

Государственное
бюджетное учреждение
здравоохранения
«Санкт-Петербургский
клинический научно-
практический центр
специализированных видов
медицинской помощи

(онкологический)»
(Санкт-Петербург, Россия)

Частное образовательное
учреждение высшего
образования «Санкт-
Петербургский медико-
социальный институт»
(Санкт-Петербург, Россия)

ВОЗМОЖНОСТИ И ПЕРСПЕКТИВЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ПАТОМОРФОЛОГИЧЕСКОЙ ДИАГНОСТИКЕ РАКА

К.В. Шелехова

OPPORTUNITIES AND PROSPECTS OF ARTIFICIAL INTELLIGENCE IN THE PATHOLOGICAL IDENTIFICATION OF CANCER

К.В. Шелехова

*доктор медицинских наук, заведующая кафедрой патологической анатомии
факультета дополнительного профессионального образования
ЧОУ ВПО «СПБМСИ», заведующая патологоанатомическим отделением
ГБУЗ «СПбКНПЦСВМП(о)»,
197758, Санкт-Петербург, пос. Песочный, Ленинградская ул., д. 68А.*

K.V. Shelekhova

*MD, PhD, SciD, Head of Pathology Department of Clinical research and practical Center for
Specialized Oncological Care, Saint-Petersburg; Professor of Pathology Department
of Saint-Petersburg Medico-Social Institute,
197758, Saint-Petersburg, Pesochny-2, Leningradskaya str. 68a Lit A.*

Цифровизация, искусственный интеллект и машинное обучение все больше и активнее внедряются в медицину и, в частности, патологическую анатомию. В обзоре представлены определения и современные концепции этого направления, возможности и ограничения, преимущества и недостатки его использования. Обсуждаются примеры из репрезентативных публикаций.

Ключевые слова: *искусственный интеллект, машинное обучение, патология.*

Digitalization, artificial intelligence and machine learning are increasingly being introduced into medicine and pathology in particular. The review presents the definitions and modern concepts of this field, the possibilities and limitations, the advantages and disadvantages of their use. Examples from representative publications are discussed.

Key words: *artificial intelligence, machine learning, pathology.*

В современной онкологии патологоанатомическая диагностика играет ключевую роль. Ведение многочисленных онкологических нозологий в соответствии с рекомендациями требует не только морфологического подтверждения диагноза, но и молекулярно-патологической идентификации прогностических и предиктивных биомаркеров, количество которых постоянно увеличивается. В повседневной работе врач-патологоанатом объединяет и анализирует большие объемы информации, поступающие из различных источников – это и клиническая информация, и данные изображений гистологического, гистохимического, иммуногистохимического окрашивания, и данные молекулярного анализа. Прогресс IT-технологий постепенно внедряется и в патологоанатомическую дисциплину. Цифровизация патологической анатомии планомерно происходит в течение последних десятилетий, хотя и не нашла еще столь широкого применения, как в других дисциплинах.

Достаточно быстрое развитие сканирования микроскопических препаратов позволило оцифровать типично аналоговую микроскопическую информацию, что стимулировало дальнейшие разработки в области ком-

пьютерной обработки данных изображений, получившие название машинного или компьютерного зрения. Основная идея, заложенная в основу этих сложных алгоритмов, заключается в математическом воспроизведении биологических нейронных сетей. Под искусственной нейронной сетью понимают компьютерную модель, имитирующую биологические нейронные сети. Она состоит из искусственных нейронов, которые можно расположить слоями/плоскостями друг за другом. Они часто используются в приложениях глубокого обучения. Как правило, эти модели обучаются с использованием большого количества изображений хорошо охарактеризованных наборов данных, например, таких как контрольный набор данных ImageNet [1, 2].

Искусственный интеллект описывает способность компьютерной системы воспроизводить интеллектуальное/человеческое поведение. Для этого используются методы машинного обучения. Машинное обучение включает в себя компьютерные алгоритмы, которые обладают способностью улучшаться по мере приобретения опыта и генерировать искусственные знания. Это позволяет решать проблемы без необходимости явно указывать (например, программировать) каждый отдельный шаг решения. Глубокое обучение – это область машинного обучения, которая, в частности, связана с использованием искусственных нейронных сетей [1].

Подходы с использованием искусственного интеллекта все чаще тестируются в медицинских учреждениях, особенно в направлениях, связанных с визуализацией. В частности, ряд компаний предлагает продукты для решения конкретных рутинных диагностических задач в радиологии, однако ни один из них пока не рекомендован к широкому использованию в какой-либо стране. В патологии этот процесс идет очень медленно, что, вероятно, связано с высокой стоимостью оборудования (в первую очередь, сканеров), необходимостью крупных инвестиций и с вопросами безопасности. Тем не менее, в ряде ведущих институтов мира цифровая интерпретация частично или преимущественно внедрена, а некоторые компании предлагают сертифицированные патологоанатомические продукты по определенным направлениям. Так, в частности, по результатам ретроспективного анализа стало возможным дифференцировать рак и нормальную паренхиму молочной железы с точностью 98,8% в нескольких контрольных наборах данных, используя извлечение признаков и классификатор на основе модели машины опорных векторов (SVM) [3]. Метод опорных векторов описывает математический метод распознавания образов и обычно приводится в качестве примера классического машинного обучения. Классификация объектов достигается путем отнесения объектов к классам таким образом, что область между отдельными классами становится как можно больше.

Аналогичные исследования проводились для рака предстательной железы и плоскоклеточного рака полости рта [4, 5]. Другое ретроспективное исследование было направлено на выявление метастазов в лимфатических узлах больных раком молочной железы [4, 5]. Авторы сравнивали различные модели, представленные участниками ежегодного конкурса по программированию в области биомедицинской визуализации. Наилучшие результаты показал алгоритм – сверточная сеть на основе технологии Google – с частотой ложноположительных ответов 1,25. Сверточная сеть – это подтип искусственных нейронных сетей, который уменьшает ввод предположительно релевантных функций, используя определенные матрицы складывания. Сверточные сети в основном используются в приложениях автоматического распознавания изображений. Кроме того, авторы сравнивали результаты лучших моделей с заключениями врачей-патологоанатомов. При этом опытные патологи при отсутствии временных ограничений достигли наивысших значений точности по сравнению с алгоритмами. Однако эти показатели снижались, если врач находился под временным давлением. В другом исследовании, посвященном той же теме, точность диагностики оценивалась как с поддержкой искусственного интеллекта, так и без нее. В случаях с трудно идентифицируемыми микрометастазами точность повышалась с 83,3% (только врач) до 91,2% (при поддержке машины). По результатам этого исследования авторы сообщают о более высокой точности при сочетании человека и машины по сравнению с таковой только человека или только машины [6].

Стратегии, существующие на сегодняшний день, в основном основаны на методе обучения с учителем. Это означает, что каждое изображение, представленное сети во время обучения, сначала должно быть аннотировано патологом экспертного уровня. Поскольку для эффективного обучения требуются тысячи изображений, это приводит к значительным затратам труда и времени. В то же время патологи-эксперты – единичные специалисты, а получение подобного опыта из других источников маловероятно.

Подходы (алгоритмы) с применением искусственного интеллекта и машинного обучения видятся особенно востребованными для решения повторяющихся задач с ограниченной сложностью. В рутинной патологоанатомической диагностике это касается, прежде всего, анализа образцов тканей при скрининговых программах (к примеру, колоректального рака) или анализа случаев с множественными повторяющимися образцами для скрининга рака предстательной железы. В таких ситуациях перенос трудоемкого и монотонного процесса анализа препаратов на алгоритмы искусственного интеллекта позволил бы патологам посвящать больше времени

другим действиям, более сложным в творческом и аналитическом отношении.

Другим направлением, где применение искусственного интеллекта видится перспективным и многообещающим, являются сферы прогноза биологического поведения опухоли и предсказание возможного ответа на терапию. Для определения прогностических и предиктивных биомаркеров используют дополнительные иммуногистохимические и/или молекулярно-генетические тесты. И здесь применение искусственного интеллекта в сочетании с цифровым анализом привело бы к смене парадигмы, поскольку алгоритмы искусственного интеллекта способны различать и количественно оценивать структуры ниже порогового уровня человеческого глаза. Тем самым, это перспективно в отношении создания новых классов морфологических биомаркеров с прогностической и предиктивной достоверностью. В ретроспективном исследовании Beck A.H с соавт. прогноз для пациентов с раком молочной железы на основе срезов, окрашенных гематоксилином и эозином, был успешно оценен с использованием выделения признаков машинного обучения [7]. Свойства изображения извлекались автоматически, с определением тысяч характеристик на клеточном и тканевом уровнях. Логарифмический ранговый тест обнаружил очень значимую связь между предсказанием модели и выживаемостью пациентов ($p \leq 0,001$), независимо от других клинических или патологических факторов риска. У пациентов, отнесенных моделью к группе высокого риска, 5-летняя выживаемость составила 68,7%. Напротив, 84,5% пациентов из группы низкого риска были живы через пять лет. Кроме того, с помощью статистического анализа были определены морфологические критерии, вносящие наибольший вклад в соответствующую классификацию. Аналогичный метод, основанный на глубоком обучении, был описан для колоректального рака [8] и других заболеваний. Сначала для обработки изображений использовалась глубокая сверточная сеть. Затем ее выход служил входом для второй сети, так называемой рекуррентной нейронной сети с долговременной функцией кратковременной памяти. В сетях такого типа взвешенные связи существуют не только в направлении выходного слоя, но и в предыдущих слоях/уровнях. Это может создать своего рода память, которая полезна, например, при распознавании последовательностей. В ретроспективных исследованиях, подтверждающих концепцию, этот метод смог обнаружить микросателлитную нестабильность в желудочно-кишечных опухолях и идентифицировать молекулярный подтип рака мочевого пузыря [9, 10].

Концепция сопряжения генотипа и фенотипа предполагает, что в опухолевой ткани, например, при мутации ключевого гена, также обнаруживаются изменения морфологической картины. Помимо способности прогнозировать молекулярные или клинические

параметры на основе гистоморфологических данных, приложения искусственного интеллекта будут играть все более важную роль в анализе молекулярно-патологических особенностей. Одним из примеров может являться классификация профилей метилирования ДНК плоскоклеточного рака легкого с использованием глубокого обучения, для дифференциальной диагностики между метастазами карциномы головы/шеи и первичного рака легкого [11]. Аналогичным образом методы искусственного интеллекта играют все более важную роль в классификации молекулярных опухолей на основе мультиомики [12]. Несмотря на то, что непосредственное клиническое применение еще не установлено, уже появляются доказательства будущего дополнительного значения сложных молекулярных и морфологических профилей новообразований [13].

Многообещающим подходом для получения более точного прогноза опухолевого процесса и ответа его на терапию может быть использование в качестве вводных для искусственного интеллекта комбинации различных модальностей данных, объединяющих гистологические, иммуногистохимические и молекулярные параметры. Примером этому может служить ретроспективное исследование по глиомам Mobaradpour с соавт., в котором авторы продемонстрировали достоверно лучшую прогностическую вероятность комбинации анализа изображений и данных о мутациях по сравнению с использованием только данных изображений или только информации о мутациях [14]. В персонализированной медицине будущего мультимодальный анализ биомаркеров, интегрирующий морфологические, рентгенографические, лабораторные, геномные, протеомные и клинические параметры, может представлять большую проблему, преодоление которой возможно только с помощью подходов искусственного интеллекта. Однако для этого потребуются значительные разработки в области IT-технологий и клинической практики, а также высочайшее качество вводных данных.

Хотя искусственный интеллект и машинное обучение могут произвести революцию в патологии, имеется ряд серьезных проблем с их трансляционной реализацией. Существует сильная положительная корреляция между точностью алгоритма искусственного интеллекта и количеством используемых данных. В исследовании Campanella et al. ошибка проверки уменьшилась примерно в 10 раз, когда было проанализировано в 100 раз больше случаев [15]. Однако лишь часть гистологических образцов доступна в цифровом формате, позволяющем проводить компьютерный анализ. Кроме того, цифровизация всегда требует значительных первоначальных инвестиций. Несмотря на то, что доля оцифрованной информации значительно возрастет в среднесрочной перспективе, то есть самое позднее в ближайшее десятилетие, детальный обзор и точное описание этих данных экспертами-патологами пока отсутствует. Это ока-

зывает серьезное влияние на качество данных – еще один фактор, существенно влияющий на точность модели искусственного интеллекта (проблема «мусор на входе – мусор на выходе») [1].

В настоящее время самым большим препятствием для использования методов на основе искусственного интеллекта является почти полное отсутствие проспективных, рандомизированных, многоцентровых исследований, оценивающих с одной стороны, пользу для патологоанатомов, а с другой стороны – для пациентов. Такие исследования срочно необходимы для опреде-

ления решений искусственного интеллекта, которые действительно улучшают результаты, связанные с пациентами. Помимо данных изображений, для такого рода исследований необходимы высококачественные клинические данные. В первую очередь это относится к потенциальному прогностическому применению методов на основе обсуждаемого метода. Если эти проблемы будут решены, использование искусственного интеллекта и цифровой патологии может изменить специальность и помочь патологам выполнять свою работу быстрее и с большей точностью.

Список литературы

1. Försch S., Klauschen F., Hufnagl P., Roth W. Artificial intelligence in pathology // *Dtsch Arztebl Int.* – 2021. – № 118. – P. 199–204.
2. Mukhopadhyay S., Feldman M.D., Abels E., et al. Whole slide imaging versus microscopy for primary diagnosis in surgical pathology: a multicenter blinded randomized noninferiority study of 1992 cases (pivotal study) // *Am J Surg Pathol.* – 2018. – № 42. – P. 39–52.
3. Osareh A., Shadgar B. A computer aided diagnosis system for breast cancer // *IJCSI.* – 2011. – № 8. – P. 233–40.
4. Lee G., Sparks R., Ali S., et al. Co-occurring gland angularity in localized subgraphs: predicting biochemical recurrence in intermediate-risk prostate cancer patients // *PLoS One.* – 2014. – № 9. – e97954.
5. Lu C., Lewis J.S. Jr., Dupont W.D., Plummer W.D. Jr., Janowczyk A., Madabhushi A. An oral cavity squamous cell carcinoma quantitative histomorphometric-based image classifier of nuclear morphology can risk stratify patients for disease-specific survival // *Mod Pathol.* – 2017. – № 30. – P. 1655–65.
6. Steiner D.F., Macdonald R., Liu Y., et al. Impact of deep learning assistance on the histopathologic review of lymph nodes for metastatic breast cancer // *Am J Surg Pathol.* – 2018. – № 42. – P. 1636–46.
7. Beck A.H., Sangoi A.R., Leung S., et al. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival // *Sci Transl Med.* – 2011. – № 3. – P. 108ra113.
8. Bychkov D., Linder N., Turkki R., et al. Deep learning based tissue analysis predicts outcome in colorectal cancer // *Sci Rep.* – 2018. – № 8. – P. 3395.
9. Kather J.N., Pearson A.T., Halama N., et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer // *Nat Med.* – 2019. – № 25. – P. 1054–6.
10. Woerl A.C., Eckstein M., Geiger J., et al. Deep learning predicts molecular subtype of muscle-invasive bladder cancer from conventional histopathological slides // *Eur Urol.* – 2020. – № 78. – P. 256–64.
11. Jurmeister P., Bockmayr M., Seegerer P., et al. Machine learning analysis of DNA methylation profiles distinguishes primary lung squamous cell carcinomas from head and neck metastases // *Sci Transl Med.* – 2019. – № 11. – P. eaaw8513.
12. Hoadley K.A., Yau C., Hinoue T., et al. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer // *Cell.* – 2018. – № 173. – P. 291–304.
13. Hoberger M., von Laffert M., Heim D., Klauschen F. Histomorphological and molecular profiling: friends not foes! Morpho-molecular analysis reveals agreement between histological and molecular profiling // *Histopathology.* – 2019. – № 75. – P. 694–703.
14. Mobadersany P., Yousefi S., Amgad M., et al. Predicting cancer outcomes from histology and genomics using convolutional networks // *Proc Natl Acad Sci USA.* – 2018. – № 115. – E2970-9.
15. Campanella G., Hanna M.G., Geneslaw L., et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images // *Nat Med.* – 2019. – № 25. – P. 1301–9.

References

1. Försch S., Klauschen F., Hufnagl P., Roth W. Artificial intelligence in pathology. *Dtsch Arztebl Int.* 2021; 118: P. 199-204.
2. Mukhopadhyay S., Feldman M.D., Abels E., et al. Whole slide imaging versus microscopy for primary diagnosis in surgical pathology: a multicenter blinded randomized noninferiority study of 1992 cases (pivotal study). *Am J Surg Pathol.* 2018; 42: 39-52.
3. Osareh A., Shadgar B. A computer aided diagnosis system for breast cancer. *IJCSI.* 2011; 8: 233-40.
4. Lee G., Sparks R., Ali S., et al. Co-occurring gland angularity in localized subgraphs: predicting biochemical recurrence in intermediate-risk prostate cancer patients. *PLoS One.* 2014; 9: e97954.

5. Lu C., Lewis J.S. Jr., Dupont W.D., Plummer W.D. Jr., Janowczyk A., Madabhushi A. An oral cavity squamous cell carcinoma quantitative histomorphometric-based image classifier of nuclear morphology can risk stratify patients for disease-specific survival. *Mod Pathol.* 2017; 30: 1655-65.
6. Steiner D.F., Macdonald R., Liu Y., et al. Impact of deep learning assistance on the histopathologic review of lymph nodes for metastatic breast cancer. *Am J Surg Pathol.* 2018; 42: 1636-46.
7. Beck A.H., Sangoi A.R., Leung S., et al. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Sci Transl Med.* 2011; 3: 108ra113.
8. Bychkov D., Linder N., Turkki R., et al. Deep learning based tissue analysis predicts outcome in colorectal cancer. *Sci Rep.* 2018; 8: 3395.
9. Kather J.N., Pearson A.T., Halama N., et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat Med.* 2019; 25: 1054-6.
10. Woerl A.C., Eckstein M., Geiger J., et al. Deep learning predicts molecular subtype of muscle-invasive bladder cancer from conventional histopathological slides. *Eur Urol.* 2020; 78: 256-64.
11. Jurmeister P., Bockmayr M., Seegerer P., et al. Machine learning analysis of DNA methylation profiles distinguishes primary lung squamous cell carcinomas from head and neck metastases. *Sci Transl Med.* 2019; 11: eaaw8513.
12. Hoadley K.A., Yau C., Hinoue T., et al. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell.* 2018; 173: 291-304.
13. Hoberger M., von Laffert M., Heim D., Klauschen F. Histomorphological and molecular profiling: friends not foes! Morpho-molecular analysis reveals agreement between histological and molecular profiling. *Histopathology.* 2019; 75: 694-703.
14. Mobadersany P., Yousefi S., Amgad M., et al. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc Natl Acad Sci USA.* 2018; 115: E2970-9.
15. Campanella G., Hanna M.G., Geneslaw L., et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med.* 2019; 25: 1301-9.