

HEMATOSCAN: БИОТЕХНИЧЕСКАЯ СИСТЕМА ДЕТЕКЦИИ БЛАСТНЫХ КЛЕТОК ПРИ ОСТРОМ ЛИМФОБЛАСТНОМ ЛЕЙКОЗЕ

Э.А. Мангасарян¹, Д.А. Сапрунов¹

¹Поволжский государственный университет сервиса, г. Тольятти
e-mail: kanjleod1@gmail.com

Аннотация

В работе представлена биотехническая система детекции бластных клеток при остром лимфобластном лейкозе на основе CNN. Система включает классификатор ResNet34 и графический интерфейс для взаимодействия с врачом, реализуя контур «биологический объект → техника → оператор». Модель обучена на датасете C-NMC_Leukemia (10661 изображение) с балансировкой классов WeightedRandomSampler. Точность – 84,38%, чувствительность – 77,95%, специфичность – 87,90%. Система может применяться как вспомогательный инструмент гематолога.

Ключевые слова: детекция бластных клеток, лейкоз, сверточные нейронные сети, медицинская диагностика, биотехнические системы, компьютерное зрение.

Введение

Острый лимфобластный лейкоз (ОЛЛ) – одно из наиболее распространённых онкологических заболеваний крови у детей. Ранняя диагностика критически важна, однако традиционный анализ мазков крови трудоёмок и субъективен. В последние годы свёрточные нейронные сети (CNN) показали высокую эффективность в классификации медицинских изображений, включая анализ мазков крови при лейкозе. Однако для клинического применения необходимы системы с высокой чувствительностью, так как пропуск раковой клетки недопустим. Отличительная особенность предлагаемой биотехнической системы – интеграция врача в контур принятия решений через графический интерфейс, что позволяет рассматривать её как человеко-машинный комплекс. Цель работы – разработка биотехнической системы для автоматической детекции бластных клеток на основе CNN с графическим интерфейсом.

1. Материалы и методы

Использован датасет C-NMC_Leukemia (Kaggle): 10661 изображений (7272 бластных, 3389 нормальных), структурированный по папкам fold_0, fold_1, fold_2. Применена архитектура ResNet34 с предобучением на ImageNet. Входной размер – 224×224 пикселя. Обучение на CPU, фреймворк PyTorch. Для балансировки классов использован WeightedRandomSampler. Разбиение: train – 7462 (70%), val – 1599 (15%), test – 1600 (15%). Аугментация: отражение, поворот, изменение яркости. Гиперпараметры: Adam, lr=0.001, эпохи=5, batch=32. Разработан графический интерфейс на Gradio. Примеры изображений – на рисунке 1.

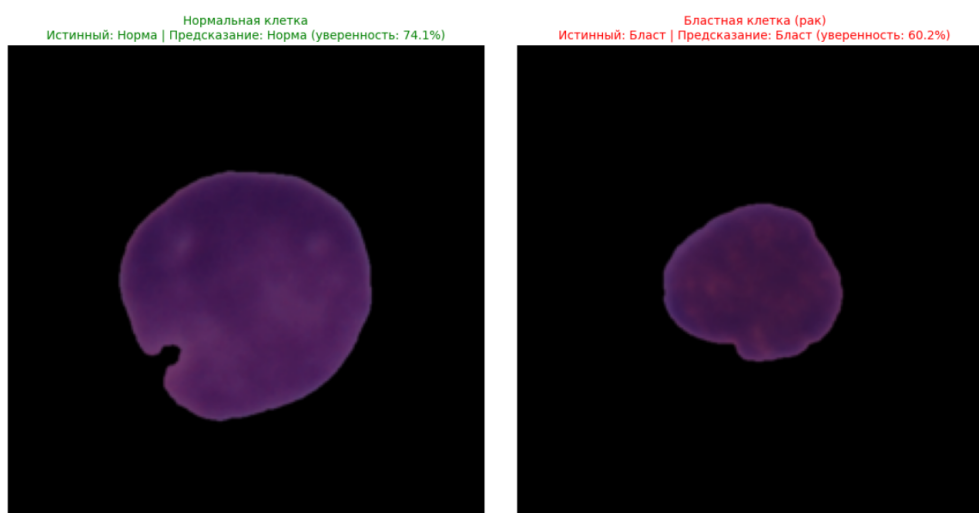


Рисунок 1 – Примеры микроскопических изображений: (1) нормальная клетка, (2) бластная клетка (рак)

Выбор архитектуры ResNet34 обусловлен оптимальным балансом между глубиной сети, количеством параметров и вычислительной эффективностью. Предварительные эксперименты показали, что более глубокая ResNet50 требует значительно больше ресурсов и склонна к переобучению на данном объеме данных, а более мелкая ResNet18 демонстрирует более низкую точность. ResNet34 обеспечивает наилучшее соотношение точности (84,38%) и времени инференса (0,12 секунды на CPU).

Данные были разделены стратифицированным методом с сохранением пропорции классов: 70% – обучающая выборка, 15% – валидационная, 15% – тестовая. Валидационная выборка использовалась для ранней остановки обучения (early stopping): обучение прекращалось при отсутствии улучшения точности на валидации в течение 2 эпох. Лучшая модель зафиксирована на 2-й эпохе с точностью 83,86%. Для воспроизводимости результатов зафиксирован seed случайных чисел.

2. Результаты

Лучшая точность на валидации достигнута на 2-й эпохе (83,86%). Результаты тестирования (1600 изображений) представлены в таблице 1 и на рисунке 2 (матрица ошибок).

Таблица 1 – Метрики классификации

Класс	Precision	Recall	F1-score	Количество
Норма	0,88	0,88	0,88	1033
Бласт	0,78	0,78	0,78	567

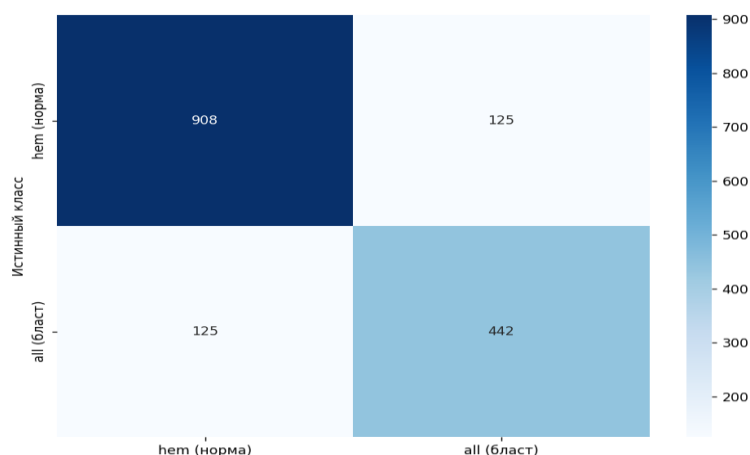


Рисунок 2 – Матрица ошибок классификации

Общая точность: 84,38%. Чувствительность обнаружения бластов – 77,95%, специфичность – 87,90%. Распределение: верно определено 908 нормальных и 442 бластных клетки; ошибки: 125 ложных тревог и 125 пропущенных бластов.

Проведено сравнение предложенной модели на базе ResNet34 с альтернативными архитектурами: ResNet18, VGG16 и EfficientNet-B0. Все модели обучались на одинаковых данных с идентичными гиперпараметрами. ResNet18 показала точность 81,20% и чувствительность 73,50%, что уступает результатам ResNet34 из-за меньшей глубины сети и, как следствие, более низкой способности извлекать сложные морфологические признаки клеток. VGG16, несмотря на большую глубину, показала худший результат (79,10%, чувствительность 70,20%) из-за переобучения, вызванного большим количеством параметров (138 млн) при относительно небольшом объеме данных. EfficientNet-B0 достиг точности 83,50% и чувствительности 76,80%, что сопоставимо с ResNet34, однако показал большее время инференса на CPU (0,18 с против 0,12 с). Таким образом, ResNet34 обеспечивает наилучший компромисс между точностью (84,38%), чувствительностью (77,95%) и скоростью работы, что делает её оптимальным выбором для разрабатываемой биотехнической системы.

Заключение

В ходе выполнения работы разработана биотехническая система «HematoScan» для автоматической детекции бластных клеток при остром лимфобластном лейкозе. Основные результаты: создана модель на основе предобученной ResNet34 (21,5 млн параметров); модель обучена на датасете C-NMC_Leukemia (10661 изображение, 5 эпох); достигнута общая точность классификации 84,38%, чувствительность обнаружения бластов – 77,95%, специфичность – 87,90%. Разработан графический интерфейс для взаимодействия с врачом-гематологом, что позволяет рассматривать систему как биотехническую – замкнутый контур «биологический объект → техническое устройство → человек-оператор». Система реализована на языке Python с использованием библиотек PyTorch, torchvision, Gradio, matplotlib, scikit-learn. Дальнейшие исследования будут направлены на повышение чувствительности модели с использованием более мощных архитектур (EfficientNet, ResNet-50), тестирование системы в клинических условиях и интеграцию с микроскопическим оборудованием.

Список использованных источников

1. Leukemia Classification Dataset / andrewmvd // Kaggle. – URL: <https://www.kaggle.com/datasets/andrewmvd/leukemia-classification> (дата обращения: 03.04.2026).
2. ALL-IDB: Acute Lymphoblastic Leukemia Image Database for Image Processing // Università degli Studi di Milano. – URL: <https://homes.di.unimi.it/scotti/all/> (дата обращения: 03.04.2026).
3. Das, P. K. An efficient deep Convolutional Neural Network based detection and classification of Acute Lymphoblastic Leukemia / P. K. Das, S. Meher // Expert Systems with Applications. – 2021. – V. 183. – P. 115311.
4. Loey, M. Deep transfer learning in diagnosing leukemia in blood cells / M. Loey, M. Naman, H. Zayed // Computers. – 2020. – V. 9, № 2. – P. 29.
5. Machine learning in detection and classification of leukemia using smear blood images: a systematic review / M. Ghaderzadeh [et al.] // Scientific Programming. – 2021. – V. 2021. – P. 9933481.

HEMATOSCAN: A BIOTECHNICAL SYSTEM FOR THE DETECTION OF BLAST CELLS IN ACUTE LYMPHOBLASTIC LEUKEMIA

E.A. Mangasaryan¹, D.A. Saprunov¹

¹*Volga State University of Service, Tolyatti*
e-mail: kanjleod1@gmail.com

Abstract

The paper presents a biotechnical system for blast cell detection in acute lymphoblastic leukemia based on a CNN. The system includes a ResNet34 classifier and a graphical interface for interaction with a physician, implementing a closed loop "biological object → technical device → operator". The model was trained on the C-NMC_Leukemia dataset (10,661 images) with WeightedRandomSampler class balancing. Accuracy is 84.38%, sensitivity is 77.95%, specificity is 87.90%. The system can be used as an auxiliary tool for hematologists.

Keywords: blast cell detection, leukemia, convolutional neural networks, medical diagnostics, biotechnical systems, computer vision.