

ЭЛЕКТРОНИКА, ФОТОНИКА,
ПРИБОРОСТРОЕНИЕ
И СВЯЗЬ

ELECTRONICS, PHOTONICS,
INSTRUMENT MAKING
AND COMMUNICATIONS

УДК 621-822

DOI: 10.17212/2782-2001-2026-1-125-140

Интеллектуальная система анализа данных в онкогематологии*

Е.В. ПОЛЯКОВ

115409, г. Москва, Каширское шоссе, 31, Национальный исследовательский ядерный университет «МИФИ»

115478, г. Москва, Каширское шоссе, 24, Национальный медицинский исследовательский центр онкологии имени Н.Н. Блохина МЗ РФ

EVPolyakov@mephi.ru

Онкогематология сталкивается с растущей нагрузкой на диагностические лаборатории из-за роста заболеваемости и высокой трудоемкости морфологического анализа. Традиционная микроскопия характеризуется субъективностью интерпретации, межлабораторной вариабельностью и дефицитом узких специалистов, что негативно сказывается на качестве диагностики. Одной из ключевых проблем внедрения искусственного интеллекта в клиническую практику остается нехватка стандартизированных, качественно аннотированных данных и слабая интеграция с лабораторными информационными системами. В работе представлена интеллектуальная система анализа данных – модульный аппаратно-программный комплекс для анализа биомедицинских изображений, охватывающий полный цикл обработки: от загрузки изображений до формирования диагностических выводов. Система включает модули предобработки, сегментации и классификации клеток на основе гибридных архитектур глубоких нейронных сетей, извлечения морфометрических, текстурных и фрактальных признаков, а также интерактивную платформу аннотирования с поддержкой подхода «человек в цикле» (human-in-the-loop). Реализованы контроль качества разметки, версионность аннотаций, автоматический расчет коэффициентов согласия между экспертами и интеграция с внешними системами через API. Система обеспечивает ролевой доступ, деидентификацию данных и совместную работу врачей-морфологов и инженеров (специалистов по машинному обучению и разработке ИИ-систем), что способствует формированию репрезентативных и стандартизированных обучающих выборок. Система интегрирована с внешними источниками данных, что обеспечивает взаимодействие с автоматическими сканерами, открытыми репозиториями и медицинскими информационными системами. Экспериментальные исследования продемонстрировали высокую точность сегментации (Dice-коэффициент = 0,91 – стандартная метрика качества сегментации) и классификации (F1 = 0,88), а также сокращение времени на первичный анализ и аннотирование на 60...70 %. Предложенное решение способствует стандартизации, повышению объективности и воспроизводимости морфологической диагностики и может служить основой для построения единой цифровой экосистемы в онкогематологии.

* Статья получена 09 октября 2025 г.

Ключевые слова: аппаратно-программный комплекс, онкогематология, цифровая патология, анализ биомедицинских изображений, глубокое обучение, компьютерное зрение, система поддержки принятия решений, human-in-the-loop, DICOM, медицинский искусственный интеллект, морфологический анализ

ВВЕДЕНИЕ

Онкогематология как область медицины требует высокой точности, скорости и объективности при диагностике заболеваний системы крови, где ключевым элементом остается морфологический анализ микроскопических препаратов периферической крови и костного мозга [1, 2]. Современными вызовами в онкогематологии являются системные трудности, связанные с ростом заболеваемости, дефицитом квалифицированных специалистов и высокой нагрузкой на диагностические лаборатории: субъективность интерпретации, межлабораторная вариабельность и трудоемкость ручного анализа, что негативно сказывается на качестве диагностики и доступности экспертной помощи в отдаленных регионах [3, 4].

Достижения в области искусственного интеллекта, компьютерного зрения и цифровой патологии открывают возможности для автоматизации этих процессов [5, 6], однако их клиническое внедрение сдерживается рядом фундаментальных проблем: вариативностью качества изображений, неоднородностью окрашивания, сложностью морфологии клеток, их слипанием и, что наиболее критично, – дефицитом качественно аннотированных данных [7, 8]. Существующие решения, такие как QuPath, 3D Slicer и CVAT (Computer Vision Annotation Tool) – открытые платформы для анализа медицинских изображений и аннотирования данных, зачастую не адаптированы под специфику онкогематологии, не обеспечивают полноценной интеграции с клиническими процессами и не поддерживают механизмов согласования экспертных мнений [9, 10].

Создание комплексных систем для сквозной обработки биомедицинских изображений – от сбора данных до формирования диагностического заключения – является необходимым шагом для преодоления указанных ограничений. В настоящей работе представлена веб-платформа, реализующая принципы цифровой морфологии и стандартизированного управления данными, а также подход «человек в цикле» (human-in-the-loop) [11].

1. ПОСТАНОВКА ЗАДАЧИ

Целью работы является разработка единой защищенной и масштабируемой веб-платформы для анализа биомедицинских изображений в онкогематологии, способной преодолеть ключевые барьеры внедрения искусственного интеллекта в клиническую практику. Система предназначена для обеспечения полного цикла обработки данных – от загрузки изображений до автоматизированного анализа и формирования диагностического заключения с участием врача. Основные проблемы, на решение которых направлено исследование: вариативность качества изображений, неоднородность окрашивания, сложность морфологии клеток, их слипание, дефицит качественно аннотированных данных [12, 13], несбалансированность обучающих выборок

и отсутствие сквозной интеграции с лабораторными информационными системами [14].

Для устранения этих ограничений проектируется унифицированная программная архитектура, соответствующая международным стандартам цифровой медицины и обеспечивающая кроссплатформенность, отказоустойчивость и безопасность. Существующие решения зачастую привязаны к конкретным производителям оборудования, поэтому не обеспечивают должной защиты персональных данных и не поддерживают гибкую интеграцию с клиническими системами [15]. В связи с этим разрабатываемая платформа строится на основе веб-доступа через унифицированный API (Application Programming Interface – интерфейс программирования приложений), что позволяет организовать надежное взаимодействие с различными источниками данных – от цифровых микроскопов до лабораторных информационных систем, независимо от их производителя и формата хранения [16–20].

Особое внимание уделено решению проблемы дефицита аннотированных данных, критически важных для обучения и валидации моделей глубокого обучения [13]. Предлагается специализированная система аннотирования, реализованная в виде приложения с поддержкой ролевого доступа, версионности разметки, полной деидентификации данных и автоматизированного расчета коэффициентов согласия между экспертами. Интеграция предобученных моделей позволяет генерировать начальную разметку, значительно ускоряя процесс и повышая его качество [21]. Это обеспечивает формирование репрезентативных и стандартизированных, клинически релевантных обучающих выборок для задач распознавания клеточных структур в препаратах периферической крови и костного мозга [22].

Для повышения точности анализа в условиях высокой клеточной плотности и неоднородности окраски разработаны гибридные архитектуры глубокого обучения, сочетающие сверточные нейронные сети (convolutional neural networks, CNN), трансформеры с механизмом пространственного внимания (Vision Transformer, ViT) и адаптивные модели детекции (Deformable DETR) [23–26]. Такая комбинация направлена на компенсацию недостатков отдельных подходов: U-Net обеспечивает высокую точность сегментации, но теряет эффективность при высокой плотности клеток; YOLO (You Only Look Once) – архитектура нейронной сети, позволяющая быстро детектировать объекты, но склонная к пропуску мелких или частично перекрывающихся структур, в то время как ViT и Deformable DETR улучшают контекстное восприятие и обработку изображений с высокой плотностью клеток. Интеграция этих моделей в единую ансамблевую систему позволила повысить Dice-коэффициент сегментации до 0,91 и снизить количество ложных срабатываний при идентификации бластов. Согласование выводов достигается за счет механизма взвешенного голосования (weighted voting) и калибровки уверенности (Temperature Scaling). В результате система обеспечивает надежное обнаружение, сегментацию и классификацию клеток даже в клинически трудно интерпретируемых случаях.

Центральным элементом системы является реализация подхода «человек в цикле» (human-in-the-loop), при котором врачи-эксперты участвуют в анализе, корректируя результаты автоматической сегментации и классификации клеток. Эти правки не только повышают точность текущей интерпретации, но и направляются в конвейер дообучения моделей, обеспечивая их непрерывное

улучшение и адаптацию к новым данным и условиям [21, 27]. Система также включает модуль структурированного описания морфологических признаков и пространственной организации клеток, интегрированный с платформой формирования экспертного мнения, что гарантирует полноту, согласованность и воспроизводимость данных [28].

Разработанная система интегрирована в единую веб-платформу, предназначенную для централизованного хранения, аннотирования и анализа гематологических изображений с использованием методов искусственного интеллекта. Платформа находится на стадии исследовательского тестирования, однако финальный адрес развертывания может быть скорректирован в зависимости от условий внедрения и инфраструктурных требований.

Предусмотрена комплексная экспериментальная оценка разработанных методов и архитектуры по критериям точности, производительности, клинической полезности и воспроизводимости. Планируется также разработка рекомендаций по интеграции технологии в практику здравоохранения, включая возможности ее использования в рамках программ импортозамещения и цифровой трансформации медицинских процессов.

2. АРХИТЕКТУРА И ФУНКЦИОНАЛЬНЫЕ КОМПОНЕНТЫ СИСТЕМЫ

Разработанная система представляет собой модульный программно-аппаратный комплекс, реализованный на основе веб-архитектуры и охватывающий все этапы анализа – от загрузки изображений до формирования структурированных диагностических выводов (рис. 1)¹.

Система построена по многоуровневому принципу и включает четыре основных уровня: пользовательский интерфейс, бизнес-логику, обработку данных и интеграцию с внешними источниками. Пользовательский интерфейс построен на основе React и обеспечивает доступ к функциям загрузки, просмотра, аннотирования и анализа изображений. Бизнес-логика реализована на Python с использованием фреймворка Django и отдельных модулей на PHP для тестирования; она отвечает за управление сессиями, ролями пользователей, версию данных и аудит операций.

Уровень обработки данных включает модули предобработки (OpenCV, scikit-image), сегментации (U-Net, Mask R-CNN, YOLO) и классификации (ResNet, EfficientNet и другие архитектуры), а также извлечение морфометрических и текстурных признаков.

В работе реализован гибридный подход, объединяющий два направления анализа:

- глубокое обучение с применением архитектур U-Net, YOLO, ViT и Deformable DETR, где признаки извлекаются автоматически нейросетевыми слоями;
- классическое машинное обучение (SVM, Random Forest, XGBoost), использующее более 200 морфометрических, текстурных и фрактальных признаков, извлекаемых с помощью OpenCV и scikit-image.

¹ Система развёрнута в тестовом режиме и доступна по адресу <https://imageset.iunknown.ru>

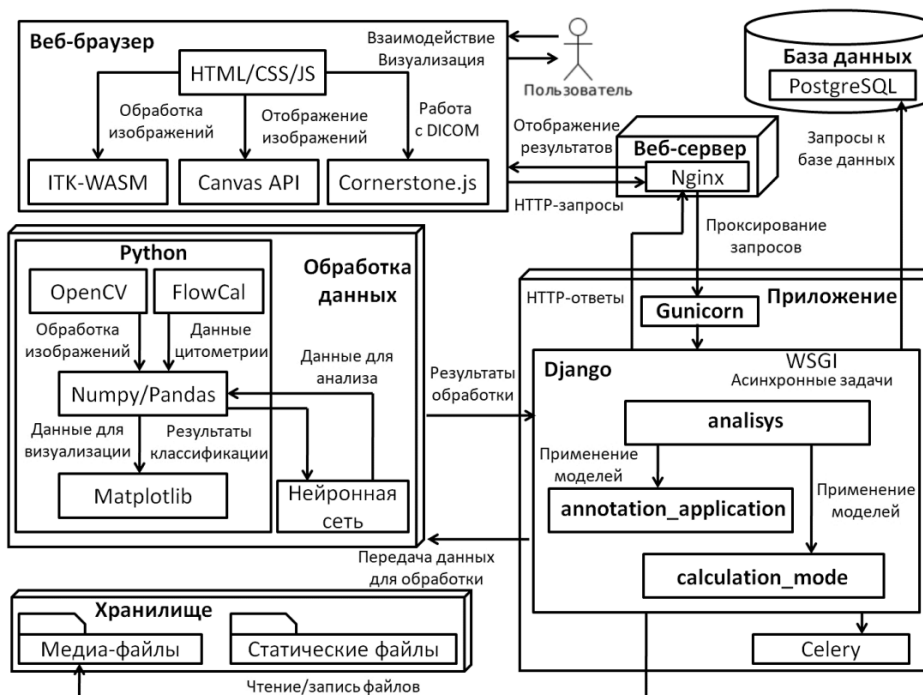


Рис. 1. Интеллектуальная система обработки и анализа биомедицинских данных в онкогематологии

Fig. 1. An intelligent system for processing and analysis of biomedical data in onco hematology

Такой подход позволяет не только сравнивать эффективность современных и традиционных методов, но и обеспечивать интерпретируемость результатов для врача, а также поддерживать работу системы в условиях ограниченных вычислительных ресурсов. При разработке системы учитывались требования ГОСТ^{2,3}.

Уровень интеграции обеспечивает взаимодействие с автоматическими сканерами, открытыми репозиториями, лабораторными (ЛИС) и медицинскими информационными системами (МИС) через поддержку международных стандартов обмена медицинской информацией: DICOM (Digital Imaging and Communications in Medicine) для изображений, HL7 (Health Level Seven) как основной стандарт передачи клинических данных и FHIR (Fast Healthcare Interoperability Resources) – современный профиль HL7 для веб-ориентированных приложений [29], а также протоколов RESTful API [19].

Такая модульная и стандартизированная архитектура обеспечивает масштабируемость, гибкость и соответствие требованиям клинической практики.

² ГОСТ Р 71671–2024. Системы поддержки принятия врачебных решений с применением искусственного интеллекта. Основные положения: национальный стандарт Российской Федерации. М.: Росстандарт, 2024. 28 с.

³ ГОСТ Р 71737–2024. Системы искусственного интеллекта в здравоохранении. Применение менеджмента риска к медицинским изделиям. Общие положения: национальный стандарт Российской Федерации. М.: Росстандарт, 2024. 36 с. (Введен в действие приказом Федерального агентства по техническому регулированию и метрологии от 28 октября 2024 г. № 1518-ст; дата введения 1 января 2025 г.).

2.1. МЕТОДЫ АНАЛИЗА БИОМЕДИЦИНСКИХ ИЗОБРАЖЕНИЙ

Анализ биомедицинских изображений в системе реализован в виде конвейера, охватывающего все этапы – от предобработки до интерпретации (рис. 2).

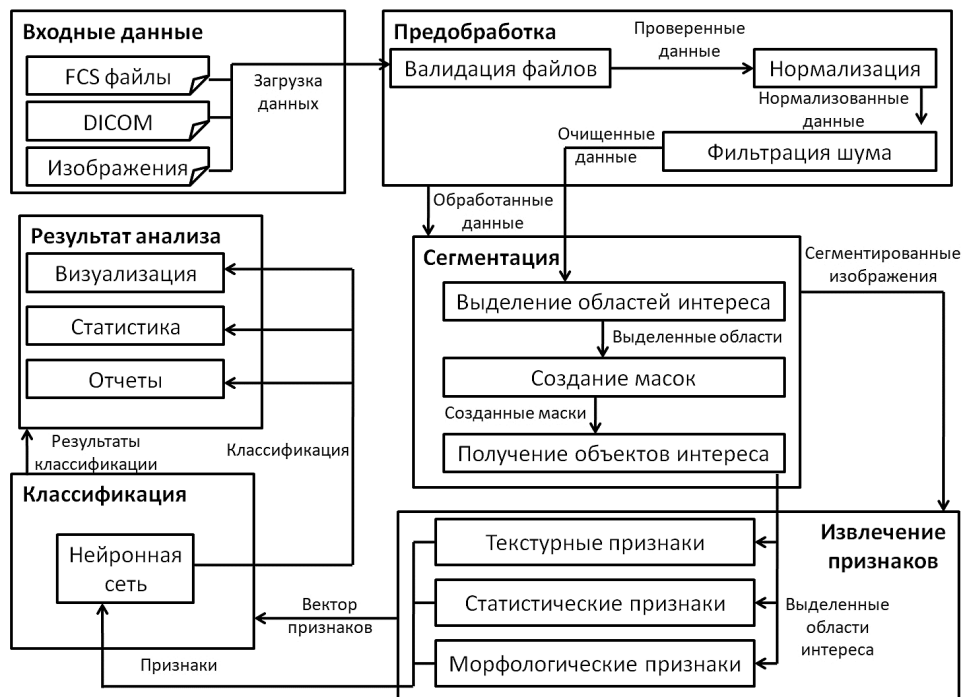


Рис. 2. Модель анализа биомедицинских изображений в онкогематологии

Fig. 2. Model for the analysis of biomedical images in oncohematology

На этапе предобработки выполняется нормализация цвета (методы Масенко, Vahadane), фильтрация шумов (медианный, гауссов), коррекция освещения и устранение артефактов. Сегментация ядродержащих клеток осуществляется с использованием комбинированного подхода: классические методы (watershed, пороговая сегментация) применяются для первичного выделения объектов, а глубокие нейронные сети (U-Net, Mask R-CNN) – для уточнения границ, особенно в условиях слипания клеток.

Классификация проводится с помощью YOLO, адаптированной для задач медицинской визуализации, и архитектур на основе ResNet и EfficientNet. Особое внимание уделено интерпретируемости решений ИИ: для визуализации «важных» областей используются методы Grad-CAM и LIME, что повышает доверие врачей к системе [30, 31]. Извлечение признаков включает расчет более 200 параметров, описывающих морфологию клеток: морфометрические (площадь, периметр, компактность, соотношение ядро/цитоплазма), текстурные (GLCM, GLRLM) и фрактальные признаки по методикам Метце и Кошта. Эти данные формируют основу для построения моделей машинного обучения и объективной количественной оценки морфологии клеток [28, 32, 33].

2.2. СИСТЕМА АННОТИРОВАНИЯ И УПРАВЛЕНИЕ ПРАВАМИ

Система аннотирования является ключевым элементом экосистемы, обеспечивающим накопление высококачественной, клинически релевантной обучающей выборки. Ранняя версия реализована в виде веб-приложения на базе PHP и MySQL, а после модернизации – на Python и PostgreSQL [21], что позволяет экспертам проводить совместное аннотирование, контролировать качество разметки, запускать модели анализа и корректировать их результаты. Пользователи с разными ролями (врач, администратор ЛПУ, инженер по знаниям) имеют соответствующие уровни доступа, что обеспечивает безопасность и целостность данных.

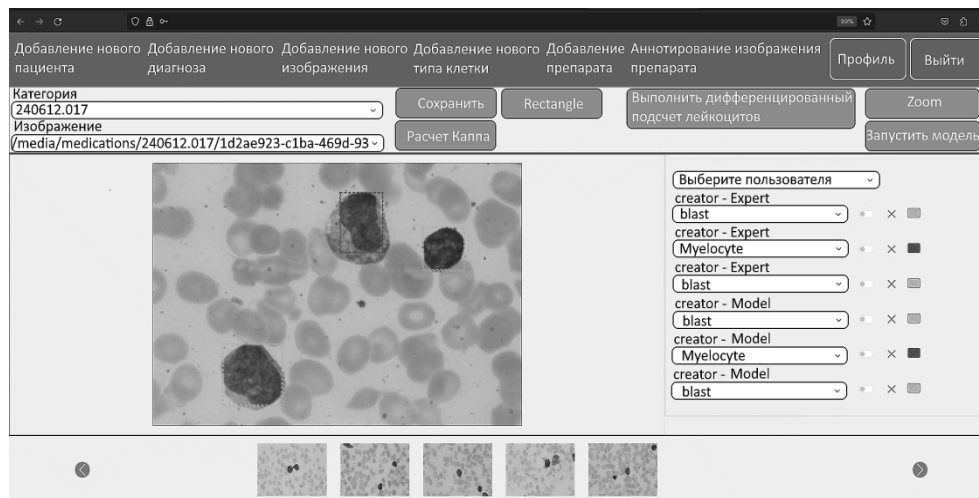


Рис. 3. Пример установления типа выбранной клетки посредством графического примитива и отображения результатов проведенной экспертами аннотации

Fig. 3. Example of determining a cell type through a graphic primitive and displaying results of expert annotation

Система поддерживает несколько слоев аннотаций (эксперт А, эксперт В, модель), что позволяет визуально сравнивать метки и выявлять расхождения. После аннотирования система автоматически рассчитывает коэффициенты согласия (Cohen's Каппа), что критически важно для оценки надежности данных [3].

На рис. 3 показан фрагмент структуры интерфейса и практическое использование платформы для обработки клинических данных. Дополнительно реализован модуль статистики, позволяющий анализировать распределение клеточных популяций по каждому изображению, что помогает в формировании диагностического заключения.

2.3. РЕАЛИЗАЦИЯ И ВЗАИМОДЕЙСТВИЕ С ПЛАТФОРМОЙ

Практическая реализация выполнена в виде веб-платформы, объединяющей все этапы анализа. Клиентская часть (frontend) разработана с использованием JavaScript-библиотеки React, что обеспечивает высокую отзывчивость интерфейса при работе с изображениями высокого разрешения. Серверная часть (backend) реализована на фреймворке Django (Python) и включает тестовые модули на языке PHP, обеспечивая обработку запросов, управление сессиями

и взаимодействие с базой данных. В качестве хранилища используется PostgreSQL для структурированных данных и MinIO (S3-совместимое хранилище) для бинарных файлов. Для асинхронной обработки ресурсоемких задач (например, сегментации изображений высокого разрешения) используется Celery – система распределенного выполнения фоновых задач. Celery передает задания через Redis – специализированную службу обмена сообщениями, которая временно хранит задачи и распределяет их между вычислительными узлами. Это обеспечивает высокую отзывчивость интерфейса и масштабируемость обработки. Реализованы RESTful API для поддержания возможности интеграции с ЛИС, МИС и автоматическими сканерами.

Поддерживается экспорт данных в форматах JSON, TXT и DICOM SR для совместимости с международными стандартами. Интерфейс поддерживает каскадную загрузку изображений (tile-based rendering), что позволяет эффективно работать с WSI (Whole Slide Images) [16, 31, 34–36]. Платформа также поддерживает подход «человек в цикле» (human-in-the-loop): врач может скорректировать результаты работы искусственного интеллекта, и эти правки автоматически направляются в систему дообучения для непрерывного улучшения моделей. Схема реализации принципа «человек в цикле» (human-in-the-loop) в интеллектуальной системе анализа данных в онкогематологии представлена на рис. 4.

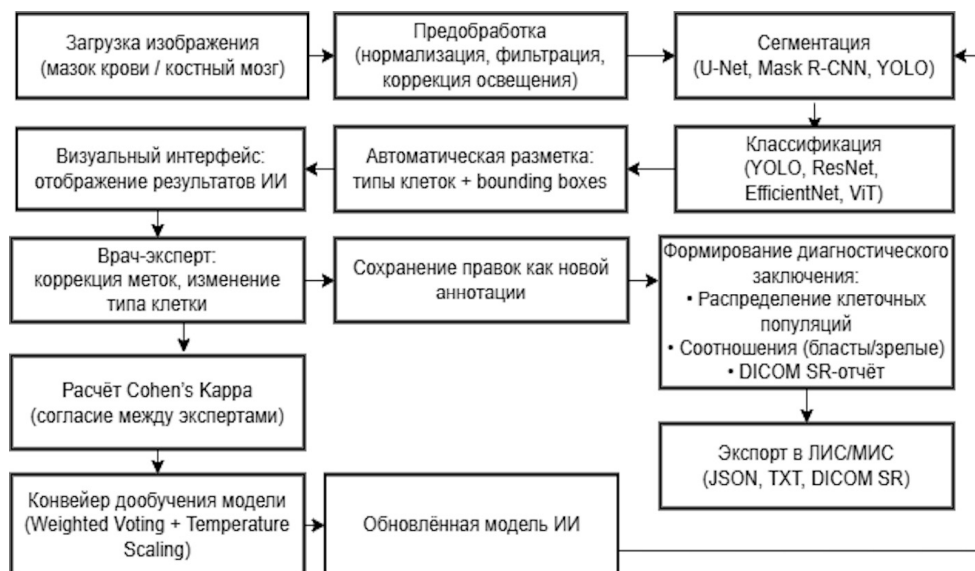


Рис. 4. Схема реализации принципа «человек в цикле» (human-in-the-loop) в интеллектуальной системе анализа данных в онкогематологии

Fig. 4. Scheme of the human-in-the-loop principle implementation in the intelligent data analysis system for oncohematology

Для обучения и валидации моделей глубокого обучения использовалась вычислительная инфраструктура, включающая сервер с GPU NVIDIA RTX 4060 16GB, распределенную файловую систему и фреймворки PyTorch, TensorFlow и MONAI [22].

Система решает задачу автоматической классификации клеток периферической крови и костного мозга на 22 морфологических типа, включая патологические формы (бласты – миелоидные и лимфоидные, промиелоциты, атипичные лимфоциты), зрелые клетки (нейтрофилы, эозинофилы, базофилы, лимфоциты, моноциты) и клетки-предшественники. Такой подход позволяет одновременно дифференцировать более двух диагностических категорий, что принципиально отличается от простой бинарной классификации («здоровый/больной»). Классификация соответствует международным стандартам FАВ и рекомендациям ВОЗ по диагностике онкогематологических заболеваний.

Процесс обучения организован с использованием инструментов отслеживания экспериментами (Weights & Biases) – это обеспечивает воспроизводимость и контроль за гиперпараметрами. Обучение всех моделей проводилось на аннотированных выборках, сформированных в платформе, размещенной на домене iunknown.ru (поддомены для каждой задачи свои), с применением аугментации и стратифицированного разделения на обучающую, валидационную и тестовую выборки.

3. ЭКСПЕРИМЕНТАЛЬНЫЕ РЕЗУЛЬТАТЫ

Экспериментальные исследования проводились на выборке из 1200 изображений мазков костного мозга пациентов с острыми лейкозами, сформированной в НМИЦ онкологии им. Н.Н. Блохина. В аннотировании данных участвовали 5 экспертов-морфологов. Оценка качества сегментации показала среднее значение Dice-коэффициента = 0,91 (IoU = 0,84), что свидетельствует о высокой точности выделения клеток системой.

Для классификации 10 типов клеток достигнута F1-мера = 0,88 с наилучшими результатами для миелоидных и лимфоидных линий. Среднее время обработки одного изображения составило 3...5 секунд. После трех циклов дообучения на коррекциях экспертов-гематологов F1-мера выросла на 7 %, что подтверждает эффективность подхода «человек в цикле» (human-in-the-loop). Уровень согласия между экспертами (Cohen's Kappa) составил 0,85, что указывает на высокую надежность разметки.

На выходе система формирует комплексный диагностический отчет, включающий:

- количественный подсчет клеток по 22 морфологическим типам;
- расчет ключевых соотношений (бласты/зрелые клетки, ядродержащие/эритроидные и др.);
- диагностические метки с указанием вероятности каждого подтипа лейкоза;
- структурированное заключение в формате DICOM SR, совместимое с ЛИС и МИС (опционально);
- визуальные карты Grad-CAM и LIME для объяснения решения модели (опционально).

Таким образом, платформа позволяет формировать не только вероятностную оценку патологии, но и полноценный морфологический профиль миелограммы, сопоставимый с заключением врача-морфолога.

В ходе пилотного внедрения в НМИЦ онкологии им. Н.Н. Блохина было установлено, что модель корректно классифицирует клетки в 91 % случаев (F1-мера = 0,88) и сокращает время на этапе первичного анализа и аннотирования на 60...70 %, что подтверждается оценкой экспертов-морфологов. Оценка времени проводилась в ходе исследования с участием врачей-морфологов (опыт работы от 5 до 20 лет). Каждому эксперту предлагалось проанализировать 10 случайных мазков крови (100 изображений для каждого мазка) дважды: с использованием традиционной микроскопии и с применением разработанной системы. Среднее время на первичный анализ и аннотирование одного мазка составило $42,3 \pm 5,1$ минуты вручную и $13,7 \pm 2,8$ минуты с использованием системы, что соответствует сокращению времени на анализ препарата на 67,6 %.

ЗАКЛЮЧЕНИЕ

Разработанная интеллектуальная система анализа данных в онкогематологии представляет собой законченное масштабируемое и клинически интегрированное решение, объединяющее передовые методы искусственного интеллекта, стандартизированное хранение данных и принцип «человек в цикле». Система эффективно решает ключевые барьеры внедрения ИИ в медицинскую практику: дефицит кадров, дефицит качественно аннотированных данных, субъективность интерпретации результатов диагностики, слабую интеграцию ИИ-решений с лабораторными информационными системами и недостаточную прозрачность алгоритмов.

Практическая апробация подтвердила высокую точность, воспроизводимость и объективность результатов, а также позволила сократить время на этапах первичного анализа и аннотирования изображений на 60...70 %. Система обеспечивает формирование стандартизированных и репрезентативных обучающих выборок, что является важным шагом к построению единой федеральной системы цифровой морфологии в онкогематологии.

Реализация системы будет способствовать цифровой трансформации здравоохранения и развитию отечественной инфраструктуры медицинских информационных систем.

В дальнейшем планируется расширение функциональности платформы за счет интеграции в нее данных с проточной цитометрии и геномики, а также развитие механизмов активного обучения. Уникальная комбинация защищенной платформы, накопленная база данных и адаптивных алгоритмов искусственного интеллекта позволяет рассматривать ее как экосистему высокотехнологичной, объективной и доступной диагностики гематологических заболеваний в масштабах страны.

СПИСОК ЛИТЕРАТУРЫ

1. Application of artificial intelligence in laboratory hematology: advances, challenges, and prospects / H. Liao, F. Zhang, F. Chen, Y. Li, Y. Sun, D.D. Sloboda, Q. Zheng, B. Ying, T. Hu // *Acta Pharmaceutica Sinica B*. – 2025. – Vol. 15 (11). – P. 5702–5733. – DOI: 10.1016/j.apsb.2025.05.036.

2. An overview and a roadmap for artificial intelligence in hematology and oncology / W. Rösler, M. Altenbuchinger, B. Baeßler, T. Beissbarth, G. Beutel, R. Bock, N. von Bubnoff, J.-N. Eckardt, S. Foersch, C.M.L. Loeffler, J.M. Middeke, M.-L. Mueller, T. Oellerich, B. Risse, A. Scherag, C. Schliemann, M. Scholz, R. Spang, C. Thielscher, I. Tsoukakis, J.N. Kather // *Journal of Cancer Research and Clinical Oncology*. – 2023. – Vol. 149 (10). – P. 7997–8006. – DOI: 10.1007/s00432-023-04667-5.
3. Machine learning and artificial intelligence in haematology / R. Shouval, J.A. Fein, B. Savani, M. Mohty, A. Nagler // *British Journal of Haematology*. – 2020. – Vol. 192. – DOI: 10.1111/bjh.16915.
4. Machine learning in detection and classification of leukemia using smear blood images: a systematic review / M. Ghaderzadeh, F. Asadi, A. Hosseini, D. Bashash, H. Abolghasemi, A. Roshanpour // *Scientific Programming*. – 2021. – P. 9933481. – DOI: 10.1155/2021/9933481.
5. A guide to deep learning in healthcare / A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, J. Dean // *Nature Medicine*. – 2019. – Vol. 25. – P. 24–29. – DOI: 10.1038/s41591-018-0316-z.
6. A survey on deep learning in medical image analysis / G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A.W.M. van der Laak, B. van Ginneken, C.I. Sánchez // *Medical Image Analysis*. – 2017. – Vol. 42. – P. 60–88. – DOI: 10.1016/j.media.2017.07.005.
7. Lewis J.E., Pozdnyakova O. Digital assessment of peripheral blood and bone marrow aspirate smears // *International Journal of Laboratory Hematology*. – 2023. – Vol. 45. – P. 47–55. – DOI: 10.1111/ijlh.14082.
8. A large-scale multi domain leukemia dataset for the white blood cells detection with morphological attributes for explainability / A. Rehman, T. Meraj, A.M. Minhas, A. Imran, M. Ali, W. Sultani // *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*. – Cham: Springer, 2024. – P. 553–563. – DOI: 10.1007/978-3-031-72384-1_52.
9. QuPath: Open source software for digital pathology image analysis / P. Bankhead, M.B. Loughrey, J.A. Fernández, Y. Dombrowski, D.G. McArt, P.D. Dunne, S. McQuaid, R.T. Gray, L.J. Murray, H.G. Coleman, J.A. James, M. Salto-Tellez, P.W. Hamilton // *Scientific Reports*. – 2017. – Vol. 7. – P. 16878. – DOI: 10.1038/s41598-017-17204-5.
10. A review of the automated detection and classification of acute leukaemia: coherent taxonomy, datasets, validation and performance measurements / M.A. Alsalem, A.A. Zaidan, B.B. Zaidan, M. Hashim, H.T. Madhloom, N.D. Azeez, S. Alsysisuf // *Computer Methods and Programs in Biomedicine*. – 2018. – Vol. 158. – P. 93–112. – DOI: 10.1016/j.cmpb.2018.02.005.
11. Topol E.J. High-performance medicine: the convergence of human and artificial intelligence // *Nature Medicine*. – 2019. – Vol. 25. – P. 44–56. – DOI: 10.1038/s41591-018-0300-7.
12. Artificial intelligence in hematology / A. Nazha, O. Elemento, S. Ahuja, B. Lam, M. Miles, R. Shouval, S. McWeeney, S. Sirhan, A. Srisuwananukorn, T. Haferlach // *Blood*. – 2025. – Vol. 146 (19). – P. 2283–2292. – DOI: 10.1182/blood.2025029876.
13. Wilm F.E. Cross-domain generalization of deep learning-based image analysis algorithms in histopathology: Dissertation. – Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany, 2025.
14. Validating whole slide imaging for diagnostic purposes in pathology / L. Pantanowitz, J.H. Sinaard, W.H. Henricks, L.A. Fatheree, A.B. Carter, L. Contis, B.A. Beckwith, A.J. Evans, A. Lal, A.V. Parwani // *Archives of Pathology & Laboratory Medicine*. – 2013. – Vol. 137. – P. 1710–1722. – DOI: 10.5858/arpa.2013-0093-CP.
15. European Society of Radiology (ESR). Methodology for ESR iGuide content // *Insights into Imaging*. – 2019. – Vol. 10 (1). – P. 32. – PMID: 30868370; PMCID: PMC6419665. – DOI: 10.1186/s13244-019-0720-z.
16. Farahani N., Parwani A.V., Pantanowitz L. Whole slide imaging in pathology: advantages, limitations, and emerging perspectives // *Pathology and Laboratory Medicine International*. – 2015. – Vol. 7. – P. 23–33. – DOI: 10.2147/PLMIS59826.
17. Hover-Net: simultaneous segmentation and classification of nuclei in multi-tissue histology images / S. Graham, Q.D. Vu, S.E.A. Raza, A. Azam, Y.W. Tsang, J.T. Kwak, N. Rajpoot // *Medical Image Analysis*. – 2019. – Vol. 58. – P. 101563. – DOI: 10.1016/j.media.2019.101563.
18. Fine-grained imbalanced leukocyte classification with global-local attention transformer / B. Chen, F. Qin, Y. Shao, J. Cao, Y. Peng, R. Ge // *Journal of King Saud University – Computer and Information Sciences*. – 2023. – Vol. 35 (8). – P. 101661. – DOI: 10.1016/j.jksuci.2023.101661.

19. SliDL: A toolbox for processing whole-slide images in deep learning / A.G. Berman, W.R. Orchard, M. Gehrung, F. Markowetz // PLoS ONE. – 2023. – Vol. 18. – P. e0289499. – DOI: 10.1371/journal.pone.0289499.
20. *Cohen J.* A coefficient of agreement for nominal scales // Educational and Psychological Measurement. – 1960. – Vol. 20 (1). – P. 37–46. – DOI: 10.1177/001316446002000104.
21. Система аннотирования биомедицинских изображений на основе интернеттехнологий в онкогематологии / В.В. Дмитриева, Я.А. Магманова, В.И. Цыпльак, Е.В. Поляков // International Journal of Open Information Technologies. – 2025. – Т. 13, № 3. – С. 52–57.
22. Исследование одноэтапной нейронной сети глубокого обучения для анализа биомедицинских изображений / Е.В. Поляков, В.В. Дмитриева, Е.В. Шувалова, Н.А. Филатова, О.П. Колбацкая, И.А. Климанов // Фундаментальные и прикладные проблемы техники и технологии. – 2025. – Т. 371, № 3. – С. 118–125.
23. *Kouzehkanan, Z.M., Saghari, S., Tavakoli, S., Rostami, P., Abaszadeh, M., Mirzadeh, F., Shahabi Satsar, E., Gheidishahran, M., Gorgi, F., Mohammadi, S., & Hosseini, R.* A Large dataset of white blood cells containing cell locations and types, along with segmented nuclei and cytoplasm / Z.M. Kouzehkanan, S. Saghari, S. Tavakoli, P. Rostami, M. Abaszadeh, F. Mirzadeh, E.S. Satsar, M. Gheidishahran, F. Gorgi, S. Mohammadi, R. Hosseini // Scientific Reports. – 2022. – Vol. 12. – P. 1123. – DOI: 10.1038/s41598-021-04426-x.
24. *Tsutsui S., Pang W., Wen B.* WBCAtt: a white blood cell dataset annotated with detailed morphological attributes // arXiv. 2023. – arXiv:2306.13531.
25. End-to-end object detection with transformers / N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, S. Zagoruyko // Computer Vision – ECCV 2020. – Cham: Springer, 2020. – P. 213–229. – (Lecture Notes in Computer Science; vol. 12346). – DOI: 10.1007/978-3-030-58452-8_13.
26. Deformable DETR: deformable transformers for end-to-end object detection / X. Zhu, W. Su, L. Lu, B. Li, X. Wang, J. Dai // ICLR 2021 Conference Paper. – DOI: 10.48550/arXiv.2010.04159.
27. Causability and explainability of artificial intelligence in medicine / A. Holzinger, G. Langs, H. Denk, K. Zatloukal, H. Müller // WIREs Data Mining and Knowledge Discovery. – 2019. – Vol. 9 (4). – P. e1312. – DOI: 10.1002/widm.1312.
28. *Поляков Е.В., Дмитриева В.В.* Система анализа текстуры биомедицинских изображений // Модели, системы, сети в экономике, технике, природе и обществе. – 2025. – № 2. – С. 85–94.
29. Whole slide image representation in bone marrow cytology / Y. Mu, H.R. Tizhoosh, T. Dehkharghanian, C.J. Campbell // bioRxiv. 2022. – DOI: 10.1101/2022.12.06.519318.
30. Grad-CAM: visual explanations from deep networks via gradient-based localization / R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra // Proceedings of the IEEE International Conference on Computer Vision (ICCV). – IEEE, 2017. – P. 618–626. – DOI: 10.1109/ICCV.2017.74.
31. *Ribeiro M.T., Singh S., Guestrin C.* "Why Should I Trust You?": Explaining the predictions of any classifier // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. – ACM, 2016. – P. 1135–1144. – DOI: 10.1145/2939672.2939778.
32. *Haralick R.M., Shanmugam K., Dinstein I.* Textural features for image classification // IEEE Transactions on Systems, Man, and Cybernetics. – 1973. – Vol. SMC-3 (6). – P. 610–621. – DOI: 10.1109/TSMC.1973.4309314.
33. *Costa L. da F., Cesar R.M.* Shape classification and analysis: theory and practice. – 2nd ed. – CRC Press, 2009.
34. *Ronneberger O., Fischer P., Brox T.* U-Net: convolutional networks for biomedical image segmentation // Medical Image Computing and Computer Assisted Intervention – MICCAI 2015. Pt. 3. – Cham: Springer, 2015. – P. 234–241. – DOI: 10.1007/978-3-319-24574-4_28.
35. Deep residual learning for image recognition / K. He, X. Zhang, S. Ren, J. Sun // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – IEEE, 2016. – P. 770–778. – DOI: 10.1109/CVPR.2016.90.
36. *Settles B.* Active learning. – Springer, 2012. – 114 p. – (Synthesis Lectures on Artificial Intelligence and Machine Learning). – DOI: 10.2200/S00429ED1V01Y201207AIM018.

Поляков Евгений Валерьевич, кандидат технических наук, доцент кафедры компьютерных систем и технологий Национального исследовательского ядерного университета. Основное направление научных исследований – формирование, обоснование и разработка технологических решений для проведения исследований в области медицины и техники. Имеет более 170 печатных работ. E-mail: EVPolyakov@mephi.ru

Polyakov Evgeny V., PhD (Eng.), associate professor at the Department of Computer Systems and Technologies, National Research Nuclear University MEPHI (Moscow Engineering Physics Institute). His main research focuses on formation, justification, and development of technological solutions for conducting studies in medicine and technology. He is the author of more than 170 publications. E-mail: EVPolyakov@mephi.ru

DOI: 10.17212/2782-2001-2026-1-125-140

Intelligent data analysis system in oncohematology*

E.V. POLYAKOV

*National Research Nuclear University «MEPhI» (Moscow Engineering Physics Institute), 31 Kashirskoye Shosse, Moscow, 115409, Russian Federation
National Medical Research Center of Oncology named after N.N. Blokhin, Ministry of Health of the Russian Federation, Russia, 115478, Moscow, Kashirskoye Shosse, house 24*

EVPolyakov@mephi.ru

Abstract

Oncological hematology is facing an increasing pressure on diagnostic laboratories due to rising disease incidence and the high labor intensity of morphological analysis. Traditional microscopy is characterized by subjective interpretation, inter-laboratory variability, and a shortage of specialized experts, which negatively impacts diagnostic quality. A key challenge in implementing artificial intelligence (AI) in clinical practice remains the lack of standardized, high-quality annotated data and poor integration with laboratory information systems. This paper presents an intelligent data analysis system – a modular hardware-software complex for biomedical image analysis that covers the full processing cycle from image loading to generating diagnostic conclusions. The system incorporates modules for image preprocessing, cell segmentation, and classification based on hybrid deep neural network architectures, as well as extraction of morphometric, textural, and fractal features. It also includes an interactive annotation platform supporting the human-in-the-loop approach. The platform implements quality control of annotations, versioning, automatic calculation of inter-expert agreement coefficients (e.g., Cohen's Kappa), and integration with external systems via APIs. Role-based access, data de-identification, and collaborative workflows between morphologists and knowledge engineers (i.e., machine learning specialists and AI developers) facilitate the creation of representative and standardized training datasets. The system is integrated with external data sources, enabling seamless interaction with automated scanners, open repositories, and medical information systems. Experimental studies demonstrated high segmentation accuracy (Dice coefficient = 0.91 is a standard segmentation quality metric based on the overlap between predicted and ground-truth masks) and classification performance (F1 = 0.88), along with a 60–70 % reduction in time required for primary analysis and

* Received 09 October 2025.

annotation. The proposed solution promotes standardization, enhances objectivity and reproducibility of morphological diagnostics, and can serve as a foundation for a unified digital ecosystem in oncohematology.

Keywords: hardware-software complex, oncohematology, digital pathology, biomedical image analysis, deep learning, computer vision, decision support system, human-in-the-loop, DICOM, medical artificial intelligence, morphological analysis

REFERENCES

1. Liao H., Zhang F., Chen F., Li Y., Sun Y., Sloboda D.D., Zheng Q., Ying B., Hu T. Application of artificial intelligence in laboratory hematology: advances, challenges, and prospects. *Acta Pharmaceutica Sinica B*, 2025, vol. 15 (11), pp. 5702–5733. DOI: 10.1016/j.apsb.2025.05.036.
2. Rösler W., Altenbuchinger M., Baeßler B., Beissbarth T., Beutel G., Bock R., von Bubnoff N., Eckardt J.-N., Foersch S., Loeffler C.M.L., Middeke J.M., Mueller M.-L., Oellerich T., Risse B., Scherag A., Schliemann C., Scholz M., Spang R., Thielscher C., Tsoukakis I., Kather J.N. An overview and a roadmap for artificial intelligence in hematology and oncology. *Journal of Cancer Research and Clinical Oncology*, 2023, vol. 149 (10), pp. 7997–8006. DOI: 10.1007/s00432-023-04667-5.
3. Shouval R., Fein J.A., Savani B., Mohty M., Nagler A. Machine learning and artificial intelligence in haematology. *British Journal of Haematology*, 2020, vol. 192. DOI: 10.1111/bjh.16915.
4. Ghaderzadeh M., Asadi F., Hosseini A., Bashash D., Abolghasemi H., Roshanpour A. Machine Learning in detection and classification of leukemia using smear blood images: a systematic review. *Scientific Programming*, 2021, p. 9933481. DOI: 10.1155/2021/9933481.
5. Esteva A., Robicquet A., Ramsundar B., Kuleshov V., DePristo M., Chou K., Cui C., Corrado G., Thrun S., Dean J. A guide to deep learning in healthcare. *Nature Medicine*, 2019, vol. 25, pp. 24–29. DOI: 10.1038/s41591-018-0316-z.
6. Litjens G., Kooi T., Bejnordi B.E., Setio A.A.A., Ciampi F., Ghafoorian M., van der Laak J.A.W.M., van Ginneken B., Sánchez C.I. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 2017, vol. 42, pp. 60–88. DOI: 10.1016/j.media.2017.07.005.
7. Lewis J.E., Pozdnyakova O. Digital assessment of peripheral blood and bone marrow aspirate smears. *International Journal of Laboratory Hematology*, 2023, vol. 45, pp. 47–55. DOI: 10.1111/ijlh.14082.
8. Rehman A., Meraj T., Minhas A.M., Imran A., Ali M., Sultani W. A large-scale multi domain leukemia dataset for the white blood cells detection with morphological attributes for explainability. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*. Cham, Springer, 2024, pp. 553–563. DOI: 10.1007/978-3-031-72384-1_52.
9. Bankhead P., Loughrey M.B., Fernández J.A., Dombrowski Y., McArt D.G., Dunne P.D., McQuaid S., Gray R.T., Murray L.J., Coleman H.G., James J.A., Salto-Tellez M., Hamilton P.W. QuPath: Open source software for digital pathology image analysis. *Scientific Reports*, 2017, vol. 7, p. 16878. DOI: 10.1038/s41598-017-17204-5.
10. Alsalem M.A., Zaidan A.A., Zaidan B.B., Hashim M., Madhloom H.T., Azeez N.D., Al-syisuf S. A review of the automated detection and classification of acute leukaemia: coherent taxonomy, datasets, validation and performance measurements. *Computer Methods and Programs in Biomedicine*, 2018, vol. 158, pp. 93–112. DOI: 10.1016/j.cmpb.2018.02.005.
11. Topol E.J. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 2019, vol. 25, pp. 44–56. DOI: 10.1038/s41591-018-0300-7.
12. Nazha A., Elemento O., Ahuja S., Lam B.D., Miles M., Shouval R., McWeeney S.K., Sirhan S., Srisuwananukorn A., Haferlach T. Artificial intelligence in hematology. *Blood*, 2025, vol. 146 (19), pp. 2283–2292. DOI: 10.1182/blood.2025029876.
13. Wilm F.E. *Cross-domain generalization of deep learning-based image analysis algorithms in histopathology*. Dissertation. Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany, 2025.
14. Pantanowitz L., Sinard J.H., Henricks W.H., Fatheree L.A., Carter A.B., Contis L., Beckwith B.A., Evans A.J., Otis C.N., Lal A., Parwani A.V. Validating whole slide imaging for diagnostic

purposes in pathology. *Archives of Pathology & Laboratory Medicine*, 2013, vol. 137, pp. 1710–1722. DOI: 10.5858/arpa.2013-0093-CP.

15. European Society of Radiology (ESR). Methodology for ESR iGuide content. *Insights into Imaging*, 2019, vol. 10 (1), p. 32. PMID: 30868370. PMCID: PMC6419665. DOI: 10.1186/s13244-019-0720-z.

16. Farahani N., Parwani A.V., Pantanowitz L. Whole slide imaging in pathology: advantages, limitations, and emerging perspectives. *Pathology and Laboratory Medicine International*, 2015, vol. 7, pp. 23–33. DOI: 10.2147/PLMI.S59826.

17. Graham S., Vu Q.D., Raza S.E.A., Azam A., Tsang Y.W., Kwak J.T., Rajpoot N. Hover-Net: simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Analysis*, 2019, vol. 58, p. 101563. DOI: 10.1016/j.media.2019.101563.

18. Chen B., Qin F., Shao Y., Cao J., Peng Y., Ge R. Fine-grained imbalanced leukocyte classification with global-local attention transformer. *Journal of King Saud University – Computer and Information Sciences*, 2023, vol. 35 (8), p. 101661. DOI: 10.1016/j.jksuci.2023.101661.

19. Berman A.G., Orchard W.R., Gehrung M., Markowetz F. SliDL: A toolbox for processing whole-slide images in deep learning. *PLoS ONE*, 2023, vol. 18, p. e0289499. DOI: 10.1371/journal.pone.0289499.

20. Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 1960, vol. 20 (1), pp. 37–46. DOI: 10.1177/001316446002000104.

21. Dmitrieva V.V., Magmanova Ya.A., Tsypliyak V.I., Polyakov E.V. Sistema annotirovaniya biomeditsinskikh izobrazhenii na osnove internettekhologii v onkogematologii [A system for annotating biomedical images based on internet technology in oncohematology]. *International Journal of Open Information Technologies*, 2025, vol. 13, no. 3, pp. 52–57.

22. Polyakov E.V., Dmitrieva V.V., Shuvalova E.V., Filatova N.A., Kolbatskaya O.P., Klimanov I.A. Issledovanie odnoetapnoi neuronnoi seti glubokogo obucheniya dlya analiza biomeditsinskikh izobrazhenii [Study of one-stage deep learning neural network for biomedical image analysis]. *Fundamental'nye i prikladnye problemy tekhniki i tekhnologii = Fundamental and Applied Problems of Technics and Technology*, 2025, vol. 371, no. 3, pp. 118–125.

23. Kouzehkanan Z.M., Saghari S., Tavakoli S., Rostami P., Abaszadeh M., Mirzadeh F., Shahabi E.S., Gheidishahran M., Gorgi F., Mohammadi S., Hosseini R. A large dataset of white blood cells containing cell locations and types, along with segmented nuclei and cytoplasm. *Scientific Reports*, 2022, vol. 12, p. 1123. DOI: 10.1038/s41598-021-04426-x.

24. Tsutsui S., Pang W., Wen B. WBCAtt: a white blood cell dataset annotated with detailed morphological attributes. *arXiv*. 2023. arXiv:2306.13531.

25. Carion N., Massa F., Synnaeve G., Usunier N., Kirillov A., Zagoruyko S. End-to-end object detection with transformers. *Computer Vision – ECCV 2020. Lecture Notes in Computer Science*. Vol. 12346. Cham, Springer, 2020, pp. 213–229. DOI: 10.1007/978-3-030-58452-8_13.

26. Zhu L., Su W., Lu L., Li B., Wang X., Dai J. Deformable DETR: deformable transformers for end-to-end object detection. *ICLR 2021 Conference Paper*. DOI: 10.48550/arXiv.2010.04159.

27. Holzinger A., Langs G., Denk H., Zatloukal K., Müller H. Causability and explainability of artificial intelligence in medicine. *WIREs Data Mining and Knowledge Discovery*, 2019, vol. 9 (4), p. e1312. DOI: 10.1002/widm.1312.

28. Polyakov E.V., Dmitrieva V.V. Sistema analiza tekstury biomeditsinskikh izobrazhenii [Biomedical image texture analysis system]. *Modeli, sistemy, seti v ekonomike, tekhnike, prirode i obshchestve = Models, Systems, Networks in Economics, Technology, Nature, and Society*, 2025, no. 2, pp. 85–94.

29. Mu Y., Tizhoosh H., Dehkharghanian T., Campbell C. Whole slide image representation in bone marrow cytology. *bioRxiv*. 2022. DOI: 10.1101/2022.12.06.519318.

30. Selvaraju R.R., Cogswell M., Das A., Vedantam R., Parikh D., Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 618–626. DOI: 10.1109/ICCV.2017.74.

31. Ribeiro M.T., Singh S., Guestrin C. "Why Should I Trust You?": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016, pp. 1135–1144. DOI: 10.1145/2939672.2939778.

32. Haralick R.M., Shanmugam K., Dinstein I. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 1973, vol. SMC-3 (6), pp. 610–621. DOI: 10.1109/TSMC.1973.4309314.
33. Costa L. da F., Cesar R.M. *Shape classification and analysis: theory and practice*. 2nd ed. CRC Press, 2009.
34. Ronneberger O., Fischer P., Brox T. U-Net: convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2015*. Pt. 3. Cham, Springer, 2015, pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28.
35. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.
36. Settles B. *Active learning. Synthesis Lectures on Artificial Intelligence and Machine Learning*. Springer, 2012. 114 p. DOI: 10.2200/S00429ED1V01Y201207AIM018.

Для цитирования:

Поляков Е.В. Интеллектуальная система анализа данных в онкогематологии // Системы анализа и обработки данных. – 2026. – № 1 (101). – С. 125–140. – DOI: 10.17212/2782-2001-2026-1-125-140.

For citation:

Polyakov E.V. Intellectual'naya sistema analiza dannykh v onkogematologii [Intelligent data analysis system in oncohematology]. *Sistemy analiza i obrabotki dannykh = Analysis and Data Processing Systems*, 2026, no. 1 (101), pp. 125–140. DOI: 10.17212/2782-2001-2026-1-125-140.