reduced Mucosal View, Erythematous*.

Датасет включает видеозаписи от 117 пациентов, прошедших капсульную эндоскопию в период с февраля 2016 по январь 2018 года в отделении медицины больницы Бэрум (Вестре Викен, Норвегия). Детальная демографическая информация (распределение по полу и возрасту) не приводится в открытом описании датасета в целях соблюдения требований анонимизации и нормативных актов Норвегии и GDPR. Все исследования выполнены с использованием системы капсульной эндоскопии Olympus Endocapsule 10, включая капсулу Olympus EC-S10 и рекордер Olympus RE-10; видеозаписи осуществлялись с частотой 2 кадра в секунду в разрешении 336×336 пикселей с кодированием H.264.

### B. Предобработка данных для исследования

Исходный датасет Kvasir-Capsule характеризуется выраженным дисбалансом классов: класс *Normal Mucosa* составляет 78,2% размеченных изображений, тогда как редкие патологии (например, *Polyp, Hematin*) представлены менее чем 0,3% [15]. Для обеспечения репрезентативности выборки при ограниченных вычислительных ресурсах был применён метод пропорционального сокращения: каждый класс уменьшен в одинаковой пропорции до общего объёма 12 000 изображений с сохранением клинически значимых патологий.

Данные разделены на обучающую (80%) и тестовую (20%) выборки с соблюдением принципа видеоразнообразия кросс-валидации: все кадры из одного видео принадлежат только одной подвыборке, что исключает эффект data leakage.

Предобработка изображений включала:

- изменение размера до 640×640 пикселей с сохранением пропорций;
- нормализацию значений пикселей с использованием статистик ImageNet;
- аугментацию данных с применением библиотеки Albumentations: геометрические трансформации (повороты  $\pm 15^\circ$ , отражения), фотометрические изменения (яркость  $\pm 20\%$ , контраст  $\pm 20\%$ ), имитация артефактов

капсульной эндоскопии (размытие по Гауссу, CLAHE для коррекции освещения).

Для компенсации дисбаланса классов использована комбинация взвешенной выборки (weighted sampling) и переоценки весов классов в функции потерь.

### C. Архитектура модели и алгоритм исследования

Для решения задачи одновременной детекции и классификации патологий выбрана модифицированная архитектура YOLOv8 OBB (Oriented Bounding Boxes). Ключевые особенности адаптации для медицинской визуализации:

1. Увеличенное входное разрешение (640×640 вместо стандартных 336×336) для сохранения деталей мелких патологий.
2. Голова OBB, генерирующая не только координаты ограничивающей рамки, но и угол её вращения, что позволяет точно описывать вытянутые структуры (ангиоэктазии, линейные кровоизлияния).
3. Механизм визуализации внимания на основе модифицированного Grad-CAM для повышения интерпретируемости результатов.

Обучение модели проводилось с использованием оптимизатора AdamW (начальная скорость обучения 0,001, cosine decay, weight decay =  $5 \times 10^{-4}$ ), размер батча — 32 изображения. Применена стратегия early stopping с терпением 10 эпох для предотвращения переобучения.

Алгоритм исследования включал следующие этапы: (1) загрузка и пропорциональное сокращение датасета до 12 000 изображений с сохранением клинического баланса; (2) применение аугментации и нормализации; (3) обучение модифицированной архитектуры YOLOv8 OBB с параметрами AdamW (lr=0,001, batch=32, early stopping=10 эпох); (4) валидация на независимой тестовой выборке (20%) и расчёт метрик.

### D. Метрики оценки

Для оценки качества модели использованы стандартные метрики задач детекции:

- mAP@0.5 — средняя точность при пороге IoU = 0,5;
- mAP@0.5:0.95 — усреднённая mAP по порогам IoU от 0,5 до 0,95 с шагом 0,05;
- Precision (точность) и Recall (полнота);
- F1-score (macro) — гармоническое среднее точности и полноты, усреднённое по классам;
- MCC (Matthews Correlation Coefficient) — устойчивая метрика для несбалансированных данных.

Дополнительно рассчитано время инференса на одном изображении для оценки применимости в реальном времени.

## III. РЕЗУЛЬТАТЫ

Обучение модели проводилось на рабочей станции с GPU NVIDIA RTX 3090 (24 ГБ видеопамяти) в течение 27 эпох (ранняя остановка). Общее время обучения составило 112 минут.

ТАБЛИЦА 1. ОСНОВНЫЕ МЕТРИКИ КАЧЕСТВА МОДЕЛИ НА ТЕСТОВОЙ ВЫБОРКЕ (2400 ИЗОБРАЖЕНИЙ)

Метрика	Значение
mAP@0.5	0,905
mAP@0.5:0.95	0,688
Precision	0,892
Recall	0,827
F1-score (macro)	0,836
MCC	0,820
Время инференса, мс	4,1

Результаты подтверждают высокую эффективность модели: значение  $mAP@0.5 = 0,905$  существенно превышает порог 0,80, принятый в качестве минимального требования для клинического применения [16]. Показатель  $mAP@0.5:0.95 = 0,688$  свидетельствует о точной локализации объектов различной формы и размера.

Анализ по классам показал наилучшие результаты для клинически значимых патологий:

- *Blood* (кровотечение):  $mAP@0.5 = 0,918$ ;
- *Angiectasia* (ангиоэктазии):  $mAP@0.5 = 0,937$ ;
- *Ulcer* (язвы):  $mAP@0.5 = 0,872$ .

Анатомические ориентиры (*Pylorus*, *Ileo-Cecal valve*) детектируются с точностью  $>0,95$ , что важно для пространственной привязки патологий.

Скорость обработки 244 кадра/с (4,1 мс на изображение) позволяет анализировать полное исследование капсульной эндоскопии (~50 000 кадров) менее чем за 5 минут, что в 9–18 раз быстрее ручного анализа врачом.

Визуализация результатов (рис. 1) демонстрирует корректную детекцию патологий различной морфологии: модель точно определяет ориентированные рамки для вытянутых ангиоэктазий и выделяет мелкие кровозлияния на фоне нормальной слизи.

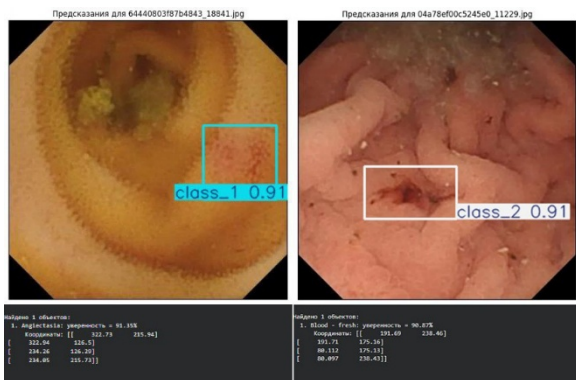


Рис. 1. Примеры детекции патологий: а) ангиоэктазии; б) кровотечение. Ограничительные рамки — предсказания модели

#### IV. ЗАКЛЮЧЕНИЕ

В ходе исследования были выполнены все поставленные задачи. Проанализированы существующие подходы к детекции патологий ЖКТ, адаптирована архитектура YOLOv8 ОБВ для учёта ориентации объектов, проведена балансировка и предобработка датасета Kvasir-Capsule, модель успешно обучена и провалидирована. Достигнутые метрики ( $mAP@0,5 = 0,905$ , скорость 244 кадра/с) подтверждают эффективность предложенного решения для

автоматизированного анализа капсульной эндоскопии. К ограничениям подхода относится снижение точности на кадрах с выраженными артефактами (пузыри, слизь), что требует внедрения модулей предварительного улучшения качества изображений. Также для повышения клинического доверия необходима дальнейшая интеграция механизмов объяснимого ИИ (XAI). Внедрение разработанной модели позволит сократить время анализа исследований, повысить точность диагностики и снизить нагрузку на врачей-эндоскопистов.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] Федеральный закон от 21 ноября 2011 г. № 323-ФЗ "Об основах охраны здоровья граждан в Российской Федерации".
- [2] Приказ Министерства здравоохранения РФ от 5 августа 2022 г. N 530н "Об утверждении унифицированных форм медицинской документации, используемых в медицинских организациях, оказывающих медицинскую помощь в стационарных условиях, в условиях дневного стационара и порядков их ведения" (с изменениями и дополнениями).
- [3] Приказ Министерства здравоохранения РФ от 7 сентября 2020 г. N 947н "Об утверждении Порядка организации системы документооборота в сфере охраны здоровья в части ведения медицинской документации в форме электронных документов".
- [4] Sung H., Ferlay J., Siegel R. L. et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries // CA: A Cancer Journal for Clinicians. 2021. vol. 71, № 3. pp. 209–249. DOI: 10.3322/caac.21660
- [5] Министерство здравоохранения Российской Федерации. Статистика заболеваемости злокачественными новообразованиями в России за 2023 год. М., 2024. 45 с.
- [6] ASGE Standards of Practice Committee. Endoscopic diagnosis of gastrointestinal diseases // Gastrointestinal Endoscopy. 2021. vol. 93, № 6. pp. 1245–1255. DOI: 10.1016/j.gie.2020.11.013
- [7] LeCun Y., Bengio Y., Hinton G. Deep learning // Nature. 2020. vol. 521. pp. 436–444. DOI: 10.1038/nature14539
- [8] Zhou Z., Siddiquee M. M. R., Tajbakhsh N., Liang J. UNet++: A nested U-Net architecture for medical image segmentation // IEEE Transactions on Medical Imaging. 2020. vol. 39, № 6. pp. 1856–1867. DOI: 10.1109/TMI.2019.2959609
- [9] Pogorelov K., Randel K. R., Griwodz C. et al. Kvasir: A multi-class image dataset for computer-aided gastrointestinal disease detection // Proc. 8th ACM Multimedia Systems Conf. 2020. pp. 164–169. DOI: 10.1145/3339825.3394925
- [10] Zhang R., Zheng Y., Poon C. C. Y. Polyp detection in colonoscopy images using deep learning // IEEE Transactions on Medical Imaging. 2022. vol. 41, № 7. pp. 1902–1913. DOI: 10.1109/TMI.2022.3145762
- [11] Borgli H., Thambawita V., Smedsrud P. H. et al. HyperKvasir: A comprehensive multi-class image and video dataset for gastrointestinal endoscopy // Scientific Data. 2020. vol. 7, № 1. p. 283. DOI: 10.1038/s41597-020-00622-y
- [12] Jha D., Smedsrud P. H., Riegler M. A. Kvasir-SEG: A segmented polyp dataset // Lecture Notes in Computer Science. 2020. vol. 12261. pp. 451–460. DOI: 10.1007/978-3-030-59710-7\_44
- [13] Ronneberger O., Fischer P., Brox T. U-Net: Convolutional networks for biomedical image segmentation // Int. Conf. on Medical Image Computing and Computer-Assisted Intervention. 2020. pp. 234–241. DOI: 10.1007/978-3-319-24574-4\_28
- [14] Thambawita V., Halvorsen P., Johansen D. et al. An extensive study on cross-dataset bias and evaluation metrics interpretation for machine learning applied to gastrointestinal tract abnormality classification // ACM Transactions on Computing for Healthcare. 2020. vol. 1, № 2. pp. 1–22. DOI: 10.1145/3386295
- [15] Smedsrud P. H., Gjestang H., Nedrejord O. O. et al. Kvasir-Capsule, a video capsule endoscopy dataset // Scientific Data. 2020. vol. 7, № 1. p. 283. DOI: 10.1038/s41597-020-00622-y
- [16] Leenhardt R., Vasseur P., Li C. et al. A neural network algorithm for detection of GI angiectasia during small-bowel capsule endoscopy // Gastrointestinal Endoscopy. 2019. vol. 89, № 1. pp. 189–194. DOI: 10.1016/j.gie.2018.09.001