

КОМПЬЮТЕРНЫЕ ПРОГРАММЫ С ОТКРЫТЫМ КОДОМ ДЛЯ АНАЛИЗА МАЗКОВ ШЕЙКИ МАТКИ НА ОНКОЦИТОЛОГИЮ

Кононова И.В.¹, Софронова С.И.¹, Кириллина М.П.¹, Мамаева С.Н.²

¹ФГБНУ «Якутский научный центр комплексных медицинских проблем», Якутск, e-mail: irinakon.07@mail.ru;

²ФГАОУ ВО «СВФУ имени М.К. Аммосова», Якутск

В стратегии ВОЗ, утвержденной 3 августа 2020 года, которая признает рак шейки матки (РШМ) глобальной проблемой, изложена срочная необходимость осуществления и расширения масштабов программ скрининга и активизации научных исследований в этом направлении для разработки экономически эффективных и инновационных мероприятий. К стандартным методам скрининга РШМ на протяжении многих лет относится цитологическое исследование шейки матки, которое снизило уровень смертности от РШМ во многих странах мира. Однако лишь некоторые лаборатории имеют в своем оснащении автоматизированные системы анализа и компьютеризированного просмотра. Это не может не сказываться на эффективности выявления патологии шейки матки. Поэтому, не только в России, но и в других странах мира, в текущее время вырос интерес к разработке прикладных компьютерных программ (КП) с открытым доступом для анализа изображений клеток. Эти КП могут также уменьшить зависимость от трудоемкого ручного микроскопического исследования. В статье описаны восемь КП с открытым исходным кодом, предназначенных для анализа изображений клеток, полученных методом световой микроскопии, и которые могут использоваться для анализа мазков шейки матки, приготовленных методами традиционной и жидкостной цитологии. В этих КП используются компьютерные методы и модели, основанные на глубоком обучении. КП апробированы и рекомендованы исследователями, которые не являются их разработчиками (авторами) и правообладателями. Несмотря на то что КП рекомендуются, в том числе пользователям без знания программирования, все-таки их использование требует знаний и умений в области компьютерных технологий, выходящих за обычные пользовательские рамки, принятые в текущее время в российской биомедицине. И это требует в настоящее время от российского биомедицинского сообщества укрепления и расширения знаний и умений в сфере компьютерных технологий.

Ключевые слова: рак шейки матки, скрининг, цитология, компьютерное зрение, автоматизация.

OPEN SOURCE COMPUTER PROGRAMS FOR ONCOCYTOLOGY CERVICAL SMEARS ANALYSIS

Kononova I.V.¹, Sofronova S.I.¹, Kirillina M.P.¹, Mamaeva S.I.²

¹Yakut Scientific Center of Complex Medical Problems, Yakutsk, e-mail: irinakon.07@mail.ru;

²M.K. Ammosov North-Eastern Federal University", Yakutsk

In the strategy dated the 3 August 2020 the WHO recognized Cervical Cancer (CC) as a global problem and highlighted the urgent need to implement and scale up screening programs and intensify research in this direction to develop cost-effective and innovative activities. For many years, the standard methods of CC screening include cytological examination of the cervix, which has reduced the CC mortality in many countries of the world. However, only a few laboratories are equipped with automated analysis and computer vision systems. This can affect the effectiveness of detecting the pathology of the cervix. Therefore, not only in Russia, but also in other countries of the world, interest in the development of open access computer applications (CAs) for the analysis of cell images has grown. These CAs could also help reduce time-consuming manual microscopic examinations. This article describes eight open-sources CAs for the analysis of cell images obtained by light microscopy and which can be used for the analysis of cervical smears prepared by conventional and liquid cytology methods. These CAs use deep learning-based computer methods and models. The CAs have been tested and recommended by researchers who are not their developers (authors) and copyright holders. Despite the fact that CAs are recommended, including without the use of programming, nevertheless, their use requires knowledge and skills in computer technology that go beyond the current scope of ordinary users in Russian biomedical environment. And that now requires the Russian medical community to strengthen and expand its knowledge and skills in computer technology.

Keywords: cervical cancer, screening, cytology, computer vision, automation.

В докладе генерального директора Всемирной организации здравоохранения (ВОЗ) по ускорению элиминации рака шейки матки (РШМ), выпущенном 30 ноября 2018 года, отмечается, что, наряду с вакцинацией против вируса папилломы человека (ВПЧ), лечением предраковых состояний, ранним выявлением, безотлагательным лечением инвазивного рака на ранних стадиях и оказанием паллиативной помощи, важной частью проверенной эффективной стратегии в отношении борьбы с РШМ на всех этапах оказания медицинской помощи является скрининг. В докладе также указывается, что, по прогнозам, несмотря на эффективность этой стратегии, без срочного расширения медицинского обслуживания смертность от РШМ продолжит значительно расти, и рост будет неравномерным – наибольший относительный прирост ежегодного числа случаев заболевания в период до 2040 гг. придется на страны с более низким уровнем дохода, что еще больше усугубит текущие значительные различия в показателях заболеваемости РШМ, когда 90% случаев смерти от него происходят в странах с низким и средним уровнем доходов [1].

Такой сценарий со значимой вероятностью может развернуться в Российской Федерации (РФ), относящейся к странам со средним уровнем доходов [2]. Действительно, в РФ уровни заболеваемости РШМ и смертности от него на протяжении многих лет стабильно составляют значительную долю в структуре общей заболеваемости злокачественными новообразованиями (ЗНО) и смертности от них. Например, в 2020 году общероссийские уровни заболеваемости РШМ и смертности от него все еще имеют высокие значения - 4-е место в структуре всех ЗНО у женщин [3].

3 августа 2020 года, во время проведения семьдесят третьей сессии, принимая к сведению доклад генерального директора, ВОЗ утвердила Глобальную стратегию по ускорению элиминации РШМ, признав его проблемой общественного здравоохранения, и установила цели и задачи на период 2020-2030 гг. В стратегии указывается, что скрининг (с вакцинацией) является эффективным мероприятием по его профилактике. Также в стратегии признается неотложная необходимость осуществления и расширения масштабов программ скрининга и активизации научных исследований для разработки экономически эффективных и инновационных мероприятий, связанных с РШМ, что позволит повысить ценовую и физическую доступность скрининга в том числе [4].

К стандартным методам скрининга РШМ на протяжении многих лет относится цитологическое исследование шейки матки, которое снизило уровень смертности от РШМ в странах с высоким уровнем доходов [5].

В России в текущее время проводится два типа цитологических исследований шейки матки: традиционная цитология (ТЦ) и жидкостная цитология (ЖЦ). ЖЦ обеспечивает более высокое качество образцов за счет большей четкости и равномерного распределения мазков.

Результаты исследований, в которых сравнивались ТЦ и ЖЦ, позволяют сделать вывод о том, что ЖЦ обладает более высокой чувствительностью [6; 7], особенно при применении автоматизированной системы анализа и компьютеризированного просмотра [8].

В России представлены несколько систем для ЖЦ, однако лишь некоторые клиничко-диагностические лаборатории располагают автоматизированным оборудованием для всех диагностических этапов. Зачастую в лабораториях отсутствуют системы компьютеризированного просмотра. Это не может не сказываться на эффективности выявления патологии шейки матки. Одним из факторов, влияющих на количество ложноотрицательных мазков, является правильная интерпретация цитологом присутствующих в мазке клеток. Лишь очень небольшая часть из них может быть злокачественной, и это требует внимательности и напряжения от цитолога. Согласно рекомендациям допустимо просматривание не более 70 мазков в день во избежание переутомления. И даже в этом случае цитологу нужно быть в постоянном тонусе, анализируя 3–4 поля зрения в секунду, не пропустив при этом патологических изменений в мазке [9]. К тому же до сих пор значительное количество российских клиничко-диагностических лабораторий не оснащены системами для ЖЦ и, соответственно, применяют ТЦ.

В связи с ограничением материальных ресурсов в здравоохранении, не только в России, но и многих других странах мира, вырос интерес к разработке прикладных компьютерных программ (КП) с открытым кодом для анализа изображений клеток (open-source software tools), которые могли бы помочь с решением проблемы материальной доступности автоматизированных системы анализа и компьютеризированного просмотра. КП для анализа изображений клеток призваны уменьшить зависимость от трудоемкого ручного микроскопического исследования биоматериала, одновременно требующего от специалиста значительного опыта в этой области биомедицинских исследований.

Конечно, для того чтобы такие КП использовались в клинической практике, необходимо пройти процедуру их регистрации как медицинского изделия, которая сопровождается определенными трудностями [10]. Однако стремительное развитие цифровых технологий во всех сферах человеческой жизни, которое мы имеем в текущее время, неизбежно приведет и к цифровой трансформации медицинских услуг, в том числе в клинической лабораторной диагностике, и, соответственно, трудности, в первую очередь административные, должны быть преодолены. Темпы цифровизации диктуют специалистам клинической лабораторной диагностики, в том числе работающим с изображениями биоматериала, быть уверенными и продвинутыми пользователями разных КП и знать основы распространенных языков программирования, чтобы поддерживать эффективность своей профессиональной деятельности. К тому же во многих клиничко-диагностических лабораториях ведется активная

исследовательская работа, которая в настоящее время невозможна без использования цифровых технологий.

Цель исследования. Учитывая наличие проблемы материальной доступности автоматизированных систем анализа и компьютеризированного просмотра для анализа мазков шейки матки на онкоцитологию, мы сделали обзор КП, актуальных в настоящий момент. КП для нашего обзора должны были соответствовать следующим критериям: иметь открытый код, уметь анализировать изображения клеток, полученных методом световой микроскопии, и обладать потенциалом для анализа мазков шейки матки, приготовленных методами ТЦ и ЖЦ. Также эти КП должны быть апробированы и рекомендованы исследователями, которые не являются их разработчиками (авторами) и правообладателями.

Материалы и методы исследования. Поиск информации о КП, которые соответствуют критериям, упомянутым выше, осуществлялся в открытом полнотекстовом архиве литературы по биомедицине и наукам о жизни PubMed Central. Архив принадлежит Национальной библиотеке медицины Национального института здоровья США (U.S. National Institutes of Health's National Library of Medicine) [11]. Учитывая скорость разработок КП, их поиск велся среди публикаций только 2021 года.

Детали поиска были следующими - (Open[All Fields] AND "source"[All Fields] AND ("software"[MeSH Terms] OR "software"[All Fields]) AND bioimage[All Fields] AND ("cytology"[Subheading] OR "cytology"[All Fields] OR "cytological techniques"[MeSH Terms] OR ("cytological"[All Fields] AND "techniques"[All Fields]) OR "cytological techniques"[All Fields] OR "cytology"[All Fields] OR "cell biology"[MeSH Terms] OR ("cell"[All Fields] AND "biology"[All Fields]) OR "cell biology"[All Fields])) AND ("open access"[filter] AND ("2021/01/01"[PubDate] : "2021/12/31"[PubDate])).

Результаты исследования и их обсуждение. В результате указанного поискового запроса платформой PubMed Central были выданы 66 публикаций, из них были выбраны 2 статьи, которые, по нашему мнению, наиболее полно соответствуют задачам нашего обзора [12; 13]. В выбранных нами статьях мы нашли упоминания о восьми КП, таких как BIAFLOWS, CDeep3M2, HistomicsML2, NucleAlzer, CellPose, Microscopy Image Browser, InstantDL.

В перечисленных КП решается широкий спектр задач, которые очень важны для компьютерного анализа изображений клеток, таких как автоматическая обработка вариабельности мазков, обнаружение артефактов, сегментация отдельных ячеек и кластеров ячеек, сегментация ядер и цитоплазмы для каждой отдельной клетки и автоматическое обнаружение аномальных изменений морфологии клеток [14].

Необходимо также отметить, что в них использованы компьютерные инструменты, имеющие принятое сейчас название «глубокое обучение» (deep-learning). Глубокое обучение обещает выполнить ранее трудноразрешимые задачи анализа изображений биологических объектов. Компьютерные методы, основанные на глубоком обучении, способствуют обучению компьютерных моделей определять соответствующие шаблоны в необработанных входных данных, в то время как в классических подходах к анализу изображения применялся и применяется уже известный алгоритм для выполнения желаемой задачи с использованием набора специально разработанных правил. Определение соответствующих шаблонов в необработанных входных данных достигается путем обучения сложных моделей, известных как глубокие нейронные сети, на наборах данных, помеченных определенным результатом. Глубокое обучение позволяет вычислительным моделям нескольких уровней обработки изучать и представлять данные с несколькими уровнями абстракции, имитируя то, как мозг воспринимает и понимает информацию из множества модулей, фиксируя сложные структуры крупномасштабных данных. Глубокое обучение заняло значительное место в компьютеризированном просмотре (компьютерном зрении), когда нейронные сети, использующие глубокое обучение, начали превосходить другие методы, использовавшиеся в популярных КИ для анализа изображений. До недавнего времени считалось очень сложной задачей позволить компьютерам распознавать изображения биологических образцов, но к настоящему времени глубокие нейронные сети превзошли даже человеческие возможности в этой области. Существует следующий факт: модель глубокого обучения уменьшила вдвое один из лучших показателей ошибок в задаче классификации изображений. Методы глубокого обучения стали де-факто стандартом для решения широкого круга задач компьютерного зрения [15].

Программа BIAFLOWS (<https://biaflows.neubias.org/#/projects>) - это интернет-платформа для воспроизводимого развертывания и публичного тестирования рабочих процессов анализа изображений с упором на микроскопические изображения биологических образцов. В базе данных платформы хранятся наборы научных данных, метаданные и рабочие процессы анализа изображений с оптимизированными параметрами. Рабочие процессы запускаются удаленно, их результаты визуализируются также удаленно - в виде наложения на исходные изображения, и могут автоматически сравниваться с общепринятыми показателями. Показатели каждого запуска рабочего процесса можно просматривать для каждого изображения или в виде общей статистики по всем наборам данных. В целом BIAFLOWS обеспечивает интеграцию и оценку разнородных рабочих процессов, изначально написанных для разных программных языков и библиотек.

Программа CDeep3M2 (<https://cdeep3m.crbs.ucsd.edu/cdeep3m>) представляет собой облачный инструмент, который концептуально разработан для широкого распознавания функций с 18 миллионами обучаемых параметров и тремя моделями, обучаемыми параллельно на одном, трех и пяти последовательных кадрах изображения. В этой КП реализованы модульная структура и конвейеры пакетной обработки, что обеспечивает простоту использования и минимизацию времени простоя в облаке. Для пользователей предусмотрена немедленная возможность развертывания машины с CDeep3M для загрузки своих обучающих изображений и меток для создания собственной обученной модели и впоследствии для сегментирования своих наборов данных.

Программа HistomicsML2 (<https://histomicsml2.readthedocs.io/en/latest/index.html#>) - это КП для быстрой и интерактивной разработки классификаторов глубокого обучения на основе набора данных изображений полного слайда. Исследователи могут использовать интерфейс этой КП на основе браузера. КП обучает и проверяет классификаторы для клеточных паттернов, также помогает пользователям маркировать наиболее ценные учебные экземпляры, создавая более точные классификаторы с меньшими затратами времени и усилий.

Программа NucleAIzer (<https://www.nucleaizer.org/>) – КП, которая может быть использована и как интернет-приложение, и как локальное приложение. В КП используются разные методы анализа локализации клеточных ядер, представляющих различные реальные сценарии. В ней осуществлена адаптация к невидимым и немаркированным данным с использованием передачи стиля изображения для создания расширенных обучающих выборок, что позволяет этой модели распознавать ядра в различных реальных исследованиях.

Программа CellPose (<http://www.cellpose.org/>) – КП для ручной маркировки и обработки автоматизированных результатов с возможностью прямой загрузки в ее репозиторий. КП сегментирует широкий спектр типов изображений клеток и не требует переобучения модели или настройки параметров. CellPose работает с набором данных, содержащим более 70.000 сегментированных объектов. В этой КП происходит периодическое переобучение модели на данных, предоставленных открытым сообществом биомедицинских исследователей.

Программа CSBDeep (<https://csbdeep.bioimagecomputing.com/>) представляет собой набор цифровых инструментов для восстановления и сегментации изображений клеток с учетом их содержимого и предназначена для использования с языками программирования Python или Fiji.

Программа Microscopy Image Browser (<http://mib.helsinki.fi/index.html>) - это программный пакет на основе Matlab для расширенной обработки изображений, сегментации и визуализации многомерных (2D-4D) наборов данных световой и электронной микроскопии. КП позволяет проводить количественный анализ морфологических характеристик. Среда

этого пакета с открытым исходным кодом обеспечивает настройку и возможность добавления исследователем новых подключаемых модулей для настройки программы для конкретных нужд исследовательского проекта.

Программа InstantDL (<https://github.com/marrlab/InstantDL>) – это модуль, нацеленный на пользователей, имеющих базовые знания о машинном обучении, знающих необходимые для модуля функции и понимающих, как разделить данные на набор для обучения и набор для тестирования. Но программа хорошо документирована и для тех, кто хочет настроить ее под свои нужды, не внося изменения в код. КП может быть использована для автоматизированного анализа с использованием компьютеризированного просмотра.

После ознакомления с общей информацией об этих КП на данном этапе нам сложно рекомендовать из них какую-либо конкретную КП для исследований, содержащих методы ТЦ и ЖЦ, потому что потребности и уровни навыков программирования каждого исследователя различны. Судя по общей информации, КП имеют достаточно понятный интерфейс, поэтому можно рекомендовать опробовать их все, чтобы увидеть, удовлетворяет ли какая-либо из них потребностям конкретного исследователя.

Также хотим отметить, что, несмотря на то что КП рекомендуются, в том числе пользователям без навыков программирования, все-таки их использование требует знаний и умений в компьютерных технологиях, выходящих за рамки, принятые в текущее время за обычные пользовательские в российском биомедицинском сообществе.

Опираясь на наши знания, опыт и навыки, мы планируем провести исследование, в котором сравним характеристики этих КП для наших исследований мазков шейки матки на онкоцитологию.

Изучение характеристик КП, пусть даже ограниченное узким поиском, свидетельствует о присутствии на мировом рынке сильных инициатив в области разработки компьютерного зрения для анализа изображений в биомедицинских исследованиях. Уже известно, что не только медицинские технологические компании – от малых до крупных, но и титаны современных технологий, такие как Google (Google Brain), Microsoft и IBM, реализуют технологические проекты в этом направлении [15].

Заключение. В нашей статье мы предоставили информацию о КП с открытым исходным кодом, предназначенных для анализа изображений клеток, полученных методом световой микроскопии. По нашему мнению, они могут использоваться для анализа мазков шейки матки, приготовленных методами ТЦ и ЖЦ. Эти КП апробированы и рекомендованы исследователями, которые не являются их разработчиками (авторами) и правообладателями. Изучение характеристик этих КП позволяет нам рассматривать их использование для наших

будущих исследований шейки матки на онкоцитологию, на настоящем этапе - пока только для научных работ.

Учитывая скорость внедрения разнообразных компьютерных систем в повседневные рабочие процессы клинической практики, в кардиологическую как пример, с большой долей уверенности можно утверждать, что в скором времени автоматизированный анализ с использованием компьютеризированного просмотра цитологических препаратов шейки матки будет также рутинным процессом – а это требует уже в настоящее время от российского биомедицинского сообщества укрепления и расширения знаний и умений в области компьютерных технологий.

Список литературы

1. Доклад Генерального директора ВОЗ от 30 ноября 2018 г. EB144/28 «Ускорение элиминации рака шейки матки». [Электронный ресурс]. URL: https://apps.who.int/gb/ebwha/pdf_files/EB144/B144_28-ru.pdf (дата обращения: 21.08.2021).
2. The World Bank. [Электронный ресурс]. URL: <https://www.worldbank.org/> (дата обращения: 04.07.2021).
3. World Health Organization. International Agency for Research on Cancer. [Электронный ресурс]. URL: <https://gco.iarc.fr/today/home> (дата обращения: 21.08.2021).
4. Документ ВОЗ от 3 августа 2020 WHA73.2 «Глобальная стратегия по ускорению элиминации рака шейки матки как проблемы общественного здравоохранения и установленные в ней цели и задачи на период 2020-2030 гг.» [Электронный ресурс]. URL: https://apps.who.int/gb/ebwha/pdf_files/WHA73/A73_R2-ru.pdf (дата обращения: 21.08.2021).
5. Arbyn M., Weiderpass E., Bruni L., de Sanjosé S., Saraiya M., Ferlay J., Bray F. Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. *The Lancet Global Health*. 2020. vol. 8. E.191–203. DOI: 10.1016/S2214-109X(19)30482-6.
6. Борзенкова А.В., Байтухиева Д.Ш., Писарева Е.В., Дорошенко Е.А. Сравнение результатов цитологического скрининга патологии шейки матки традиционным и жидкостным методами // *Universum: медицина и фармакология*. 2020. №4-5 (67). URL: <http://7universum.com/ru/med/archive/item/9331> (дата обращения: 21.08.2021).
7. de Oliveira A.C.A., Domingues M.F., Neufeld P.M., Fleury M., Nogueira Neto J.F. Comparison between conventional cytology and liquid-based cytology in the Tertiary Brazilian Navy Hospital in Rio de Janeiro. *Acta Cytologica*. 2020. vol. 64. P. 539-546. DOI: 10.1159/000508018.

8. Минкина Г.Н. Цитологический скрининг рака шейки матки: от традиционного Пап-теста к компьютерным технологиям // *Акушерство, Гинекология и Репродукция*. 2017. №11(1). С. 56-63. DOI: 10.17749/2313-7347.2017.11.1.056-063.
9. Савостикова М.В., Короленкова Л.И., Федосеева Е.С., Пименова В.В. Цитологический цервикальный скрининг в Ростовской области с применением жидкостной технологии BD SUREPATH // *Новости клинической цитологии России*. 2018. №22(1-2). С. 37-40.
10. Санкт-Петербургское государственное бюджетное учреждение здравоохранения «Медицинский информационно-аналитический центр» (СПб ГБУЗ МИАЦ) [Электронный ресурс]. URL: <https://spbmiac.ru/o-procedure-registracii-medicinskogo-programmnogo-obespecheniya-po-kak-medicinskogo-izdeliya/> (дата обращения: 21.08.2021).
11. PubMed Central [Электронный ресурс]. URL: <https://www.ncbi.nlm.nih.gov/labs/pmc/> (дата обращения: 21.08.2021).
12. Paul-Gilloteaux P., Tosi S., Hériché J-K., Gaignard A., Ménager H., Marée R., Baecker V., Klemm A., Kalaš M., Zhang C., Miura K., Colombelli J. Bioimage analysis workflows: community resources to navigate through a complex ecosystem. *F1000Research*. vol. 320. no. 10. DOI:10.12688/f1000research.52569.1.
13. Lucas A.M., Ryder P.V., Li B., Cimini B.A., Eliceiri K.W., Carpenter A.E. Open-source deep-learning software for bioimage segmentation. *Molecular Biology of the Cell*. 2021. vol. 32. no.9. P.823-829. DOI: 10.1091/mbc.e20-10-0660.
14. Conceição T., Braga C., Rosado L., Vasconcelos M. A Review of Computational Methods for Cervical Cells Segmentation and Abnormality Classification. *International journal of molecular sciences*. 2019. vol. 20(20). e.5114. DOI: 10.3390/ijms20205114.
15. Lundervold A.S., Lundervold A. An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift für Medizinische Physik*. 2019. vol. 29. issue 2. P. 102-127. DOI: 10.1016/j.zemedi.2018.11.002.