DOI: 10.33667/2078-5631-2025-22-65-69

Влияние врачебного опыта на развитие машинного обучения при автоматизации общего исследования кала с использованием обработки изображений

Н. Ю. Черныш[1], В. С. Берестовская[1], А. Н. Тишко[1,2], А. А. Руднева[3], Т. В. Вавилова[1]

[1] ФГБУ «Национальный медицинский исследовательский центр имени В. А. Алмазова» Минздрава России, Санкт-Петербург, Россия
[2] ФГБУ «Всероссийский центр экстренной и радиационной медицины им А. М. Никифорова МЧС России», Санкт-Петербург, Россия
[3] ФГАОУ ВО «Балтийский федеральный университет имени Иммануила Канта», Калининград, Россия

РЕЗЮМЕ
Создание интеллектуальных программ анализа изображений кала на основе машинного обучения связано с качеством разметки изображений экспертами. В пилотном исследовании выявлена крайне низкая степень меж- и внутриэкспертной согласованности распознавания морфологических объектов по F1-score. Определены три типа расхождений: в количестве, наличии и категории объектов. Предложено использование консенсусного подхода, который позволит нивелировать субъективность и повысить надежность алгоритмов для автоматизации копрограммы.

КЛЮЧЕВЫЕ СЛОВА: *общий анализ кала, машинное обучение; разметка данных.*

КОНФЛИКТ ИНТЕРЕСОВ. *Авторы заявляют об отсутствии конфликта интересов.*

The impact of physician experience on the development of machine learning in automation of general fecal examination using image processing

N. Yu. Chernysh[1], V. S. Berestovskaya[1], A. N. Tishko[1,2], A. A. Rudneva[3], T. V. Vavilova[1]

[1] Almazov National Medical Research Centre, St. Petersburg, Russia
[2] All-Russian Center for Emergency and Radiation Medicine named after A.M. Nikiforov, Ministry of Emergency Situations of Russia, St. Petersburg, Russia
[3] Institute of Medicine and Life Sciences, Immanuel Kant Baltic Federal University, Kaliningrad, Russia

SUMMARY
The creation of intelligent fecal image analysis software based on machine-learning is dependent on the quality of image annotation by experts. A pilot study revealed an extremely low degree of inter- and intra-expert agreement in recognizing morphological objects, as measured by the F1-score. Three types of discrepancies were identified: in quantity, presence, and category of objects. It is proposed to use a consensus-based approach to mitigate subjectivity and enhance the reliability of algorithms for automation of general fecal examination.

KEY WORDS: *general fecal analysis, machine learning; data annotation.*

CONFLICT OF INTEREST. *The authors declare no conflict of interest.*

Введение

Общий анализ кала (копрограмма) представляет собой доступный способ для диагностики и мониторинга многих заболеваний желудочно-кишечного тракта (ЖКТ). Неинвазивный и менее затратный, чем колоноскопия, копрограмма является отправной точкой диагностического поиска при острых и хронических заболеваниях ЖКТ различной природы, помогает выявить возбудителей инфекционной и паразитарной природы, оценить состояние микробиома кишечника. Несмотря на то, что многие лабораторные процессы автоматизированы, исследование кала до недавнего времени оставалось исключением. Существующий метод для общего анализа кала требует технологических усовершенствований на преаналитическом (сбор/обработка) и аналитическом (исследование) этапах.

Последние годы автоматизированные системы с фотофиксацией препаратов активно выходят на рынок лабораторного оборудования [1]. Цифровая микроскопия является практичной альтернативой для визуальной оценки препарата под микроскопом и может обеспечить повышение эффективности работы химико-микроскопического подразделения медицинских лабораторий. Традиционный препарат для проведения копрограммы невозможно сохранить, пересмотреть позже, интегрировать изображение в электронную медицинскую карту и образовательные модули. Внедрение автоматизированной цифровой микроскопии создает условия для совершенствования традиционного способа исследования кала за счет унификации процедуры подготовки препарата, качеству и количеству полей зрения, стандартизации требований к качеству изображений.

Ключевым фактором для стандартизации результатов цифровой микроскопии является программное обеспечение, разработанное с использованием технологии машинного обучения (МО), как важного инструмента программ искусственного интеллекта [2].

Цель

В пилотном исследовании оценить согласованность результатов и выявить ключевые параметры разногласий при оценке цифровых фотографий препаратов кала врачами клинической лабораторной диагностики.

Материалы и методы

Материалы. Для проведения исследования была сформирована база изображений из 70 снимков, выполненных на автоматической многофункциональной станции для анализа кала SCIENDOX 2000R. В базу включены изображения, с которыми исследователи знакомились в процессе обучения работе на SCIENDOX 2000R, и снимки, которые исследователи видели впервые. Все снимки содержали объекты, которые требуют идентификации и классификации в рамках микроскопии кала при проведении копрограммы. Снимки были представлены в оригинальном формате без добавления шумов и поворота изображений. В качестве инструмента разметки поля зрения и программы для оценки результатов использовались собственные инструменты и программы на основе решений с открытым исходным кодом компании ООО «АврораАИ», являющейся резидентом Фонда Сколково, Российская Федерация; расчеты проводились с использованием языка программирования Python.

Методы. Всего в пилотном исследовании приняли участие 10 врачей клинической лабораторной диагностики. Для промежуточного аудита, результаты которого представлены в данной работе, сравнивали результаты 4 врачей клинической лабораторной диагностики с опытом проведения микроскопических исследований кала более 10 лет у каждого (далее – эксперты). Экспертам было предложено независимо друг от друга провести разметку, т.е. классифицировать параметры, идентифицируемые в рамках общего анализа кала, на снимках, включенных в базу изображений.

Таблица 1

Результаты среднего значения F1 по результатам распознавания объектов из базы изображений, выполненных разными экспертами

Объект	F1
Непереваримая клетчатка	0,21
Переваримая клетчатка	0,20
Яйцо аскариды человека неоплодотворённое	0,16
Переваренные мышечные волокна	0,16
Непереваренные мышечные волокна	0,14
Кристаллы холестерина	0,09
Капли жира	0,06
Слабопереваренные мышечные волокна	0,02
Глыбки жира	0,01
Оксалаты кальция	0,00
Мыла	0,00
Кристаллы билирубина	0,00
Слизь	0,00
Жирные кислоты	0,00
Зерна крахмала	0,00
Клетки цилиндрического эпителия	0,00
Неидентифицированные гельминты	0,00
Трипельфосфаты	0,00
Дрожжевые клетки	0,00

Обработка данных. Одна из метрик, используемых в машинном обучении: F1 – score (далее – F1), позволяет одновременно оценивать возможность разрабатываемой модели обеспечить баланс между обнаружением правильно размеченных объектов и предотвращением ошибочной идентификации [3]. Для количественной оценки согласованности результатов разметки рассчитывали F1 (*рис. 1*), исходя из того, что в данной работе в качестве назначенного эталона (правильно размеченные объекты) при попарном сравнении использовались результаты одного из экспертов. Вначале проводили расчёт F1 по результатам идентификации каждого морфологического объекта, полученного при разметке изображения, при сравнении данных пары экспертов. Далее результаты, полученные в парах, усредняли.

$$F1 = \frac{2\,TP}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}},$$ где

TP (True Positive) – количество правильно размеченных объектов;
FP (False Positive) – количество ложных правильно размеченных объектов;
FN (False Negative) – количество неразмеченных или неправильно размеченных объектов.

Рисунок 1. Формула для расчёта показателя F1

Результаты

В *таблице 1* представлены средние значения F1 по результатам разметки изображений в базе данных четырьмя экспертами, всего 6 сравнений.

Значения F1 могут находиться в диапазоне от 0 до 1, где: 0 – наихудший возможный результат, т.е. отсутствие согласованности в результатах идентификации, 1 – наилучший возможный результат, т.е. полное совпадение результатов идентификации, при целевом значении для согласованности результатов – 0,90–0,95.

Данные, приведенные в таблице 1 демонстрируют крайне низкую степень согласования результатов разметки, проведенной четырьмя независимыми специалистами. В данном случае значение F1 равное 0,00 означает, что при разметке снимков среди всех экспертов по данной морфологической категории ни одного согласования не зафиксировано. Наиболее высокий уровень согласования наблюдался для непереваримой клетчатки и переваримой клетчатки, F1 = 0,21 и 0,20, соответственно. Самыми проблемными для распознавания, т.е. полное отсутствие согласованности при оценке морфологических параметров в кале, оказались 10 объектов.

Оценка снимков из базы изображений позволила выявить три основных типа несогласованности при разметке объектов на изображении, определяющих высокий уровень разброса результатов:

Тип 1. При разметке снимка эксперт не обозначил то же количество объектов определенной морфологической категории, который отметил другой эксперт при попарном сравнении данных;

Тип 2. При разметке снимка эксперт не обозначил объект определенной морфологической категории, который отметил другой эксперт при попарном сравнении данных;

Тип 3. При разметке снимка эксперт обозначил объект, как относящийся к определенной морфологической категории, в то время как другой эксперт при попарном сравнении данных отнес этот объект к другой морфологической категории.

На *рисунке 2* приведена разметка одного изображения четырьмя экспертами с выявленными несоответствиями. В частности, для эксперта 3 выявлены следующие типы несоответствий:

Тип 1 (расхождение в числе выявленных объектов). Эксперт 3 идентифицировал 7 объектов «Переваренные мышечные волокна», при этом эксперт 1 не обнаружил ни одного объекта, эксперт 2 – разметил 3, а эксперт 4 выделил более 30 объектов соответствующей категории.

Тип 2 (расхождение в числе категорий выявленных объектов). Эксперт 3 идентифицировал на снимке объекты «Оксалаты кальция», «Переваримая клетчатка», «Мыла», которые отсутствуют при разметке изображения другими экспертами.

Тип 3 (расхождение в идентификации категории одного и того же объекта при разметке разными экспертами). Эксперт 3 и эксперт 4 разметили крупные объекты как «Непереваренные мышечные волокна», а эксперт 4 идентифицировал эти же объекты только как «Переваренные мышечные волокна».

Дизайн исследования предусматривал возможность оценки изображений, используемых при обучении работе на системе SCIENDOX 2000R без информирования об этом обстоятельстве экспертов, так как эти снимки не имели отдельной маркировки. Таким образом, появилась возможность сравнить разметку изображения одним и тем же экспертом в разные периоды времени, в ситуации, когда эксперт не знал, что проводит повторную разметку изображения. Условия эксперимента выявили те же три типа несоответствия, что и при разметке оригинальных изображений (*рис. 3*).

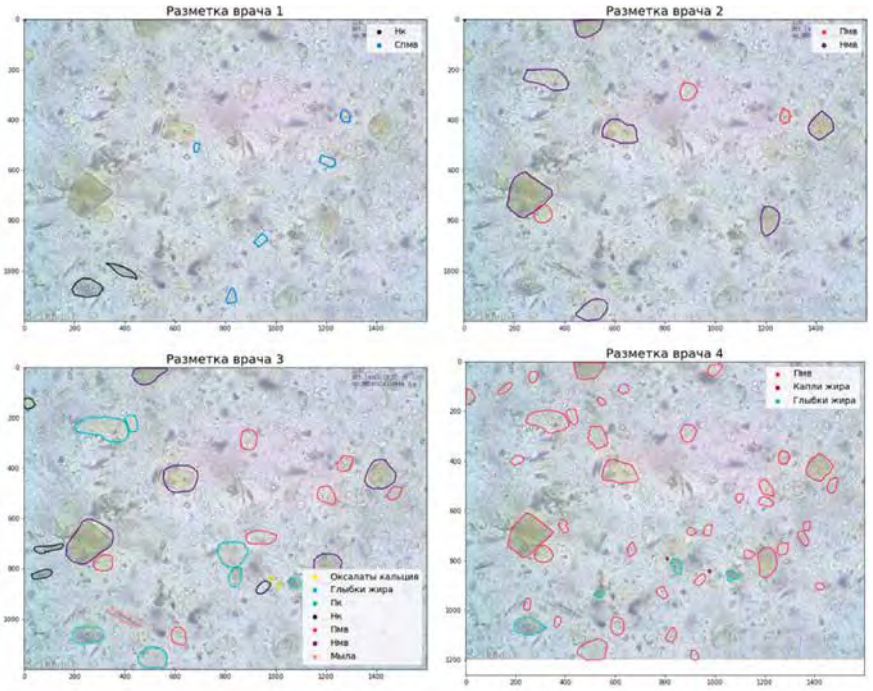

Рисунок 2. Результаты разметки одного изображения 4 экспертами

Таблица 2

Результаты среднего значения F1 по результатам распознавания объектов при попарном сравнении результатов одного эксперта

Эксперт 1		Эксперт 2	
Объект	F1	Объект	F1
Непереваримая клетчатка	0.34	Непереваримая клетчатка	0.11
Капли жира	0.00	Капли жира	0.86
Переваренные мышечные волокна	0.33	Переваренные мышечные волокна	0.21
Перевариваемая клетчатка	0.00	Перевариваемая клетчатка	0.75
Непереваренные мышечные волокна	0.00	Непереваренные мышечные волокна	0.31
Жирные кислоты	0.00	Зерна крахмала	0.15
Слабопереваренные мышечные волокна	0.00	Слабопереваренные мышечные волокна	0.00

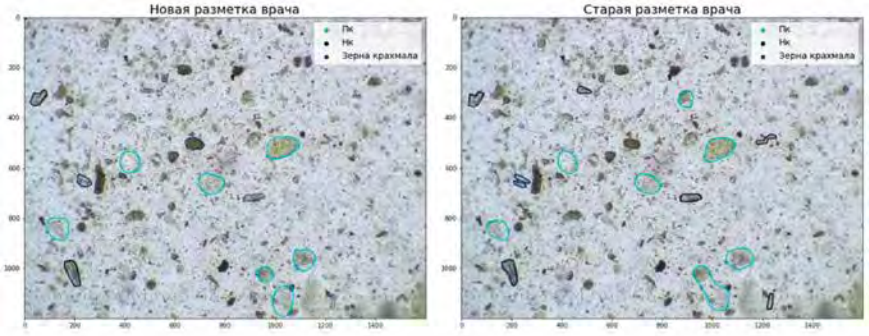

Рисунок 3А. Результат разметки одного эксперта

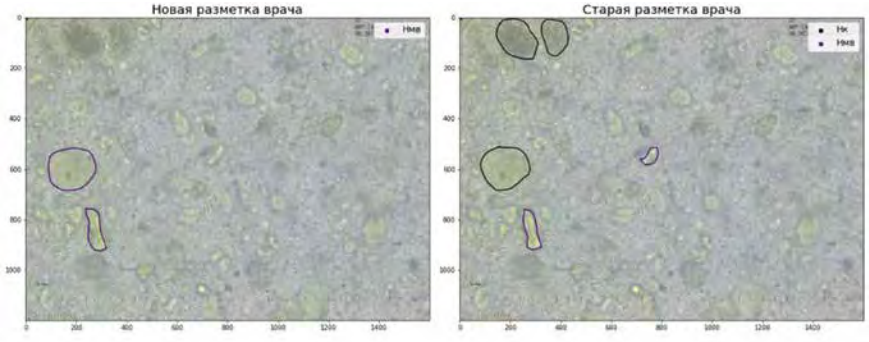

Рисунок 3В. Результат разметки одного эксперта

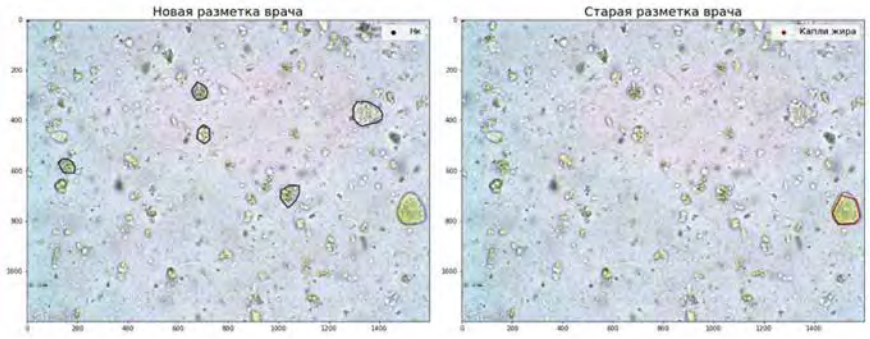

Рисунок 3С. Результат разметки одного эксперта

В *таблице 2* приведены значения F1 при попарном сравнении результатов разметки одного изображения, выполненной в разное время экспертами 1 и 2. Для количественной оценки согласованности результатов разметки также рассчитывали F1 (*рис. 1*), исходя из того, что в качестве назначенного эталона (правильно размеченные объекты) при попарном сравнении использовались результаты первичной разметки, выполненной экспертом. Далее результаты, полученные при двух разметках одного слайда, усредняли. При анализе данных в таблице можно отметить, что для приведенных примеров, воспроизводимость при разметке определялась не морфологических параметром, а индивидуальным опытом эксперта.

В данном случае значение F1 равное 0.00 означает, что при повторной разметке изображений, эксперт данный объект не выделил, хотя при первичной разметке объект присутствовал.

При последовательной разметке изображений одним экспертом также отмечены три типа несогласованности, представленные на *рисунке 3*, части A-C.

Несогласованность *типа 1* (расхождение в числе выявленных объектов). Объекты «Перевариваемая клетчатка» выделены на обоих изображениях, но их абсолютное число различается.

Несогласованность *типа 2* (расхождение в числе категорий размеченных объектов). При первичной разметке эксперт разметил объекты «Неперевариваемая клетчатка» и «Неперевариваемые мышечные волокна», при повторной оценке снимка выделены только «Неперевариваемые мышечные волокна».

Несогласованность *тип 3* (расхождение в идентификации одного и того же объекта). При первичной разметке эксперт идентифицировал на изображении объект «Капли жира», при повторной этот же объект был размечен как «Неперивариваемая клетчатка».

Заключение

Модели машинного обучения – движущая сила искусственного интеллекта. Они точны настолько, насколько точны данные, на которых они обучаются [4]. При обучении с участием МО функционал системы опирается на метки, размечаемые при обработке базы данных экспертами, которые модель искусственного интеллекта пытается предсказать. Эффективная обработка больших данных и сбор меток в нужном масштабе сопряжены со множеством сложностей. Одним из ключевых вопросов, отмечаемых при оценке медицинских изображений, являются ошибки в разметке. Неточные или непоследовательные метки приводят к созданию моделей, которые плохо работают в реальной практике. Эта проблема усугубляется при решении задач, базирующихся на субъективных суждениях, таких как диагностика изображений кала, где даже эксперты могут существенно расходиться во мнениях. Таким образом, выявление согласований и ошибок, а также приемлемых критериев при разметке объектов, необходимы для оценки эффективности выбранного метода, квалификации сотрудников и формировании экспертной группы [5]. Наше исследование по оценке согласованности результатов распознавания экспертами объектов на изображениях для копрограммы продемонстрировало высокую долю разногласий. Этот вывод не позволяет рекомендовать описанный подход для машинного обучения при анализе изображений препаратов кала.

Международное руководство по надежному и применимому искусственному интеллекту в здравоохранении указывает возможность использования консенсусного подхода при машинном обучении [6]. Маркировка изображений на основе консенсуса использует коллективные суждения экспертов для окончательной разметки. Этот подход основан на теории «мудрости толпы», использующий гипотезу, что совокупные ответы группы точнее ответов отдельных её членов за счет статистической компенсации индивидуальных ошибок и предвзятостей. Усредняя субъективные предубеждения и ошибки, консенсусная разметка может значительно повысить согласованность.

Аргументами в пользу использования в медицине маркировки на основе консенсуса выступают повышенная точность (снижение индивидуальных предубеждений), устойчивость к неоднозначности (отражение сбалансированного мнения), повышенная мотивация и производительность экспертов (состязательная нацеленность), а также гибкость и масштабируемость (вариация изображений в базе данных, задач, числа и уровня экспертов) [5]. Следующим шагом в разработке интеллектуальной программы анализа изображений кала для автоматической многофункциональной станции для анализа кала SCIENDOX 2000R будет реализация консенсусного подхода.

В последние годы наблюдается рост внедрений автоматизированных систем для диагностических исследований, требующих осмысления больших объемов данных. Машинное обучение – не волшебство, а серьезная работа в междисциплинарном пространстве, объединяющем различные методы и приемы для выявления закономерностей и принятия решений на основе баз данных.

Список литературы/ References

1. *Евгина С. А., Гусев А. В., Шаманский М. Б., Годков М. А. Искусственный интеллект на пороге лаборатории. Лабораторная служба. 2022; 11 (2): 18–26. https://doi.org/10.17116/labs20221102118*
   Evgina SA, Gusev AV, Shamanskiy MB, Godkov MA. Artificial intelligence on the doorstep of the laboratory. Laboratory Service. 2022; 11 (2): 18–26. (In Russ.) https://doi.org/10.17116/labs20221102118
2. *Nkamgang O.T., Tchiotsop D., Fotsin H.B., Talla P.K., Dorr V.L., Wolf D. Automating the clinical stools exam using image processing integrated in an expert system. Informatics in Medicine Unlocked. 2019; 15: 100165. https://doi.org/10.1016/j.imu.2019.100165*
3. *Takahashi K., Yamamoto K., Kuchiba A., Shintani A., Koyama T. Hypothesis testing procedure for binary and multi-class F1 -scores in the paired design. Stat Med.2023; 42 (23): 4177–4192. doi: 10.1002/sim.9853*
4. *Никитин Е.Д. Моделирование индивидуального стиля разметки врача-рентгенолога для улучшения точности нейронных сетей // Digital Diagnostics. 2023;4 (1S): 99–101. Nikitin ED. Learning radiologists' annotation styles with multi-annotator labeling for improved neural network performance. Digital Diagnostics. 2023; 4 (1S): 99–101. DOI: https://doi.org/10.17816/DD430358*
5. *Alonso O. Challenges with Label Quality for Supervised Learning. Journal of Data and Information Quality (JDIQ), 2015; 6 (1): 1–3. https://doi.org/10.1145/2724721.*
6. *Lekadir K., et al., FUTURE-AI Consortium FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. 2025. BMJ; 5:388: e081554. DOI: 10.1136/bmj-2024–081554.*

Статья поступила / Received 01.09.2025
Получена после рецензирования / Revised 04.09.2025
Принята в печать / Accepted 12.09.2025

Сведения об авторах

Черныш Наталия Юрьевна, *к.м.н., доцент кафедры лабораторной медицины с клиникой[1]. E-mail: nycher@mail.ru. ORCID: 0000-0002-3800-2680*

Берестовская Виктория Станиславовна, *к.м.н., доцент кафедры лабораторной медицины с клиникой[1]. E-mail: viksta@inbox.ru. ORCID: 0000-0001-5916-8076*

Тишко Анна Николаевна, *к.м.н., ассистент кафедры лабораторной медицины с клиникой[1], врач клинической лабораторной диагностики[2]. E-mail: labtishan@gmail.com. ORCID: 0009–0007–6625–597X*

Руднева Анастасия Алексеевна, *заведующая циклом «Лабораторная диагностика» Медицинского колледжа высшей школы медицины ОНК «Института медицины и наук о жизни (МЕДБИО)»[3]. E-mail: tank2605@mail.ru. ORCID: 0009-0008-6901-8192*

Вавилова Татьяна Владимировна, *д.м.н., профессор, заведующий кафедрой лабораторной медицины с клиникой[1]. E-mail: vtv.lab.spb@mail.ru. ORCID: 0000-0001-8537-3639*

[1] *ФГБУ «Национальный медицинский исследовательский центр имени В. А. Алмазова» Минздрава России, Санкт-Петербург, Россия*
[2] *ФГБУ «Всероссийский центр экстренной и радиационной медицины им А. М. Никифорова МЧС России», Санкт-Петербург, Россия*
[3] *ФГАОУ ВО «Балтийский федеральный университет имени Иммануила Канта», Калининград, Россия*

Автор для переписки: *Берестовская Виктория Станиславовна. E-mail: viksta@inbox.ru*

About authors

Chernysh Nataliia. Yu., *PhD Med, associate professor at Dept of Laboratory Medicine with Clinic[1]. E-mail: nycher@mail.ru. ORCID: 0000-0002-3800-2680*

Berestovskaya Victoria S., *PhD Med, associate professor at Dept of Laboratory Medicine with Clinic[1]. E-mail: viksta@inbox.ru. ORCID: 0000-0001-5916-8076*

Tishko Anna N., *PhD Med, Assistant at Dept of Laboratory Medicine with Clinic[1], clinical laboratory diagnostics physician[2] E-mail: labtishan@gmail.com. ORCID: 0009–0007–6625–597X*

Rudneva Anastasia A., *head of Laboratory Diagnostics Cycle[3]. E-mail: tank2605@mail.ru. ORCID: 0009-0008-6901-8192*

Vavilova Tatyana V., *DM Sci (habil.), professor, head of Dept of Laboratory Medicine with a Clinic[1]. E-mail: vtv.lab.spb@mail.ru. ORCID: 0000-0001-8537-3639*

[1] *Almazov National Medical Research Centre, St. Petersburg, Russia*
[2] *All-Russian Center for Emergency and Radiation Medicine named after A.M. Nikiforov, Ministry of Emergency Situations of Russia, St. Petersburg, Russia*
[3] *Institute of Medicine and Life Sciences, Immanuel Kant Baltic Federal University, Kaliningrad, Russia*

Corresponding author: *Berestovskaya Victoria S. E-mail: viksta@inbox.ru*

Для цитирования: Черныш Н.Ю., Берестовская В.С., Тишко А.Н., Руднева А.А., Вавилова Т.В. Влияние врачебного опыта на развитие машинного обучения при автоматизации общего исследования кала с использованием обработки изображений. *Медицинский алфавит.* 2025; (22): 65–69. https://doi.org/10.33667/2078-5631-2025-22-65-69

For citation: Chernysh N. Yu., Berestovskaya V.S., Tishko A.N., Rudneva A.A., Vavilova T.V. The impact of physician experience on the development of machine learning in automation of general fecal examination using image processing. *Medical alphabet.* 2025; (22): 65–69. https://doi.org/10.33667/2078-5631-2025-22-65-69