

УДК 004.8:616.3

DOI: [10.26102/2310-6018/2026.60.9.014](https://doi.org/10.26102/2310-6018/2026.60.9.014)

Разработка моделей машинного обучения для прогнозирования заболеваний желудочно-кишечного тракта

М.Р.К. Альмусави✉, Е.В. Ляпунцова

*Национальный исследовательский технологический университет «МИСИС»,
Москва, Российская Федерация*

Резюме. В статье рассматривается задача прогнозирования патологических состояний желудочно-кишечного тракта по эндоскопическим изображениям с использованием моделей машинного обучения. Актуальность исследования связана с необходимостью повышения воспроизводимости предварительной оценки эндоскопических данных и снижения риска пропуска клинически значимых изображений при большом объеме визуальной информации. Цель работы состоит в разработке и экспериментальной проверке ресурсно-эффективной схемы классификации эндоскопических изображений, учитывающей среднее качество распознавания и надежность автоматического решения для патологических классов. В качестве экспериментальной базы использован открытый набор Kvasir-v2, включающий 8000 изображений восьми классов. Сравнивались простая сверточная нейронная сеть, MobileNetV2 и EfficientNetB0 с переносом обучения. Дополнительно предложена селективная схема, объединяющая прогнозы двух наиболее сильных моделей и направляющая сомнительные изображения на ручную проверку. Эксперимент показал, что EfficientNetB0 достигает средней F-меры 0,8497 на тестовой части, а объединенный прогноз повышает значение показателя до 0,8681. Предложенная селективная схема обеспечила среднюю F-меру 0,9581 среди автоматически принятых решений и снизила число опасных автоматических ошибок с 64 до 10. Полученные результаты показывают перспективность клинически ориентированного подхода, при котором автоматическая классификация дополняется контролем уверенности и оценкой риска пропуска патологического состояния.

Ключевые слова: машинное обучение, эндоскопические изображения, желудочно-кишечный тракт, прогнозирование заболеваний, сверточные нейронные сети, перенос обучения, селективная классификация, патологические состояния, поддержка медицинских решений.

Для цитирования: Альмусави М.Р.К., Ляпунцова Е.В. Разработка моделей машинного обучения для прогнозирования заболеваний желудочно-кишечного тракта. *Моделирование, оптимизация и информационные технологии*. 2026;14(9). URL: <https://moitvivr.ru/ru/journal/article?id=2398>
DOI: 10.26102/2310-6018/2026.60.9.014

Development of machine learning models for predicting gastrointestinal diseases

M.R.K. Almusawi✉, E.V. Lyapunтова

National University of Science and Technology "MISIS", Moscow, the Russian Federation

Abstract. The article addresses the problem of predicting pathological conditions of the gastrointestinal tract from endoscopic images using machine learning models. The relevance of the study is determined by the need to improve the reproducibility of preliminary assessment of endoscopic data and to reduce the risk of missing clinically significant images when processing large volumes of visual information. The aim of the study is to develop and experimentally evaluate a resource-efficient classification scheme for endoscopic images, taking into account not only the average recognition quality, but also the reliability of automatic decisions for pathological classes. The open Kvasir-v2 dataset containing 8000 images from eight classes was used as the experimental basis. A simple convolutional neural network,

MobileNetV2, and EfficientNetB0 with transfer learning were compared. In addition, a selective scheme was proposed, combining the predictions of the two strongest models and sending uncertain images for manual review. The experiment showed that EfficientNetB0 achieved a macro-F-score of 0.8497 on the test set, while the combined prediction increased this value to 0.8681. The proposed selective scheme provided a macro-F-score of 0.9581 among automatically accepted decisions and reduced the number of dangerous automatic errors from 64 to 10. The obtained results demonstrate the potential of a clinically oriented approach in which automatic classification is supplemented by confidence control and assessment of the risk of missing a pathological condition.

Keywords: machine learning, endoscopic images, gastrointestinal tract, disease prediction, convolutional neural networks, transfer learning, selective classification, pathological conditions, medical decision support.

For citation: Almusawi M.R.K., Lyapunsova E.V. Development of machine learning models for predicting gastrointestinal diseases. *Modeling, Optimization and Information Technology*. 2026;14(9). (In Russ.). URL: <https://moitvvt.ru/ru/journal/article?id=2398> DOI: 10.26102/2310-6018/2026.60.9.014

Введение

Прогнозирование заболеваний желудочно-кишечного тракта по эндоскопическим изображениям относится к задачам, где формальная точность модели не исчерпывает качества результата. Практическая ценность автоматического анализа определяется способностью снижать риск пропуска патологических состояний, включая воспалительные изменения, полипы, опухолевые и предопухолевые признаки.

Высокая распространенность заболеваний органов пищеварения и онкологических поражений желудочно-кишечного тракта усиливает значимость раннего анализа эндоскопических данных. По данным GLOBOCAN 2022, колоректальный рак входит в число ведущих причин онкологической заболеваемости и смертности в мире: в 2022 г. зарегистрировано около 1,93 млн новых случаев и 904 тыс. смертей [1, 2] – показатели фиксируют клиническую цену своевременного обнаружения визуально значимых изменений.

Искусственный интеллект в эндоскопии перешел из экспериментальной области в зону клинически ориентированной оценки. В позиции Европейского общества желудочно-кишечной эндоскопии отмечено, что ценность интеллектуальных систем следует соотносить с измеримыми показателями качества эндоскопии, а не ограничивать техническими метриками модели [3]. Данное методическое требование особенно существенно для задач классификации изображений. Высокая доля верных ответов теряет прикладную убедительность, если неизвестно, как модель распознает патологические классы и сколько клинически опасных ошибок допускает.

В анализе эндоскопических изображений широко применяются сверточные нейронные сети. Обзор работ по глубокому обучению в эндоскопической визуализации показывает использование таких моделей для классификации, сегментации, повышения качества изображений и обнаружения патологических областей [4]. Медицинские модели требуют размеченных выборок, чувствительны к качеству изображения, с трудом переносятся на новые условия съемки и часто нуждаются в интерпретации результата [4].

Результаты, полученные на подготовленных наборах изображений, не всегда отражают устойчивость модели в клинической среде. Обзор [5] сообщает о высоких значениях чувствительности и специфичности в сценариях верхней эндоскопии, колоноскопии и капсульной эндоскопии. Одновременно авторы указывают на низкий общий уровень доказательности части публикаций, неоднородность моделей, различия в тестовых наборах и потребность во внешней проверке [5]. Экспериментальная схема должна учитывать этот разрыв между лабораторным качеством и надежностью автоматического решения.

В задачах классификации заболеваний желудочно-кишечного тракта по изображениям часто применяется перенос обучения. При ограниченной медицинской разметке этот прием уменьшает объем обучения с нуля и повышает качество распознавания. Авторы статьи [6] показали, что архитектуры семейства EfficientNet в сочетании с расширением обучающей выборки дают высокий результат при диагностике заболеваний желудочно-кишечного тракта по изображениям. Однако рост точности необходимо сопоставлять с вычислительной стоимостью: более сложная модель может требовать больше времени, памяти и ресурсов для получения прогноза.

Клиническое внедрение интеллектуальных систем в гастроэнтерологию требует контроля неопределенности, прозрачности поведения и устойчивости к пограничным случаям. Современные обзоры подчеркивают нерешенные вопросы регулирования, ответственности, смещения данных, безопасности пациента и доказательств эффективности в реальной практике [7], а также необходимость клинической валидации и сопоставления интеллектуальных систем с работой специалиста в желудочно-кишечной эндоскопии [8], поэтому модель прогнозирования заболеваний желудочно-кишечного тракта должна выдавать не только класс изображения. Значение имеет уверенность прогноза – при недостаточной уверенности корректнее передать изображение специалисту, чем принудительно присвоить класс.

Для отдельных нозологических направлений, включая поражения пищевода, глубокое обучение рассматривается как инструмент анализа эндоскопических изображений и поддержки диагностики [9]. Для язвенного колита метаанализ CNN-подходов показал возможность автоматизированной оценки эндоскопической тяжести заболевания, однако также подтвердил необходимость стандартизированной проверки качества алгоритмов [10]. В российских исследованиях представлены результаты применения интеллектуальных систем для обнаружения и дифференцировки новообразований толстой кишки в режиме реального времени [11], классификации эндоскопических изображений методами глубокого машинного обучения [12], а также автоматизированного контроля качества колоноскопического исследования [13]. Селективная классификация решает указанную проблему через управляемый отказ от автоматического ответа. В медицинской визуализации отказ от автоматического решения безопаснее автоматического пропуска патологии. Помимо средней F-меры по классам оценивается количество опасных автоматических ошибок, когда патологическое изображение отнесено к непатологическому классу без направления на ручную проверку.

Цель исследования состоит в разработке и экспериментальной оценке моделей машинного обучения для прогнозирования заболеваний желудочно-кишечного тракта по эндоскопическим изображениям с учетом качества распознавания, вычислительной стоимости и риска пропуска патологических состояний.

Для достижения цели поставлены следующие задачи:

1. Сравнить простую сверточную нейронную сеть и модели переноса обучения по качеству классификации и вычислительным затратам.
2. Отдельно оценить качество распознавания патологических классов.
3. Разработать объединенный прогноз на основе двух наиболее сильных моделей.
4. Предложить селективное правило принятия решения, направляющее сомнительные изображения на ручную проверку.
5. Оценить снижение числа опасных автоматических ошибок по сравнению с обычной классификацией.

Научная новизна работы состоит в клинически ориентированной схеме оценки и принятия решения. Стандартное сравнение моделей по общей точности дополняется анализом патологических классов, вычислительных ограничений, отказа от автоматического ответа для сомнительных изображений.

Материалы и методы

Экспериментальная часть выполнена на открытом наборе эндоскопических изображений Kvasir-v2. Выбор набора обусловлен предметной близостью к задаче прогнозирования заболеваний желудочно-кишечного тракта и удобной структурой для воспроизводимого эксперимента. В описании набора [14] зафиксировано происхождение базы изображений и ее назначение для компьютерной поддержки выявления заболеваний желудочно-кишечного тракта.

В эксперименте использована версия Kvasir-v2, содержащая 8000 изображений восьми классов (Таблица 1). К патологическим состояниям отнесены esophagitis, polyps и ulcerative-colitis, а остальные классы отражают нормальные анатомические ориентиры или процедурные состояния эндоскопического исследования. Разделение классов на клинические группы необходимо для последующей оценки качества, поскольку ошибка на патологическом изображении имеет более высокую практическую цену, чем путаница между непатологическими или процедурными классами. Примеры эндоскопических изображений разных классов приведены на Рисунке 1.

Таблица 1 – Распределение изображений набора Kvasir-v2 по классам

Table 1 – Distribution of Kvasir-v2 images by class

Класс изображения	Клиническая группа	Количество изображений
dyed-lifted-polyps	Непатологический или процедурный класс	1000
dyed-resection-margins	Непатологический или процедурный класс	1000
esophagitis	Патологическое состояние	1000
normal-cecum	Непатологический или процедурный класс	1000
normal-pylorus	Непатологический или процедурный класс	1000
normal-z-line	Непатологический или процедурный класс	1000
polyps	Патологическое состояние	1000
ulcerative-colitis	Патологическое состояние	1000

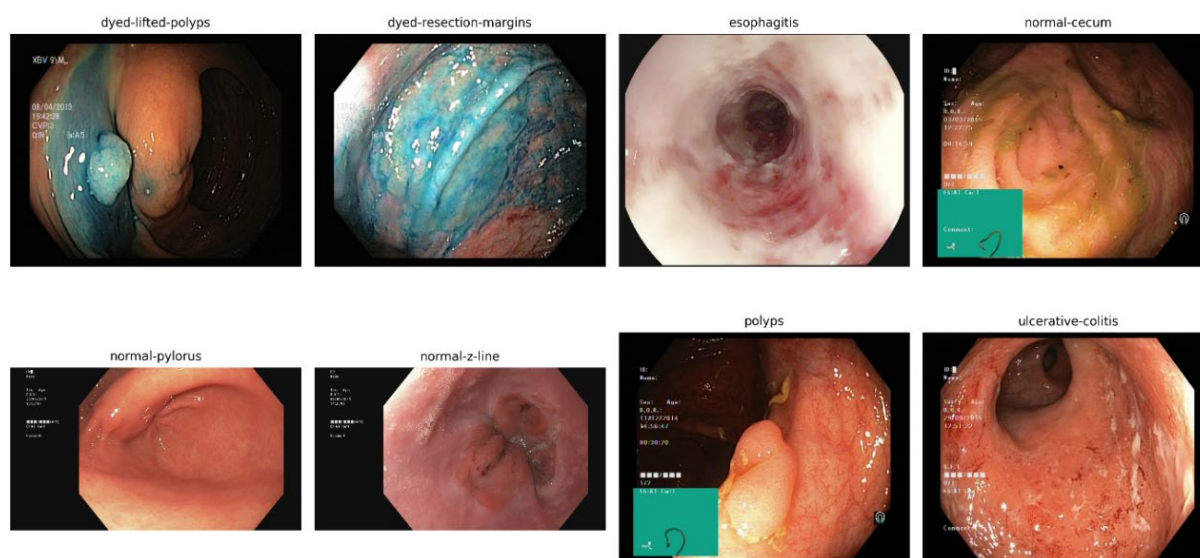


Рисунок 1 – Примеры эндоскопических изображений разных классов

Figure 1 – Examples of endoscopic images from different classes

Набор изображений разделен на обучающую, проверочную и тестовую части в соотношении 70:15:15. Разбиение выполнялось стратифицированно, с сохранением

одинакового представительства каждого класса во всех частях выборки. Обучающая часть применялась для настройки параметров моделей, проверочная часть – для выбора лучшей конфигурации и расчета весов объединенного прогноза, тестовая часть оставалась изолированной до итоговой оценки (Таблица 2).

Таблица 2 – Структура разбиения набора данных

Table 2 – Dataset split structure

Часть выборки	Количество изображений	Доля, %
Обучающая часть	5600	70,0
Проверочная часть	1200	15,0
Тестовая часть	1200	15,0

В обучающую часть вошло по 700 изображений каждого класса, в проверочную и тестовую части – по 150 изображений каждого класса. Перед подачей в модель изображения приводились к размеру 224×224 пикселя; значения пикселей обрабатывались в соответствии с требованиями конкретной архитектуры. Для обучающей части применялось расширение выборки: случайное горизонтальное отражение, небольшой поворот, изменение масштаба и контраста.

Простая сверточная нейронная сеть обучалась с нуля и задавала базовый ориентир. MobileNetV2 включена в сравнительный анализ как компактная архитектура, рассчитанная на эффективное распознавание изображений при ограниченном числе операций и параметров [15], а EfficientNetB0 представляет более сложное семейство моделей, где масштабирование сети строится через согласованное изменение глубины, ширины и разрешения [16]. В обоих вариантах переноса обучения использовалась основа, предварительно обученная на ImageNet, с заменой выходного классификационного слоя под восемь классов эндоскопических изображений (Таблица 3).

Таблица 3 – Конфигурация моделей машинного обучения

Table 3 – Configuration of machine learning models

Модель	Способ обучения	Всего параметров	Обучаемых параметров	Необучаемых параметров
Simple_CNN	Обучение с нуля	94 280	94 280	0
MobileNetV2	Перенос обучения	2 268 232	10 248	2 257 984
EfficientNetB0	Перенос обучения	4 059 819	10 248	4 049 571

В моделях переноса обучения сверточная основа была заморожена, настраивались новые классификационные слои. Простая сверточная сеть полностью настраивалась на обучающей части и проверяла достаточность признаков, извлекаемых без предварительного обучения.

Все модели обучались с одинаковыми общими параметрами. Размер пакета изображений составлял 32, максимальное число эпох ограничивалось восемью, оптимизация выполнялась алгоритмом Adam [17]. Функция потерь задавалась как категориальная перекрестная энтропия для многоклассовой классификации:

$$L = -\frac{1}{N} \sum_{n=1}^N \ln p_{n,y_n}, \quad (1)$$

где L – значение функции потерь; N – количество изображений в выборке; n – номер изображения; y_n – правильный класс изображения с номером n ; p_{n,y_n} – вероятность правильного класса, рассчитанная моделью для изображения с номером n ; $\ln$ – натуральный логарифм.

Основной сравнительной метрикой выбрана средняя F-мера по классам, которая рассчитывалась по формулам:

$$F_{\text{macro}} = \frac{1}{K} \sum_{i=1}^K F_{1,i}, \quad (2)$$

$$F_{1,i} = \frac{2P_iR_i}{P_i+R_i}, \quad (3)$$

$$P_i = \frac{TP_i}{TP_i+FP_i}, R_i = \frac{TP_i}{TP_i+FN_i}, \quad (4)$$

где F_{macro} – средняя F-мера по всем классам; K – количество классов; i – номер класса; $F_{1,i}$ – F-мера для класса i ; P_i – точность положительных решений для класса i ; R_i – полнота для класса i ; TP_i – количество изображений класса i , верно отнесенных к этому классу; FP_i – количество изображений других классов, ошибочно отнесенных к классу i ; FN_i – количество изображений класса i , ошибочно отнесенных к другим классам.

После обучения одиночных моделей был сформирован объединенный прогноз, куда вошли MobileNetV2 и EfficientNetB0, поскольку обе модели заметно превосходили простую сверточную сеть на проверочной части. Вес каждой модели рассчитывался по средней F-мере на проверочной части выборки, что не требовало дополнительного обучения метамоделей и сохраняло вычислительную простоту эксперимента.

Взвешенное объединение прогнозов моделей:

$$p_i^{\text{ens}} = \sum_{m=1}^M w_m p_i^{(m)}, \quad (5)$$

$$w_m = \frac{F_m}{\sum_{s=1}^M F_s}, \quad (6)$$

где p_i^{ens} – итоговая вероятность класса i в объединенном прогнозе; M – количество объединяемых моделей; m – номер модели; w_m – вес модели m ; $p_i^{(m)}$ – вероятность класса i , рассчитанная моделью m ; F_m – средняя F-мера модели m на проверочной части выборки; s – индекс суммирования по моделям.

В проведенном эксперименте вес модели MobileNetV2 составил 0,492, вес модели EfficientNetB0 – 0,508 (Таблица 4).

Таблица 4 – Веса моделей в объединенном прогнозе

Table 4 – Model weights in the combined prediction

Модель	Доля верных ответов на проверочной части	Средняя F-мера на проверочной части	Вес модели
MobileNetV2	0,8633	0,8620	0,492
EfficientNetB0	0,8900	0,8899	0,508

После объединения прогнозов применялось селективное правило принятия решения:

$$r(x) = \begin{cases} 1, & p_{(1)} \geq \theta_c \wedge (p_{(1)} - p_{(2)}) \geq \theta_m \wedge (\hat{y} \in C_{\text{path}} \vee \max_{i \in C_{\text{path}}} p_i < \theta_p), \\ 0, & \text{в остальных случаях} \end{cases} \quad (7)$$

где $r(x)$ – решение селективной схемы для изображения x ; при $r(x) = 1$ прогноз принимается автоматически, при $r(x) = 0$ изображение направляется на ручную проверку; $p_{(1)}$ – наибольшая вероятность среди всех классов; $p_{(2)}$ – вторая по величине вероятность среди всех классов; θ_c – порог уверенности прогноза; θ_m – порог различия между двумя наиболее вероятными классами; $\hat{y}$ – класс, предсказанный моделью; C_{path} – множество патологических классов; p_i – вероятность класса i ; θ_p – порог скрытого патологического сигнала.

В эксперименте использованы значения $\theta_c = 0,85, \theta_m = 0,00, \theta_p = 0,20$. Порог уверенности ограничивает автоматическую выдачу слабых прогнозов, а порог скрытого патологического сигнала направляет изображение на ручную проверку, если модель формально выбрала непатологический класс, но вероятность патологического класса остается достаточно высокой.

Качество селективной схемы оценивалось отдельно от качества обычной классификации. Основным дополнительным показателем – количество опасных автоматических ошибок. Под такой ошибкой понимается случай, когда изображение действительно относится к патологическому классу, но автоматическая система принимает непатологический прогноз без передачи изображения специалисту:

$$D = \sum_{n=1}^N I(r_n = 1 \wedge y_n \in C_{\text{path}} \wedge \hat{y}_n \notin C_{\text{path}}), \quad (8)$$

$$q_D = \frac{D}{N_{\text{path}}}, \quad (9)$$

где D – количество опасных автоматических ошибок; n – номер изображения; N – количество изображений в тестовой части выборки; $I(\cdot)$ – индикаторная функция, равная 1 при выполнении условия и 0 в остальных случаях; r_n – решение селективной схемы для изображения с номером n ; y_n – правильный класс изображения с номером n ; $\hat{y}_n$ – класс, предсказанный моделью для изображения с номером n ; C_{path} – множество патологических классов; q_D – доля опасных автоматических ошибок среди патологических изображений; N_{path} – количество патологических изображений в тестовой части выборки.

Результаты

Итоговая оценка одиночных моделей выполнялась на тестовой части, изолированной от обучения и подбора параметров. Результаты приведены в Таблице 5. Простая сверточная сеть достигла средней F-меры 0,6754. MobileNetV2 повысила показатель до 0,8393. Лучший результат среди одиночных моделей получила EfficientNetB0: средняя F-мера составила 0,8497, доля верных ответов – 0,8517.

Таблица 5 – Итоговая оценка одиночных моделей на тестовой части
Table 5 – Final evaluation of individual models on the test set

Модель	Доля верных ответов	Средняя точность положительных решений	Средняя полнота	Средняя F-мера	Время прогноза, с	Время на изображение, мс
Simple CNN	0,6792	0,6809	0,6792	0,6754	10,43	8,6933
MobileNetV2	0,8417	0,8455	0,8417	0,8393	18,42	15,3465
EfficientNetB0	0,8517	0,8567	0,8517	0,8497	23,67	19,7270

Патологические классы оценивались отдельно от общей многоклассовой классификации. Среднее качество по всем восьми классам может скрывать ошибки,

значимые для медицинского сценария. В Таблице 6 приведены средние значения точности положительных решений, полноты и F-меры для классов esophagitis, polyps и ulcerative-colitis. EfficientNetB0 получила среднюю F-меру 0,8315 на патологических классах, MobileNetV2 – 0,8230, Simple_CNN – 0,5934.

Таблица 6 – Средние показатели качества моделей на патологических классах

Table 6 – Average model quality metrics for pathological classes

Модель	Средняя точность положительных решений	Средняя полнота	Средняя F-мера
EfficientNetB0	0,8585	0,8111	0,8315
MobileNetV2	0,8714	0,7889	0,8230
Simple CNN	0,6392	0,5622	0,5934

Матрица ошибок лучшей одиночной модели приведена на Рисунке 2. Наиболее заметная проблема относится к классу esophagitis: часть изображений этого класса отнесена к normal-z-line. Ошибка имеет клиническую значимость, поскольку патологическое состояние получает непатологическую метку. Путаница также возникает между dyed-lifted-polyps и dyed-resection-margins; оба класса связаны с процедурными изменениями после эндоскопического вмешательства.

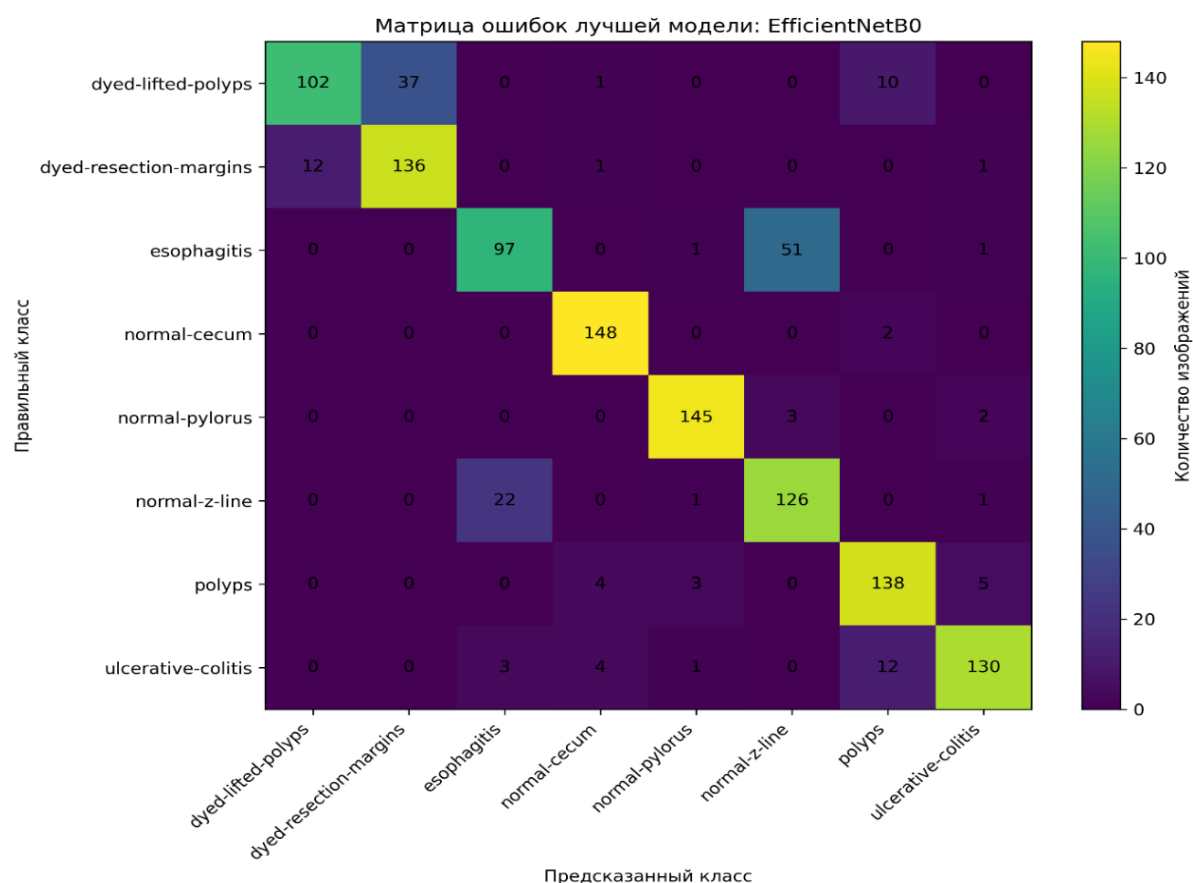


Рисунок 2 – Матрица ошибок лучшей одиночной модели

Figure 2 – Confusion matrix of the best individual model

После оценки одиночных моделей рассчитан объединенный прогноз MobileNetV2 и EfficientNetB0. Веса моделей определялись по средней F-мере на проверочной части:

0,492 для MobileNetV2 и 0,508 для EfficientNetB0. На тестовой части объединенный прогноз достиг доли верных ответов 0,8700 и средней F-меры 0,8681. Значение выше результата каждой одиночной модели, однако объединение вероятностей сохранило часть опасных автоматических ошибок.

Селективная схема применялась к объединенному прогнозу. На тестовой части 715 из 1200 изображений получили автоматический прогноз, еще 485 изображений направлены на ручную проверку. Среди автоматически принятых решений доля верных ответов составила 0,9692, средняя F-мера – 0,9581. Показатели Таблицы 7 рассчитаны для уверенной зоны работы схемы (только среди автоматически принятых решений; не по всей тестовой части).

Наиболее выраженный эффект предложенной схемы проявился при подсчете опасных автоматических ошибок. Для MobileNetV2 зафиксировано 83 случая, для EfficientNetB0 – 64 случая. Объединенный прогноз без ручной проверки сохранил 64 опасные ошибки. Применение селективного правила снизило число опасных автоматических ошибок до 10. Сравнение приведено в Таблице 8.

Причины направления изображений на ручную проверку представлены в Таблице 9. Большая часть отложенных случаев связана с недостаточной уверенностью прогноза. Еще 133 изображения направлены на проверку из-за заметной вероятности патологического состояния при непатологическом прогнозе.

Таблица 7 – Итоговая оценка предложенной селективной схемы на тестовой части

Table 7 – Final evaluation of the proposed selective scheme on the test set

Показатель	Значение
Порог уверенности	0,85
Порог различия двух лучших вероятностей	0,00
Порог скрытого патологического сигнала	0,20
Автоматически принято изображений	715
Направлено на ручную проверку	485
Доля автоматических решений	0,5958
Доля ручной проверки	0,4042
Доля верных автоматических решений	0,9692
Средняя F-мера автоматических решений	0,9581
Опасных автоматических ошибок	10
Доля опасных автоматических ошибок среди патологических изображений	0,0222

Таблица 8 – Сравнение числа опасных автоматических ошибок

Table 8 – Comparison of dangerous automatic errors

Подход	Опасных автоматических ошибок
MobileNetV2	83
EfficientNetB0	64
Объединенный прогноз без ручной проверки	64
Предложенная селективная схема	10

Таблица 9 – Причины направления изображений на ручную проверку
Table 9 – Reasons for sending images to manual review

Причина ручной проверки	Количество изображений	Доля среди направленных на проверку, %
Недостаточная уверенность прогноза	352	72,58
Заметная вероятность патологического состояния при непатологическом прогнозе	133	27,42

Вторая причина в Таблице 9 отвечает за выявление возможного скрытого патологического сигнала, даже если максимальная вероятность относится к непатологическому классу.

Обсуждение

Результаты эксперимента подтверждают преимущество переноса обучения для классификации эндоскопических изображений. Разрыв между Simple_CNN и моделями MobileNetV2, EfficientNetB0 связан с характером признаков, доступных на старте обучения. Простая сверточная сеть формировала признаки только по 5600 обучающим изображениям. Модели переноса обучения начинали с уже настроенных сверточных основ, обученных на большом массиве визуальных данных.

EfficientNetB0 дала лучший результат среди одиночных моделей. Рост качества сопровождался увеличением времени прогноза на одно изображение, поэтому EfficientNetB0 подходит для сценариев, где приоритет отдается точности, а MobileNetV2 сохраняет практическую ценность при более жестких ограничениях по времени обработки.

Матрица ошибок лучшей одиночной модели показывает сохранение клинически важных смещений классов. Наиболее заметная ошибка относится к классу esophagitis: часть изображений этого класса отнесена к normal-z-line. В медицинском сценарии ошибка получает высокий вес, поскольку патологическое изображение получает непатологическую метку.

Селективная схема изменила логику оценки результата. После введения правила уверенности и проверки скрытого патологического сигнала число опасных автоматических ошибок снизилось до 10. Снижение риска достигнуто за счет передачи 485 изображений специалисту. Результат следует трактовать как управляемую передачу неопределенности на ручную проверку.

Для интерпретации решений использованы тепловые карты, сформированные по подходу Grad-CAM [18]. На Рисунке 3 представлены тепловые карты для правильных патологических прогнозов. Визуализация не является доказательством диагноза, но помогает проверить, какие области изображения участвовали в решении модели. По этим картам можно отделить уверенные случаи, где модель ориентируется на содержательную часть эндоскопического кадра, от ситуаций, где итоговый класс мог зависеть от побочных признаков съемки.

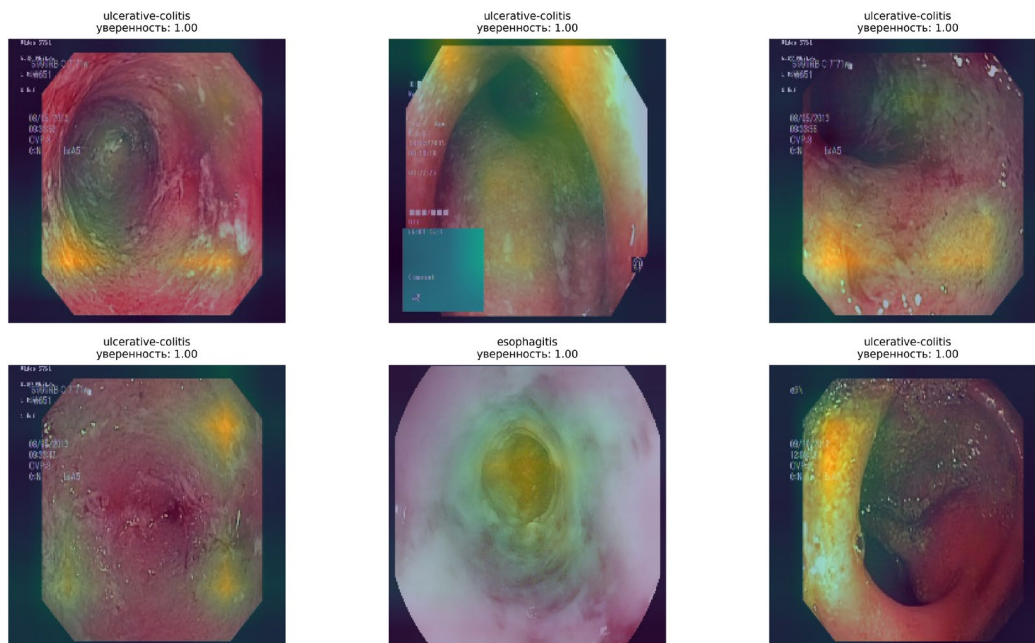


Рисунок 3 – Тепловые карты областей, повлиявших на правильные прогнозы патологических состояний

Figure 3 – Heat maps of regions influencing correct predictions of pathological conditions

Тепловые карты для изображений, направленных на ручную проверку, на Рисунке 4 показывают иную зону работы селективной схемы. В этой группе модель обнаруживает признаки, достаточные для подозрения или частичного совпадения с патологическим классом, но не достигает устойчивого автоматического решения. Принудительная классификация таких случаев нежелательна, и передача специалисту сохраняет возможность экспертной оценки там, где модель фиксирует неопределенность.

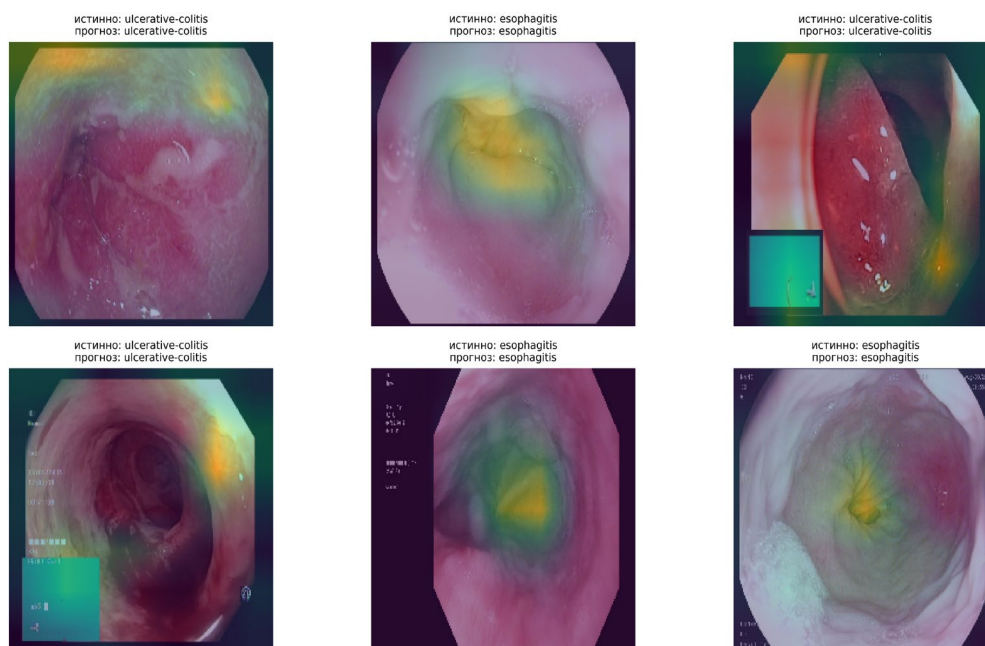


Рисунок 4 – Тепловые карты для патологических изображений, направленных на ручную проверку

Figure 4 – Heat maps for pathological images sent to manual review

Ограничения эксперимента определяются одним открытым набором изображений и фиксированным разбиением на обучающую, проверочную и тестовую части. Результаты характеризуют поведение моделей на Kvasir-v2 и требуют внешней проверки на данных из другой клинической среды. Классы набора включают нормальные, патологические и процедурные изображения, поэтому многоклассовая постановка отличается от полноценной клинической диагностики пациента. Дальнейшая работа должна включать внешнюю валидацию, расширение набора патологий и сопоставление селективной схемы с сегментацией подозрительных областей.

Заключение

Разработана и экспериментально проверена схема прогнозирования заболеваний желудочно-кишечного тракта по эндоскопическим изображениям. Эксперимент выполнен на наборе Kvasir-v2, включающем 8000 изображений восьми классов. В составе набора выделены патологические состояния esophagitis, polyps и ulcerative-colitis. Равномерное распределение классов и стратифицированное разбиение выборки обеспечили корректное сравнение моделей без дополнительной коррекции дисбаланса.

Сравнение одиночных моделей подтвердило преимущество переноса обучения перед обучением сверточной сети с нуля. Simple_CNN достигла средней F-меры 0,6754 на тестовой части, MobileNetV2 – 0,8393, EfficientNetB0 – 0,8497. На патологических классах EfficientNetB0 также показала лучший результат среди одиночных моделей: средняя F-мера составила 0,8315. Рост качества сопровождался увеличением времени прогноза на одно изображение, поэтому вычислительная стоимость остается значимым критерием выбора архитектуры. Объединение прогнозов MobileNetV2 и EfficientNetB0 повысило среднюю F-меру на тестовой части до 0,8681. Матрица ошибок при этом сохранила клинически значимые смещения, прежде всего между esophagitis и normal-z-line. Для снижения риска автоматического пропуска патологических изображений итоговая схема дополнена селективным правилом принятия решения. Правило принимает автоматический прогноз для уверенных случаев и направляет сомнительные изображения на ручную проверку.

Селективная схема автоматически обработала 715 из 1200 тестовых изображений. Среди принятых автоматических решений доля верных ответов составила 0,9692, средняя F-мера – 0,9581. Число опасных автоматических ошибок снизилось с 64 у лучшей одиночной модели и объединенного прогноза без ручной проверки до 10. Клинически ориентированная оценка зафиксировала качество модели в зоне уверенного прогноза и риск автоматического пропуска патологического состояния.

Ограничения исследования определяются одним открытым набором изображений и фиксированным разбиением на обучающую, проверочную и тестовую части. Полученные результаты требуют проверки на внешних данных из другой клинической среды. Дальнейшую работу целесообразно связать с расширением набора патологий, тестированием на независимых эндоскопических выборках и добавлением сегментации подозрительных областей для повышения интерпретируемости автоматического прогноза.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Bray F., Laversanne M., Sung H., et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*. 2024;74(3):229–263. <https://doi.org/10.3322/caac.21834>

2. Wu S., Zhang Y., Lin Z., et al. Global burden of colorectal cancer in 2022 and projections to 2050: incidence and mortality estimates from GLOBOCAN. *BMC Cancer*. 2025;25(1):1770. <https://doi.org/10.1186/s12885-025-15138-0>
3. Messmann H., Bisschops R., Antonelli G., et al. Expected value of artificial intelligence in gastrointestinal endoscopy: European Society of Gastrointestinal Endoscopy (ESGE) position statement. *Endoscopy*. 2022;54(12):1211–1231. <https://doi.org/10.1055/a-1950-5694>
4. Nie Z., Xu M., Wang Zh., et al. A review of application of deep learning in endoscopic image processing. *Journal of Imaging*. 2024;10(11):275. <https://doi.org/10.3390/jimaging10110275>
5. Zha B., Cai A., Wang G. Diagnostic accuracy of artificial intelligence in endoscopy: umbrella review. *JMIR Medical Informatics*. 2024;12:e56361. <https://doi.org/10.2196/56361>
6. Zubair Rahman A.M.J.M., Mythili R., Chokkanathan K., et al. Enhancing image-based diagnosis of gastrointestinal tract diseases through deep learning with EfficientNet and advanced data augmentation techniques. *BMC Medical Imaging*. 2024;24(1):306. <https://doi.org/10.1186/s12880-024-01479-y>
7. El-Sayed A., Lovat L.B., Ahmad O.F. Clinical implementation of artificial intelligence in gastroenterology: current landscape, regulatory challenges, and ethical issues. *Gastroenterology*. 2025;169(3):518–530. <https://doi.org/10.1053/j.gastro.2025.01.254>
8. Ali H., Muzammil M.A., Dahiya D.S., et al. Artificial intelligence in gastrointestinal endoscopy: a comprehensive review. *Annals of Gastroenterology*. 2024;37(2):133–141. <https://doi.org/10.20524/aog.2024.0861>
9. Theocharopoulos Ch., Davakis S., Ziogas D.C., et al. Deep learning for image analysis in the diagnosis and management of esophageal cancer. *Cancers*. 2024;16(19):3285. <https://doi.org/10.3390/cancers16193285>
10. Jahagirdar V., Bapaye J.A., Chandan S., et al. Diagnostic accuracy of convolutional neural network-based machine learning algorithms in endoscopic severity prediction of ulcerative colitis: a systematic review and meta-analysis. *Gastrointestinal Endoscopy*. 2023;98(2):145–154.e8. <https://doi.org/10.1016/j.gie.2023.04.2074>
11. Завьялов Д.В., Кашин С.В., Гусейнова С.Р. Система CAD EYE для детекции и дифференцировки новообразований толстой кишки в режиме реального времени. *Доказательная гастроэнтерология*. 2024;13(2):50–54. <https://doi.org/10.17116/dokgastro20241302150>
Zavyalov D.V., Kashin S.V., Guseinova S.R. CAD EYE for real-time detection and differentiation of colorectal lesions. *Russian Journal of Evidence-Based Gastroenterology*. 2024;13(2):50–54. (In Russ.). <https://doi.org/10.17116/dokgastro20241302150>
12. Хрящев В.В., Завьялов Д.В., Андержанова А.С. Классификация эндоскопических изображений устья червеобразного отростка на основе методов глубокого машинного обучения. *Цифровая обработка сигналов*. 2023;(1):35–38.
Khryashchev V.V., Zavyalov D.V., Anderzanova A.S. Classification of endoscopic images of the orifice of the appendix on the basis of deep machine learning. *Digital Signal Processing*. 2023;(1):35–38. (In Russ.).
13. Кашин С.В., Завьялов Д.В., Гусейнова С.Р. и др. Оценка ключевых показателей качества скрининговой колоноскопии с использованием системы искусственного интеллекта. *Доказательная гастроэнтерология*. 2025;14(4):56–64. <https://doi.org/10.17116/dokgastro20251404156>
Kashin S.V., Zavyalov D.V., Guseinova S.R., et al. Assessing quality indicators of screening colonoscopy using an artificial intelligence system. *Russian Journal of*

- Evidence-Based Gastroenterology*. 2025;14(4):56–64. (In Russ.). <https://doi.org/10.17116/dokgastro20251404156>
14. Pogorelov K., Randel K.R., Griwodz C., et al. KVASIR: a multi-class image dataset for computer aided gastrointestinal disease detection. In: *Proceedings of the 8th ACM on Multimedia Systems Conference, 20–23 June 2017, Taipei, Taiwan*. New York: ACM; 2017. P. 164–169. <https://doi.org/10.1145/3083187.3083212>
 15. Sandler M., Howard A., Zhu M., et al. MobileNetV2: inverted residuals and linear bottlenecks. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 18–23 June 2018, Salt Lake City, UT, USA*. IEEE; 2018. P. 4510–4520. <https://doi.org/10.1109/CVPR.2018.00474>
 16. Tan M., Le Q.V. EfficientNet: rethinking model scaling for convolutional neural networks. In: *Proceedings of the 36th International Conference on Machine Learning, 09–15 June 2019, Long Beach, CA, USA*. PMLR; 2019. P. 6105–6114.
 17. Kingma D.P., Ba J. Adam: a method for stochastic optimization. In: *3rd International Conference on Learning Representations, 07–09 May 2015, San Diego, CA, USA*. 2015. URL: <https://arxiv.org/abs/1412.6980>
 18. Selvaraju R.R., Cogswell M., Das A., et al. Grad-CAM: visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*. 2020;128(8):336–359. <https://doi.org/10.1007/s11263-019-01228-7>

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Альмусави Мустафа Рияд Кетаб, Mustafa Riyadh Ketab Almusawi,
аспирант, Национальный исследовательский технологический университет «МИСИС», Москва, Российская Федерация. Postgraduate, National University of Science and Technology "MISIS", Moscow, the Russian Federation.
e-mail: adammadam265@gmail.com

Ляпунцова Елена Вячеславовна, Elena V. Lyapuntsova, доктор технических наук, профессор, Национальный исследовательский технологический университет «МИСИС», Москва, Российская Федерация. Doctor of Engineering Sciences, Professor, National University of Science and Technology "MISIS", Moscow, the Russian Federation.
e-mail: liapuntsova.ev@misis.ru

Статья поступила в редакцию 28.05.2026; одобрена после рецензирования 31.08.2026; принята к публикации 17.09.2026.

The article was submitted 28.05.2026; approved after reviewing 31.08.2026; accepted for publication 17.09.2026.