

Информированная сегментация гистологических изображений для улучшения точности определения индекса Ki-67¹

А.В. Халин^{*,**}, С.А. Корчагин^{*,**}, Е.И. Ершов^{*,**}

** Институт проблем передачи информации имени А.А. Харкевича, Российская академия наук, Москва, Россия*

*** Московский физико-технический институт (Национальный исследовательский университет), Долгопрудный, Россия*

Поступила в редколлегию 09.07.2024 г. Принята 01.10.2024 г.

Аннотация—По данным Всемирной организации здравоохранения, за 2022 год было зарегистрировано 20 миллионов новых случаев онкологических заболеваний и почти 10 миллионов случаев смерти от раковых опухолей. Ключевым этапом в лечении рака является своевременная диагностика и корректный выбор курса лечения. Для этого в медицинской практике используются определенные маркеры, в частности, индекс Ki-67. Ручное определение этого индекса трудоёмко и времязатратно. В данной работе предлагается новый автоматический метод определения индекса, основанный на информированной сегментации – выделении клеток на оцифрованных гистопатологических снимках и их классификации. Тестирование метода на наборе данных SHIDC-B-Ki-67 демонстрирует улучшение качества сегментации клеток каждого из классов по сравнению с предыдущими решениями, в том числе улучшение F_1 -меры для класса Lymphocytes на 24%. Проводится также анализ разметки изображений набора данных SHIDC-B-Ki-67. Набор содержит не полностью размеченные изображения, на которых большое число клеток не имеет метки. Эксперименты показывают, что удаление небольшого числа таких изображений (5% от размера тестовой выборки) приводит к улучшению RMSE определения индекса на более чем 30%.

КЛЮЧЕВЫЕ СЛОВА: цифровая гистология, индекс Ki-67, полностью сверточные нейронные сети, энтропия.

DOI: 10.53921/18195822_2024_24_3_225

1. ВВЕДЕНИЕ

Антиген Ki-67 – это ядерный белок, тесно связанный с пролиферативной активностью клеток [1], то есть с их способностью к делению. Соотношение количества клеток, содержащих белок Ki-67, к общему числу клеток называется индексом Ki-67. В последнее время становится всё больше исследований о связи данного индекса со скоростью деления клеток раковых опухолей. Результаты показывают, что маркер позволяет не только предсказывать темп протекания болезни, но и прогнозировать реакцию на запланированное лечение [2, 3, 4].

Самый распространённый способ оценки индекса Ki-67 – анализ парафинового гистологического среза опухолевой ткани. В стандартной практике срезы исследуются иммуногистохимическим методом: при взаимодействии с определенным антителом клетки, содержащие белок Ki-67, окрашиваются. Далее лаборант выделяет несколько областей на микроскопическом снимке среза и вручную считает соотношение окрашенных клеток к их общему числу.

¹ Работа выполнена по теме государственного задания FFNU-2022-0028.

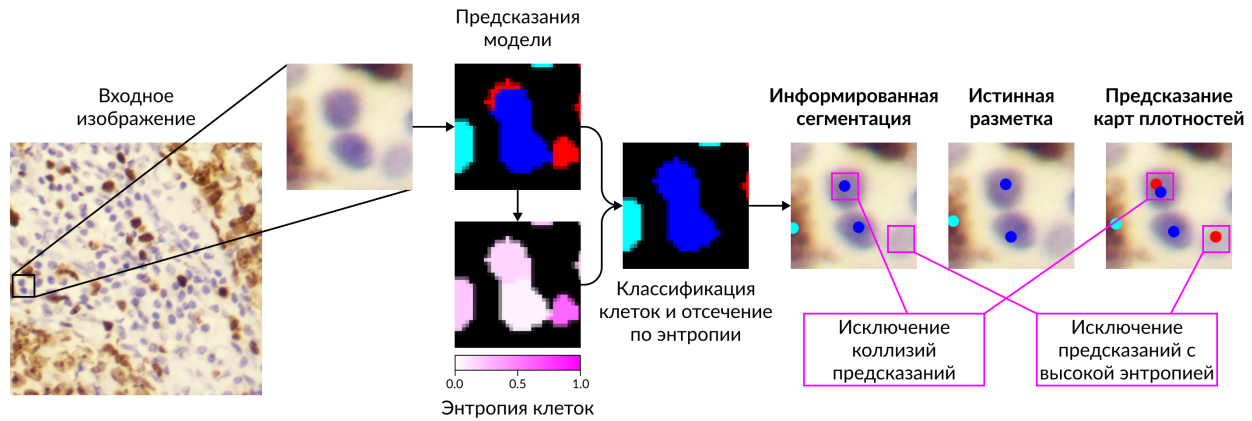


Рис. 1. Схема информированной сегментации и качественное сравнение предсказаний методов. Предлагаемый метод исключает коллизии и отвергает ответы с высокой энтропией.

Для каждого образца необходим подсчет клеток в десяти различных областях среза с высокой плотностью клеток [5]. Экспертами рекомендуется подсчет порядка 500-1000 клеток для каждой области [6], таким образом, процесс анализа является чрезвычайно времязатратным. Стоит также и проблема точности ввиду того, что анализируется не вся площадь среза.

Создание автоматической системы для определения индекса Ki-67 позволило бы решить эти проблемы. В последнее время широкое применение в медицинских задачах находят методы глубокого обучения [7]. При этом, как правило, для обучения моделей машинного обучения необходим набор размеченных экспертами данных, что затрудняет создание таких систем.

Для решения этой задачи в работе [6] был представлен набор данных SHIDC-B-Ki-67, доступный в открытом доступе. В этой же работе предлагается метод подсчёта клеток на гистологическом снимке с помощью полностью свёрточной нейронной сети (англ. fully convolutional network, FCN). Нейросеть для каждого типа клеток предсказывает карту плотностей, эти карты затем бинаризуются и обрабатываются алгоритмом водораздела.

Анализ этого метода выявил его определенный недостаток. Поскольку предсказания по различным классам формируются независимо, модель способна определять одну клетку на изображении в несколько классов (коллизия); каждое из таких предсказаний засчитывается, что приводит к неизбежному появлению ложноположительных предсказаний. Целью данной работы стало исследовать возможные подходы к устранению этого недостатка.

В данной работе предлагается новый метод определения индекса Ki-67. Он также заключается в использовании полностью свёрточной нейронной сети, но теперь для определения вероятностной карты сегментации клеток на гистологическом снимке: пикселям ставится в соответствие вероятность принадлежности каждому из четырех классов (три типа клеток и фон). Целые клетки далее классифицируются с помощью оцененных вероятностей пикселей, входящих в эту клетку, что исключает возможность возникновения коллизий. По вероятностям определяется энтропия предсказаний, и клетки с высокой энтропией исключаются для повышения точности. После этого подсчитывается их количество и определяется индекс Ki-67.

Предлагаемый метод демонстрирует лучшее значение F_1 -меры на наборе данных SHIDC-B-Ki-67 по сравнению с работой [6]. На рис. 1 представлены схема создания предсказаний предлагаемого метода, а также качественное сравнение результатов работы предлагаемого метода и метода, описанного в работе [6]. В работе, помимо этого, проводится анализ возможных недостатков используемого набора данных и их влияния на оценку качества работы методов.

2. ОБЗОР МЕТОДОВ ОПРЕДЕЛЕНИЯ ИНДЕКСА KI-67

Работы, посвященные цифровым методам определения индекса Ki-67, можно условно разделить на два типа. К первому относятся методы, определяющие индекс непосредственно по гистологическим снимкам. Так, в работе [8] снимки разбиваются на девять классов по уровню значения индекса от очень низкого до очень высокого. Для классификации предлагается объединить информацию с двух типов снимков: окрашенных на Ki-67 и окрашенных гематоксилином и эозином, что позволяет улучшить точность классификации по сравнению с методом, использующим только Ki-67-окраску. Изображения обрабатываются с помощью сверточной нейронной сети архитектуры AlexNet [9]. В статье [10] предлагается способ определения индекса Ki-67 на основании исключительно H&E-окрашенных срезов.

Методы второго типа определяют индекс Ki-67 путем подсчета клеток на изображении. Например, в работе [11] описывается процесс решения задачи детекции отдельных клеток с помощью одноэтапного множественного детектора ограничивающих прямоугольников (англ. single-shot multibox detector, SSD), обученный на синтетических данных.

В ряде работ [6, 12, 13] для этого используются нейронные сети, основанные на архитектуре U-Net [14]. При таком подходе нейросеть определяет на изображениях области, содержащие клетки, а отдельные клетки выделяются на сегментационной карте посредством пост-обработки.

Так, например, в работе [13] нейронная сеть архитектуры U-Net по гистологическому изображению предсказывает для каждого класса клеток карту плотностей. В последствии бинаризации методом Оцу [15] области, принадлежащие разным клеткам, разделяются с помощью алгоритма водораздела с преобразованием расстояния [16]. Полученные области затем подсчитываются, и определяется индекс Ki-67.

Схожая схема используется и в работе [6], однако вместо метода Оцу порог бинаризации подбирается вручную на валидационной части выборки. Также в работе проводятся эксперименты с различными архитектурами U-Net-подобных нейронных сетей, в том числе с новой архитектурой PathoNet, демонстрирующей лучшее качество сегментации клеток и определения индекса Ki-67.

В этой работе также представлен открытый набор данных SHIDC-B-Ki-67 для задачи определения индекса Ki-67. Клетки в наборе данных разделены на три класса: опухолевые клетки, содержащие Ki-67 и способные к пролиферации; опухолевые клетки, не способные к пролиферации; и опухоль-инфильтрирующие лимфоциты. Клетки последнего класса визуально слабо отличаются от клеток второго класса, но при этом не участвуют в определении индекса Ki-67. Наличие этого класса в разметке способствует способности модели нейронной сети различать эти два класса и более точно оценивать индекс.

Методы [6, 13] имеют определенную особенность. Поскольку карты плотностей, полученные от нейросети, обрабатываются независимо друг от друга, метод способен относить одну и ту же клетку к двум различным классам. В частности, это происходит между опухолевыми клетками без Ki-67 и лимфоцитами, которые, как уже было сказано, слабо отличимы визуально. Такие предсказания некорректные и неизбежно приводят к ошибкам первого рода.

В данной работе предлагается способ улучшения методов, используемых в работах [6, 13], с целью исключения коллизий предсказаний. Подробное описание предлагаемого метода представлено в разделе 3. Обучение и тестирование происходит на наборе данных SHIDC-B-Ki-67, его описание и анализ представлены в разделе 4.

3. ИНФОРМИРОВАННАЯ СЕГМЕНТАЦИЯ ДЛЯ ОПРЕДЕЛЕНИЯ ИНДЕКСА KI-67

Для каждого пациента имеется набор трехканальных изображений $\mathcal{I} = \{I^i\}_{i=1}^n \subset \mathbb{R}^{H \times W \times 3}$, где H, W – высота и ширина изображения. Изображения обучающей выборки размечены с помощью наборов координат точек $Y_c^i \subset \mathbb{Z}^2$, $c \in \{1, 2, 3\}$ на снимке. Каждый набор соответствует одному из классов, а каждая точка соответствует одной клетке данного класса. Классы Immunopositive, Immunonegative и Lymphocytes считаются классами 1, 2 и 3 соответственно.

В данной работе ставится задача создания метода, который по изображению I^i возвращает три набора координат точек $\hat{Y}_c^i \subset \mathbb{Z}^2$, $c \in \{1, 2, 3\}$, соответствующих трём классам клеток на изображении. Затем для пациента индекс Ki-67 определяется как отношение общего числа клеток первого класса к общему числу клеток первого и второго классов:

$$\hat{KI} = \frac{\sum_{i=1}^n |\hat{Y}_1^i|}{\sum_{i=1}^n |\hat{Y}_1^i| + \sum_{i=1}^n |\hat{Y}_2^i|}. \quad (1)$$

Сначала используется полностью сверточная нейронная сеть для сегментации изображения. По изображению I модель сети для каждого пикселя x предсказывает вектор $\mathbf{p}(x) = (p_0, p_1, p_2, p_3) \in [0, 1]^4$, где p_0 – степень уверенности модели в принадлежности пикселя к классу заднего фона, а p_c – степень уверенности в принадлежности к классу c , $c \in \{1, 2, 3\}$, причем $\sum_{i=0}^3 p_i = 1$.

Каждый пиксель классифицируется согласно наибольшему значению степени уверенности, и их объединение образует карту сегментации. Полученная карта сегментации затем бинаризуется: пиксели, относящиеся к любому из классов клеток, отделяются от пикселей фона. Формально

$$S(x) = \begin{cases} 1, & \max\{p_1, p_2, p_3\} > p_0, \\ 0, & \text{иначе.} \end{cases} \quad (2)$$

В последствии этого области, соответствующие различным клеткам, могут объединиться. Для разделения этих областей используется алгоритм водораздела с преобразованием расстояния [16]. Далее по карте сегментации тип каждой клетки \mathcal{C} определяется как наиболее представленный класс среди пикселей, принадлежащих к области клетки.

На данном этапе для уточнения предсказаний метода используется отсечение по энтропии $H(x) = -\sum_{i=0}^3 p_i \log p_i$. Энтропия предсказания клетки определяется как среднее значение энтропии по пикселям, относящимся к данной клетке:

$$H(\mathcal{C}) = \frac{1}{|\{x: x \in \mathcal{C}\}|} \sum_{x \in \mathcal{C}} H(x). \quad (3)$$

В случае, если энтропия клетки превышает заранее заданный порог, эта клетка исключается из предсказания. Такой подход мотивируется тем, что в методе [6] предсказания модели бинаризуются по определенному порогу. Отсечение по энтропии выступает в роли аналогичного механизма в предлагаемом методе.

Наконец, для каждой предсказанной клетки \mathcal{C} определяется её геометрический центр. Итогом работы метода является набор координат клеток, разделенный по классам. Схема метода представлена на рис. 2.

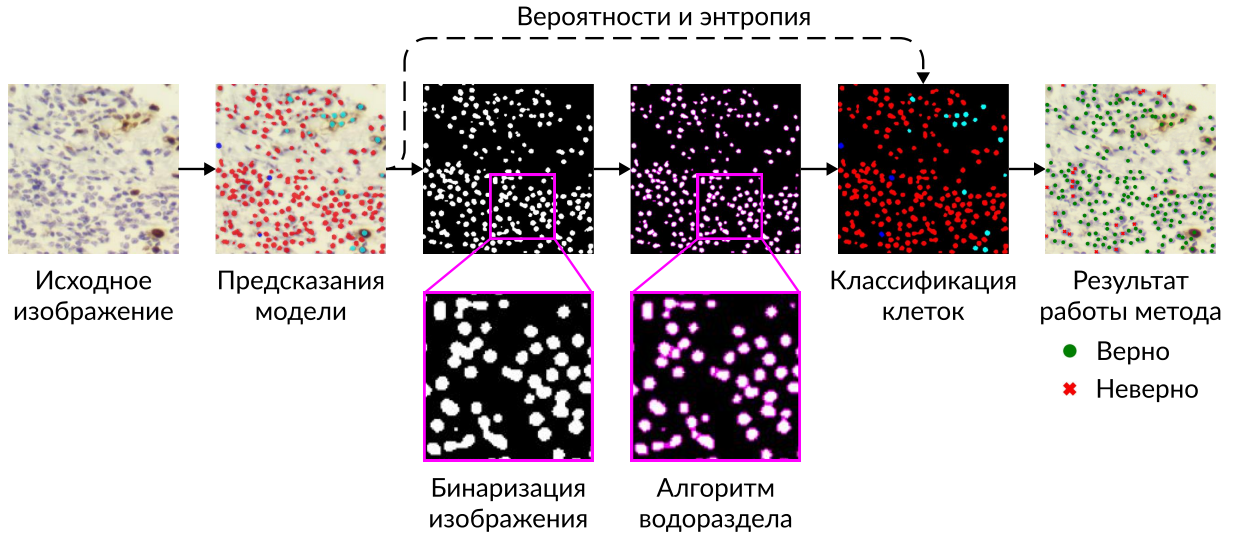


Рис. 2. Схема предлагаемого метода.

3.1. Процесс обучения нейронной сети

Разметка изображений в наборе данных SHDC-B-Ki-67 представляет собой только координаты центров клеток, что не подходит для обучения нейронной сети. Если использовать такую разметку в качестве карт сегментации, подавляющее большинство пикселей будет относиться к классу заднего фона, и графическая информация о клетках не будет сопоставлена разметке.

Для решения этой задачи каждая клетка задаётся кругом радиуса r с центром в центре клетки из разметки. Формально, каждому пикселю $x \in I$ сопоставляется вектор $\mathbf{p}^{\text{true}}(x) = (p_0^{\text{true}}, p_1^{\text{true}}, p_2^{\text{true}}, p_3^{\text{true}})$, где $p_c^{\text{true}} = 1$, если пиксель лежит на расстоянии меньше r от центра клетки c -го класса $\hat{y} \in \hat{Y}_c$, $c \in \{1, 2, 3\}$; остальные $p_j^{\text{true}} = 0$, $j \neq c$. Если клеток, удовлетворяющих этому условию, несколько, то в качестве c выбирается класс ближайшего центра клетки. Если таких клеток нет, то $p_0^{\text{true}} = 1$ и $p_1^{\text{true}} = p_2^{\text{true}} = p_3^{\text{true}} = 0$. Объединение этих векторов определяет карту сегментации $S^{\text{true}} \in \{0, 1\}^{H \times W \times 4}$.

Модель обучается предсказывать эти сегментационные карты. Пусть $\{I_i\}_{i=1}^N \subset \mathbb{R}^{H \times W \times 3}$ – обучающая выборка; $\{S_i^{\text{true}}\}_{i=1}^N \subset \{0, 1\}^{H \times W \times 4}$ – полученные описанным выше методом карты сегментации, $f_\theta: \mathbb{R}^{H \times W \times 3} \rightarrow \mathbb{R}^{H \times W \times 4}$ – нейронная сеть. Формально задача обучения нейронной сети ставится как

$$\min_{\theta} \sum_{i=1}^N \mathcal{L}(\sigma(f_\theta(I_i)), S_i^{\text{true}}), \quad (4)$$

где σ – логистическая функция, \mathcal{L} – функция потерь.

Для учета несбалансированности выборки, в качестве функции потерь используется индекс Тверского (англ. Tversky loss) [17]. Для каждого класса, включая фон, определим

$$\mathcal{L}_i(S_i^{\text{pred}}, S_i^{\text{true}}, \alpha_1, \alpha_2) = 1 - \frac{\sum p_i^{\text{true}} p_i^{\text{pred}}}{\sum p_i^{\text{true}} p_i^{\text{pred}} + \alpha_1 \sum (1 - p_i^{\text{true}}) p_i^{\text{pred}} + \alpha_2 \sum p_i^{\text{true}} (1 - p_i^{\text{pred}})}, \quad (5)$$

где p_i^{pred} – степень уверенности принадлежности пикселя к i -ому классу, предсказанная моделью, а сумма берется по всем пикселям изображения. Общая функция потерь есть взвешенная

сумма потерь по классам:

$$\mathcal{L}(S^{\text{pred}}, S^{\text{true}}, \alpha_1, \alpha_2) = \frac{\sum_{i=0}^3 w_i \mathcal{L}_i(S_i^{\text{pred}}, S_i^{\text{true}}, \alpha_1, \alpha_2)}{\sum_{i=0}^3 w_i}. \quad (6)$$

Значения α_1 , α_2 и w_i являются гиперпараметрами. Большие значения α_1 соответствуют большим штрафам за ложноположительные предсказания, а α_2 – за ложноотрицательные. При выборе $\alpha_1 + \alpha_2 = 1$ функция (5) обращается в F_γ -меру с $\alpha_1 = 1/(1 + \gamma^2)$.

После обучения определяются пороговое значение энтропии. Оно подбирается на валидационном подмножестве обучающей выборки, на котором обучение не производилось. Выбираются параметры, соответствующие наибольшему значению F_1 -меры.

Все модели обучались и тестировались с помощью фреймворка PyTorch. Для обучения использовался графический ускоритель NVIDIA A100. Каждая из моделей обучалась в течение 30 эпох. Для оптимизации использовался метод Adam [18] с шагом обучения 0,001 и параметрами $\beta_1 = 0,9$, $\beta_2 = 0,999$. Каждые 10 эпох шаг обучения уменьшался в 10 раз.

Обучающая выборка разделялась на две части в отношении 4:1; первая часть использовалась для обучения, вторая – для валидации. При генерации сегментационных карт для обучения используется $r = 7$, что приближённо соответствует радиусу клетки на изображении. В формуле (8) число пациентов равно $K = 23$. При оценке качества модели используется $d = 20$ вслед за авторами статьи [6].

3.2. Метрики оценки качества метода

Для модели оценивались как способность определять отдельные клетки на каждом изображении, так и качество предсказаний численного индекса Ki-67. Сначала необходимо определить, какие предсказания модели будут определяться как верные. Так, если координаты предсказанной клетки лежат на расстоянии не более чем d пикселей от координат истинной клетки, такое предсказание определяется как истинно положительное (англ. true positive, TP). В противном случае предсказание модели определяется как ложно положительное (англ. false positive, FP). Истинные клетки, не предсказанные моделью, определяются как ложно отрицательные предсказания (англ. false negative, FN). По количеству предсказаний каждого типа определяется F_1 -мера:

$$F_1 = \frac{2TP}{2TP + FP + FN}. \quad (7)$$

Эта метрика вычисляется для каждого класса по отдельности и для метода в целом.

Набор данных разделен на группы изображений, полученных от 23 различных пациентов. Для каждого из них по формуле (1) определяется $\hat{K}I_i$ – предсказанное методом значение индекса Ki-67 соответственно. Для оценки качества определения индекса используется корень из среднеквадратичной ошибки (англ. root mean squared error, RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{K} \sum_{i=1}^K (KI_i - \hat{K}I_i)^2}, \quad (8)$$

где KI_i – истинное значение индекса Ki-67 для соответствующего пациента, K – количество пациентов.

Нужно отметить, что наличие равного числа ошибок первого и второго рода будет приводить к верному определению индекса Ki-67, поэтому мы считаем F_1 -меру более показательной

метрикой. С другой стороны, итоговой целью в практической задаче является именно числовое значение индекса. Таким образом, низкие значения RMSE имеют большую практическую ценность, чем высокие значения F_1 -меры.

4. НАБОР ДАННЫХ SHIDC-B-KI-67

В этой статье используется набор данных SHIDC-B-Ki-67, предложенный авторами статьи [6]. Он содержит микроскопические изображения злокачественных опухолей молочной железы. Набор включает 2357 трехканальных изображений разрешения 1228×1228 , среди которых 1656 изображений обучающей выборки и 701 изображение тестовой выборки, полученных от 23 пациентов с положительным диагнозом рака молочной железы. Эти изображения были получены путём разделения на фрагменты оцифрованных срезов ткани большего разрешения. Для их получения участки ткани, покрытые парафином, окрашивались антителом SP6 и контрокрашивались гематоксилином. Срезы собирались в Ширазском медицинском университете.

Для каждого изображения экспертами была создана разметка, где каждой клетке на изображении соответствует пара чисел – координат центра клетки. Каждая клетка также относится к одному из трех классов: опухолевые клетки, содержащие Ki-67 и способные к пролиферации (Immunopositive); опухолевые клетки, не способные к пролиферации (Immunonegative); и опухоль-инфильтрирующие лимфоциты (Lymphocytes). Примеры изображений представлены на рис. 3.

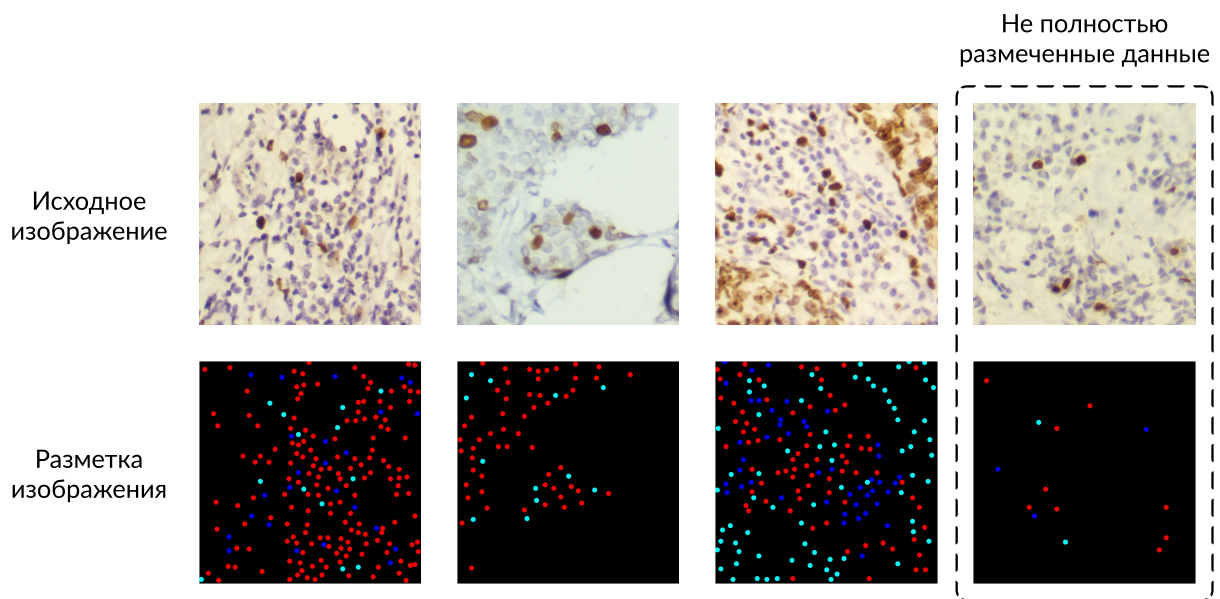


Рис. 3. Примеры изображений из набора данных SHIDC-B-KI-67 с разметкой. Классы Immunopositive, Immunonegative и Lymphocytes отмечены бирюзовым, красным и синим цветами соответственно.

Всего изображения содержат 163 тысячи размеченных клеток. Для классов Immunonegative, Immunopositive и Lymphocytes доля клеток равна 31,2%, 66,0% и 2,8% соответственно от общего числа клеток. Наблюдается выраженная несбалансированность классов.

По разметке подсчитывается общее количество клеток каждого типа, относящихся к конкретному пациенту и для каждого из них определили значение индекса Ki-67.

Среднее значение индекса Ki-67 по 23 пациентам составило 0,346 при стандартном отклонении 0,212. Наименьшее значение, представленное в наборе данных, составило 0,090, а наибольшее – 0,792. На рис. 4 представлена гистограмма значений индекса.

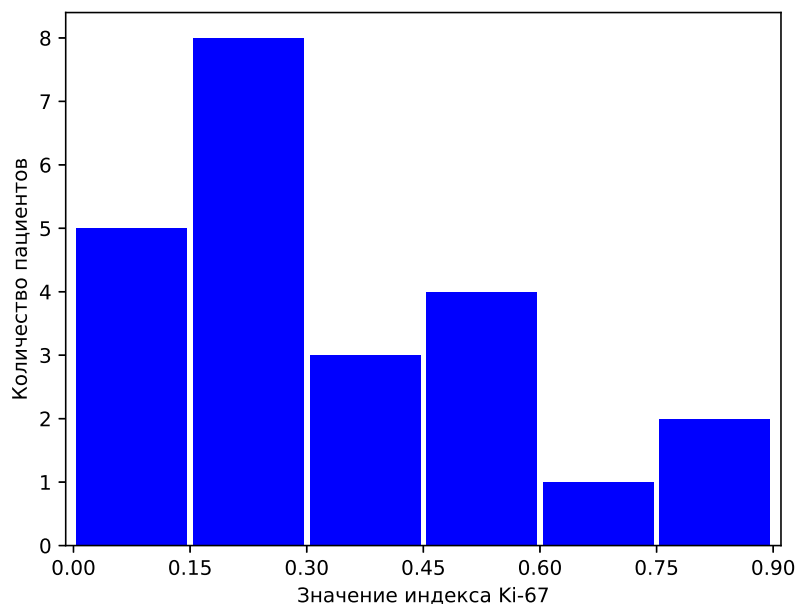


Рис. 4. Гистограмма значений индекса Ki-67 среди пациентов, представленных в наборе данных SHIDC-B-Ki-67.

При обзоре набора данных был обнаружен ряд изображений, на которых отсутствует разметка значительного числа клеток. Так, например, все клетки, имеющие бурый окрас, являются клетками, содержащими Ki-67, и должны относиться к классу Immunopositive; однако это не так (пример на рис. 3). Предполагается, что при создании набора данных гистологические изображения, полученные от пациентов, сначала размечались экспертами, а затем делились на фрагменты. Это привело к появлению областей с неразмеченными клетками.

5. ВЫЧИСЛИТЕЛЬНЫЕ ЭКСПЕРИМЕНТЫ

В данном разделе описаны эксперименты, демонстрирующие работу метода, и приведены их результаты. В первом эксперименте сравниваются различные архитектуры нейронной сети и определяем наилучшую среди рассматриваемых. Во втором эксперименте демонстрируется, что отсечение предсказаний по энтропии улучшает работу метода. В третьем эксперименте устанавливается влияние не полностью размеченных данных на оценку качества работы методов.

В первом эксперименте сравнивается качество работы метода информированной сегментации с использованием различных архитектур сверточных сетей. Для сравнения были выбраны архитектура U-Net [14], одна из наиболее распространенных архитектур FCN; а также две модификации этой архитектуры – U-Net++ [19] и DoubleU-Net [20], которые демонстрируют более высокое качество сегментации в других медицинских задачах по сравнению с U-Net.

Значения весов в формуле (6) были эмпирически выбраны равными 2, 4, 2, 1 соответственно. Каждая из архитектур обучалась с тремя парами параметров (α_1, α_2) в формуле (5): (0,3; 0,7),

(0,5;0,5) и (0,7;0,3). Такие наборы параметров соответствуют случаям, в которых точность (англ. precision) предсказаний имеет ценность ниже; одинаковую; и выше по сравнению с полнотой (англ. recall). Среди трех моделей выбиралась лучшая по F_1 -мере на валидационной части выборки.

В таблице 1 представлены результаты работы метода информированной сегментации с тремя различными архитектурами нейронных сетей. Для сравнения использовался метод предсказания карт плотностей, представленный в статье [6], использующий архитектуру PathoNet. Код, параметры и гиперпараметры модели, необходимые для воспроизведения результатов, размещены авторами статьи на веб-сервисе GitHub.

Архитектура FCN	F_1 -мера				RMSE
	Immunopositive	Immunonegative	Lymphocytes	Среднее	
UNet	0,8462	0,7700	0,3997	0,7839	0,05524
UNet++	0,8505	0,7726	0,4228	0,7876	0,05619
DoubleUNet	0,8536	0,7858	0,4556	0,7981	0,04840
PathoNet [6]	0,8526	0,7831	0,3673	0,7936	0,04735

Таблица 1. Результаты экспериментов для различных архитектур FCN на наборе данных SHIDC-B-Ki-67.

Модификация метода, основанная на архитектуре DoubleU-Net, демонстрирует наибольшие значения F_1 -меры по каждому из классов и в среднем. Метод позволяет незначительно улучшить значение F_1 -меры для первых двух классов и в среднем по сравнению с предыдущей работой. Стоит отметить существенное улучшение F_1 -меры для класса Lymphocytes – на 0,0883 (24%) по сравнению с методом PathoNet. При этом значение RMSE для нового метода несколько хуже. Таким образом, предложенный метод решает проблему коллизий и позволяет улучшить качество сегментации клеток.

Целью второго эксперимента было продемонстрировать, что исключение предсказаний с высокой энтропией повышает качество работы метода. Для этого использовалась модификация метода с архитектурой DoubleU-Net. Пороговое значение энтропии устанавливалось вручную. Далее метод оценивался на валидационной части обучающей выборки и тестовой выборке. Проверялись пороговые значения энтропии на отрезке от 0,2 до 1,0 с шагом 0,05. Зависимости F_1 -меры и RMSE от порогового значения энтропии представлены на рис. 5.

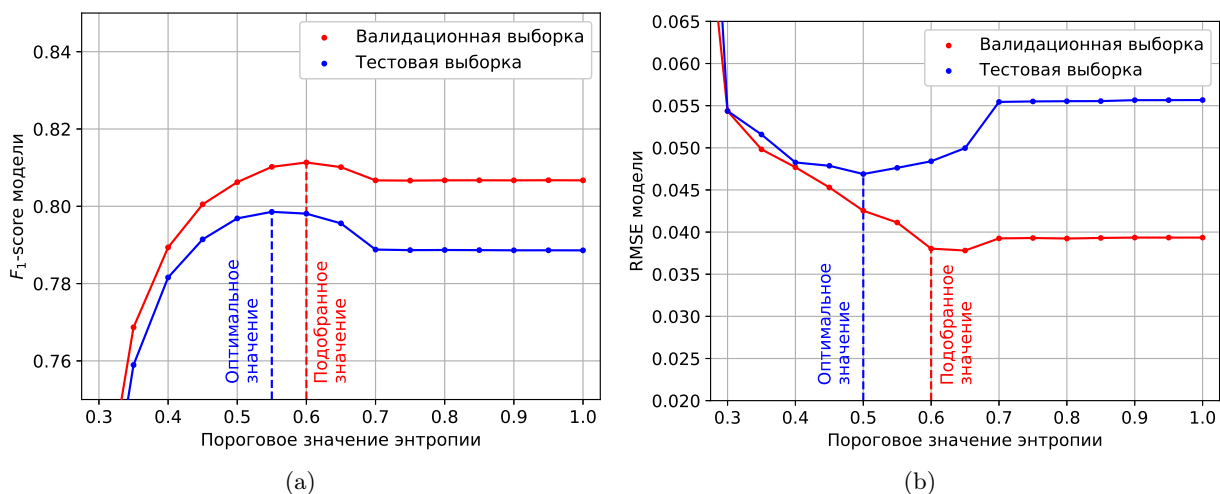


Рис. 5. График зависимости (а) F_1 -меры и (б) RMSE метода от порогового допустимого значения энтропии.

Без использования энтропии значение F_1 -меры на валидационной выборке составило 0,8067, при этом при пороговом значении энтропии, равным 0,6, метрика составила 0,8113. Для тестовой выборки в отсутствие отсеечения по энтропии F_1 -мера равна 0,7886, а при использовании выбранного порога 0,6 возрастает до 0,7981. При этом оптимальным значением порога является 0,55, при котором F_1 -мера равна 0,7986.

Для RMSE оптимальное значение порога отличается от оптимального по F_1 на 0,05. При этом подобранное для метода значение отличается от оптимального по RMSE даже на валидационной выборке. Иными словами, если бы пороговое значение выбиралось не максимизацией F_1 -меры, а минимизацией RMSE, порог был бы на 0,05 выше.

Можно сделать следующие выводы. Во-первых, отсеечение по энтропии положительно влияет на точность предсказаний. Во-вторых, подбираемое на этапе валидации значение порога может отличаться от оптимального. Наконец, оптимальные значения отличаются для разных метрик, и выбирать способ определения порога нужно в соответствии с целями работы.

Наличие не полностью размеченных изображений, описанных в разделе 4, может затруднить работу с набором данных, поскольку теоретически верные предсказания модели будут определяться как ошибочные. На этапе обучения это может затруднять подбор оптимальных параметров модели; однако этот эффект может иметь значение только в случае, если количество неразмеченных клеток сопоставимо с количеством размеченных. При тестировании модели наличие таких изображений может приводить к тому, что значения метрик занижаются и не отражают реальной способности модели к решению задачи. Третий эксперимент был проведен с целью оценить этот эффект.

Для этого из тестовой выборки были вручную удалены 5% изображений (36 изображений), выбранные как наиболее проблемные, то есть содержащие большое количество неразмеченных клеток. Далее на полученной выборке протестировались наилучшие модели из предыдущего эксперимента. Результаты представлены в таблице 2.

Архитектура FCN	F_1 -мера				RMSE
	Immunopositive	Immunonegative	Lymphocytes	Среднее	
UNet	0,8506	0,7851	0,3969	0,7951	0,04436
UNet++	0,8559	0,7924	0,4188	0,8022	0,04064
DoubleUNet	0,8583	0,8021	0,4509	0,8102	0,03347
<i>PathoNet</i> [6]	0,8578	0,7917	0,3612	0,8007	0,03739

Таблица 2. Результаты экспериментов для различных архитектур FCN на «очищенном» наборе данных.

Наиболее значимым результатом является улучшение RMSE: например, для архитектуры DoubleU-Net значение уменьшилось на 0,01456, что составляет 30,3% от изначального значения. Другим замечанием является то, что на исходном наборе данных архитектура U-Net++ проигрывала U-Net по RMSE; однако на новой тестовой выборке результаты последней хуже. Можно сделать вывод, что не полностью размеченные изображения значительно затрудняют оценку качества работы методов.

6. ЗАКЛЮЧЕНИЕ

Информированная сегментация позволяет исключить коллизии предсказаний в задаче автоматического определения индекса Ki-67 и демонстрирует лучшее качество определения клеток на наборе данных SHIDC-B-Ki-67 по сравнению с работой [6].

Использование энтропии как меры неопределенности позволяет улучшить точность метода путем исключения предсказаний с высоким значением энтропии. Оценка определенности представляет и практический интерес. Так, распространено внедрение автоматических algo-

ритмов в качестве систем предоставления второго мнения для медицинских экспертов. Тогда информация о «неуверенности» модели в её предсказаниях может указать на затруднительные места и упростить взаимодействие с системой.

Не полностью размеченные изображения, содержащиеся в наборе данных SHIDC-B-Ki-67, влияют на оценку методов. Удаление даже небольшого числа изображений из набора данных может значительно изменить наблюдаемые метрики (изменение RMSE более 30%). Более того, метод, уступающий по качеству на исходных данных, может обходить другие методы на «очищенном» наборе. Рекомендуется доработать набор данных для того, чтобы он мог стать хорошим тестом качества методов.

СПИСОК ЛИТЕРАТУРЫ

1. Scholzen T., Gerdes J. The Ki-67 protein: from the known and the unknown. *Journal of cellular physiology*, 2000, vol. 182, no. 3, pp. 311–322.
2. Yerushalmi R., Woods R., Ravdin P.M., Hayes M.M., Gelmon K.A. Ki67 in breast cancer: prognostic and predictive potential. *The lancet oncology*, 2010, vol. 11, no. 2, pp. 174–183.
3. Haroon S., Hashmi A.A., Khurshid A., Kanpurwala M.A., Mujtuba S., Malik B., Faridi N. Ki67 index in breast cancer: correlation with other prognostic markers and potential in pakistani patients. *Asian pacific journal of cancer prevention*, 2013, vol. 14, no. 7, pp. 4353–4358.
4. Melling N., Kowitz C.M., Simon R., Bokemeyer C., Terracciano L., Sauter G., Izbicki J.R., Marx A.H. High Ki67 expression is an independent good prognostic marker in colorectal cancer. *Journal of clinical pathology*, 2016, vol. 69, no. 3, pp. 209–214.
5. Jones R.L., Salter J., A'Hern R., Nerurkar A., Parton M., Reis-Filho J.S., Smith I.E., Dowsett M. Relationship between oestrogen receptor status and proliferation in predicting response and long-term outcome to neoadjuvant chemotherapy for breast cancer. *Breast cancer research and treatment*, 2010, vol. 119, pp. 315–323.
6. Negahbani F., Sabzi R., Jahromi B.P., Firouzabadi D., Movahedi F., Shirazi M.K., Majidi S., Dehghanian A. PathoNet introduced as a deep neural network backend for evaluation of Ki-67 and tumor-infiltrating lymphocytes in breast cancer. *Scientific reports*, 2021, vol. 11, no. 1, p. 8489.
7. Van der Laak J., Litjens G., Ciompi F. Deep learning in histopathology: the path to the clinic. *Nature medicine*, 2021, vol. 27, no. 5, pp. 775–784.
8. Swiderska-Chadaj Z., Gallego J., Gonzalez-Lopez L., Bueno G. Detection of Ki67 hot-spots of invasive breast cancer based on convolutional neural networks applied to mutual information of H&E and Ki67 whole slide images. *Applied Sciences*, 2020, vol. 10, no. 21, p. 7761.
9. Krizhevsky A., Sutskever I., Hinton G.E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 2012, vol. 25.
10. Liu Y., Li X., Zheng A., Zhu X., Liu S., Hu M., Luo Q., Liao H., Liu M., He Y., Chen Y. Predict Ki-67 positive cells in H&E-stained images using deep learning independently from IHC-stained images. *Frontiers in Molecular Biosciences*, 2020, vol. 7, p. 183.
11. Zhang R., Yang J., Chen C. Tumor cell identification in ki-67 images on deep learning. *Molecular & Cellular Biomechanics*, 2018, vol. 15, no. 3, p. 177.
12. Falk T., Mai D., Bensch R., Çiçek Ö., Abdulkadir A., Marrakchi Y., Böhm A., Deubner J., Jäckel Z., Seiwald K., et al. U-Net: deep learning for cell counting, detection, and morphometry. *Nature methods*, 2019, vol. 16, no. 1, pp. 67–70.
13. Geread R.S., Sivanandarajah A., Brouwer E.R., Wood G.A., Androutsos D., Faragalla H., Khademi A. Pinet—an automated proliferation index calculator framework for Ki67 breast cancer images. *Cancers*, 2020, vol. 13, no. 1, p. 11.

14. Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation. *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, proceedings, part III 18*. Munich: Springer, 2015, pp. 234–241.
15. Otsu N., et al. A threshold selection method from gray-level histograms. *Automatica*, 1975, vol. 11, no. 285–296, pp. 23–27.
16. Chen Q., Yang X., Petriu E.M. Watershed segmentation for binary images with different distance transforms. *Proceedings of the 3rd IEEE international workshop on haptic, audio and visual environments and their applications*. Ottawa: IEEE, 2004, vol. 2, pp. 111–116.
17. Salehi S.S.M., Erdogmus D., Gholipour A. Tversky loss function for image segmentation using 3d fully convolutional deep networks. *International workshop on machine learning in medical imaging*. Quebec: Springer, 2017, pp. 379–387.
18. Kingma D.P., Ba J. *Adam: A method for stochastic optimization*. Preprint. arXiv:1412.6980. 2014.
19. Zhou Z., Rahman Siddiquee M.M., Tajbakhsh N., Liang J. Unet++: A nested u-net architecture for medical image segmentation. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Proceedings 4*. Granada: Springer, 2018, pp. 3–11.
20. Jha D., Riegler M.A., Johansen D., Halvorsen P., Johansen H.D. Doubleu-net: A deep convolutional neural network for medical image segmentation. *IEEE 33rd International symposium on computer-based medical systems (CBMS)*. Rochester: IEEE, 2020, pp. 558–564.

Informed segmentation of histological images for improvement of Ki-67 index assessment

A.V. Khalin, S.A. Korchagin, E.I. Ershov

According to the World Health Organization, there were an estimated 20 million new cancer cases and almost 10 million deaths from oncological diseases in year 2022. Key steps in curing cancer include a timely diagnosis and a proper choice of treatment. In medical practice, various markers are analyzed for achieving this, Ki-67 index being one of them. Manual assessment of said index is labor-intensive and time-consuming. We propose an automatic method for index assessment based on informed cell segmentation via a fully convolutional neural network with subsequent cell classification. This method demonstrates higher segmentation quality on a SHIDC-B-Ki-67 dataset compared to previous works, with an increase of 24% of F_1 -score in Lymphocytes class. Moreover, we analyze cell markings for images presented in SHIDC-B-Ki-67 dataset. There exist images with multiple cells missing a mark. We show in our experiments that removing a small portion of such images (5% of the test set) improves RMSE of Ki-67 index assessment by more than 30%.

KEYWORDS: digital histology, Ki-67 index, fully convolutional neural networks, entropy.