ОБРАЗОВАТЕЛЬНО-ИССЛЕДОВАТЕЛЬСКАЯ ИНТЕГРИРОВАННАЯ ПЛАФОРМА ПОДДЕРЖКИ ПРИНЯТИЯ ВРАЧЕБНЫХ РЕШЕНИЙ НА ОСНОВЕ

ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ГИСТОЛОГИЧЕСКИХ ДАННЫХ. СОЗДАНИЕ И ВНЕДРЕНИЕ НЕЙРОСЕТЕВОЙ МОДЕЛИ ДЛЯ ОЦЕНКИ МОРФОЛОГИЧЕСКИХ

ИЗМЕНЕНИЙ ЛЕГКИХ ПРИ КОРОНАВИРУСНОЙ ИНФЕКЦИИ COVID-19

Хонов Владислав Русланович, Кукареко Александр Петрович

Лечебный факультет, кафедра гистологии, цитологии и эмбриологии, кафедра патологической анатомии, Белорусский государственный медицинский университет, Минск, Беларусь

# Аннотация

Создан алгоритм оценки морфологических изменений легких при инфекции COVID-19 с помощью подходов компьютерного анализа, который позволяет внедрить элементы цифровой патологии в научно-исследовательскую работу, а также образовательный процесс.

**Ключевые слова:** машинное обучение; цифровая патология; COVID-19.

# Введение

Последние достижения в области компьютерного зрения, появление возможностей быстрой оцифровки гистологических препаратов повысили интерес к приложениям искусственного интеллекта (ИИ) для патологии. Современные алгоритмы способны, к примеру, обнаруживать рак, опухоли, помогать в постановке диагноза пациенту [1]. Многие алгоритмы полагаются только на полнослайдовое изображение (WSI) и метаданные, некоторые также требуют аннотаций клеток, тканей и других объектов на WSI. Эта дополнительная информация помогает связать гистологические изобра-жения с лежащей в их основе биологией патологических изменений и может улуч-шить интерпретируемость и обобщаемость полученных алгоритмов [2]. Легкие как орган дыхательной системы играют важную роль в поддержании гомеостаза орга-низма, обеспечивая газообмен, поддерживая кислотно-щелочной баланс крови. Нарушение этих ключевых функций вследствие патологических состояний может приводить к необратимым для организма последствиям. Так, наиболее частой причиной смерти пациентов с инфекцией COVID-19 был респираторный дистресс-синдром взрослого типа (РДСВ), морфологически проявляющийся диффузным альвеолярным повреждением (ДАП). Создание алгоритма оценки морфологических изменений легких при инфекции COVID-19 с помощью подходов компьютерного ана-лиза позволяет внедрить элементы цифровой патологии в научно-исследовательскую работу, а также образовательный процесс.

# Методы и методологии

* 1. **Изучаемая выборка WSI**

Для построения нейросетевой модели был собран набор данных, состоял из 200 полнослайдовых изображений (WSI) гистологических препаратов легких от 70 пациентов (2021-2022 гг.), умерших от COVID-19. Препараты окрашены гематоксилином и эозином. Данные были предоставлены кафедрой патологической анатомии БГМУ и Objective Whole Slide Image Server. Все случаи были анонимизированы. Изображения NIH и Objective Whole Slide Image Server были оцифрованы с 40-кратным увеличением.

**1.2 Формирование наборов данных**

200 полнослайдовых изображений (WSI) c помощью аннотационных инструментов платформы и расширяемой интегрированной средой разработки (на базе QuPath) были разделены на плитки размером 256 × 256 и 512 × 512 пикселей. Все изображения были нормализованы в цвете и увеличении. Было выделено три необходимых набора данных: главный, набор редких гистологических признаков (вспомогательный), набор для тестирования программы (контрольный).

Главный набор данных содержал основные и наиболее встречающиеся морфологические признаки ДАП при коронавирусной инфекции COVID-19 для первоначальной подготовки нейросетевой модели. Данный набор состоял из 40 случаев (104 WSI) и был создан с целью построения модели для классификации графических плиток.

Вспомогательный набор для предварительной подготовки (10 случаев, 26 WSI) для извлечения редких, но важных гистопатологических признаков, таких как: сидерофаги в просветах альвеол, лимфоцитарные инфильтраты при васкулитах. Этот набор состоял из аутопсий легких, взятых в период с 2020 по 2022 год.

Тестовый набор состоял из случаев, отобранных в период с 2021 по 2022 год: 20 произвольно выбранных случаев (56 полнослайдовых изображений) аутопсий легкого при коронавирусной инфекции COVID-19. В последующем этот набор использовался для оценки точности работы программы.

**1.3 Процесс подготовки WSI**

Предметные стекла сканировали с 40-кратным увеличением в цифровые полнослайдовые изображения с помощью цифрового сканера Aperio VERSA Brightfield, Fluorescence & FISH Digital Pathology Scanner (Leica Biosystems).

На рисунке 2 показывает демонстрируются основные этапы классификации/сегментации полученных данных с использованием графовых нейронных сетей. В основном наборе для предварительной подготовки WSI были разбиты на изображения с коэффициентом перекрытия равным 0 и размерами 256 × 256 и 512 × 512 пикселей с увеличением в 2 ×, 10 ×, 20 × и 25 раз соответственно. Для данной работы мы выделили четыре различных увеличения, которые позволяют охарактеризовать различные морфологические признаки и иногда не перекрывающиеся морфологические данные, имеющие важное значение для оценки степени выраженности патоморфологических изменений. Для сегментации фона (не является регионом интереса) использовались пиксели со всеми значениями выше 220 (практически чистый белый цвет) в 24-битном цветовом пространстве RGB, изображения с более чем 90% белого фона исключались. При увеличении ×2 из одного слайда было получено более 400 плиток, при увеличении ×10 из одного слайда было получено 6410 плиток, при увеличении ×20 из одного слайда было получено 33476 плиток, при увеличении ×25 из одного слайда было получено более 64578 плиток, все плитки отбирались произвольным образом. WSI с неподходящим окрашиванием, нечетким изображением, с включениями пузырьков воздуха или жидкости, посторонних микрочастиц не подходили для анализа и были исключены. Информативность (“заполненность” WSI тканью) каждого изображения составляла не менее 70% от всей площади плитки.

**1.4 Выделение плиточных кластеров**

Тайлы в основном наборе для предварительной подготовки были преобразованы в векторы признаков, состоящие из 128 значений. Чтобы объединить похожие изображения, эти векторы признаков были сгруппированы с использованием алгоритма K-средних для каждого увеличения. Чтобы обеспечить всестороннее представление о патологических изменениях, которые характеризуют каждый кластер, была создана графовая карты путем случайного выбора 120 плиток из каждого кластера. Было протестировано различное количество кластеров: 5, 10, 30, 50 80 кластеров. Было выявлено, что в результате использования слишком малого количества кластеров получаются малоинформативные данные, а использование слишком большого количества кластеров сильно усложняет работу нейросети, и в результате не получается ожидаемого увеличения эффективности соразмерно временным затратам.

**1.5 Интеграция кластеров и трансферное обучение**

Кластеры, имеющие статистически значимые схожие патологические изменения, группировались в отдельные классы (рисунок 1).

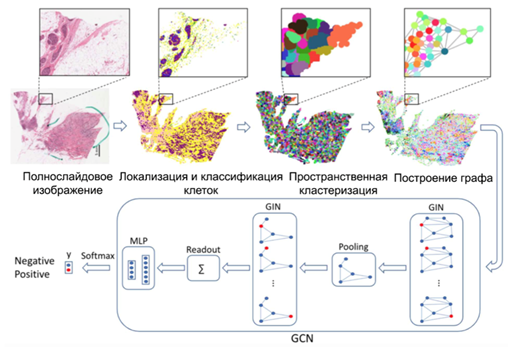


Рисунок 1 – Этапы классификации

Все морфологические признаки были разделены на “патологические” и “структурные/опорные” объекты. Морфологические данные, которые классифицировали для 2-кратного увеличения: “патологические” – ателектазы, дистелектазы полнокровие крупных сосудов, альвеолярный геморрагический синдром, фиброз, примерная норма; “структурные/опорные” компоненты – вены и артерии, бронхи/бронхиолы, хрящи. Для 10-кратного увеличения: “патологические” – гиалиновые мембраны на разных стадиях развития, н[акопление экссудата в альвеолах,](https://mypresentation.ru/documents_5/64c28c6b84ffdce276016fea6dfffa81/img21.jpg) утолщение стенок альвеол, сладж-феномен, лимфоцитарные инфильтраты при васкулитах; “структурные/опорные” компоненты – венулы и артериолы

бронхиолы. При 20-кратном увеличении: “патологические” – тромбы различных видов. Для 25-кратного увеличения: “патологические” – гигантоклеточная трансформация альвеолярного эпителия/симпласты, десквамация стенок бронхиол с полиморфноклеточной воспалительной инфильтрацией, микроангиопатия, плоскоклеточная метаплазия альвеолярного эпителия, сидерофаги в просветах альвеол; “структурные/опорные” компоненты – кровеносные капилляры. Кластеры, которые не вписывались ни в одну из групп или не поддавались однозначной трактовке, были сгруппированы в один класс «другие». Кластеры, которые нельзя определить только одним морфологическим признаком, были исключены.

**1.6 Определение стадий ДАП**

При определении стадий ДАП были выделены две наиболее клинически значимых фазы: экссудативная и пролиферативная. Работу значительно усложнил тот факт, что при короновирусной инфекции COVID-19 наблюдается некоторое несоответствие стадий диффузного альвеолярного повреждения с длительностью заболевания: в некоторых случаях после 2-х недель заболевания встречаются патоморфологические признаки, характерные для экссудативной стадии, и отсутствуют изменения, наблюдаемые в пролиферативную стадию. Для преодоление данной трудности маркерные морфологические были условно разделены на две группы: первая - признаки характерные для экссудативной фазы, вторая - признаки характерные для пролиферативной фазы (таблица 1).

Таблица 1 – Распределение морфологических признаков в зависимости от фазы ДАП

|  |  |
| --- | --- |
| **Признаки характерные для экссудативной фазы** | **Признаки характерные для пролиферативной фазы** |
| * четко выраженные нарушения кровообращения: альвеолярный геморрагический синдром, полнокровие сосудов (ветвей легочной артерии и вен, капилляров межальвеолярных перегородок), тромбы в просветах сосудов различных видов (смешанный, организующийся, фибриновый), сладж-феномен | * сидерофаги в просветах альвеол |
| * десквамация и метаплазия бронхиального и бронхиолярного эпителия | * ателектазы |
| * внутриальвеолярный и интерстициальный отек | * утолщение межальвеолярных перегородок |
| * гиалиновые мембраны на разных стадиях развития, дистелектазы | – |
| * отек и инфильтрация межальвеолярных перегородок лимфоцитами, плазматическими клетками, макрофагами, небольшим числом нейтрофилов | – |
| * десквамированные пласты уродливых клеток, иногда многоядерных, альвеолярного эпителия, пролиферация и очаговый некроз альвеоцитов II типа, встречаются симпласты | – |

Продолжение таблицы 1

**1.7 Структура платформы**

ИПППВР состоит из трех программных компонентов — модели сегментации, модели классификатора и модели активного ученика, — которые совместно поддерживают пользователя-аннотатора платформы (рисунок 2).



Рисунок 2 – Общее описание внутренних процессов системы поддержки принятия решения

Модель сегментации определяет расположение всех клеток в небольшой области. Аннотатор, обычно квалифицированный патологоанатом, начинает маркировать клетки. Эти метки используются для обучения модели классификатора, которая идентифицирует и помечает области, которые может впоследствии скорректировать аннотатор. По мере того, как классификатор учится на этих исправлениях, его прогнозы постепенно улучшаются и требуют меньшего количества исправлений. Когда интересующие патологии достаточно размечены (20–30 аннотаций на класс), модель активного ученика определяет следующую наиболее информативную область для аннотации, и процесс повторяется.

# Результаты

В результате для оценки результатов работы модели необходимо количественно определить минимальное число WSI из датасета с присутствующими и отсутствующими патоморфологическими изменениями, вызванными COVID-19, требуемых для достижения желаемого качества разметки (создания графической маски соответствующих патологий), с использованием разработанного инструмента аннотации. Качество маски измеряется F1-мерой. F1-мера есть оценка разработанной маски в сравнении с маской, размеченной квалифицированным патологоанатомом. В исследовании использованы предварительно сформированные положительные и отрицательные выборки, число WSI в которых постепенно возрастало для достижения желаемых показателей F1 (таблица 2). В этой модели выборка патологий осуществлялась случайным образом в пределах заданной площади на WSI. Аналогично, участки без патологий случайным образом выбирались за пределами области патологических изменений. Для обеспечения точности эксперимента использовалось 20 отрицательных и положительных наборов в датасете. То есть, если при моделировании проверялись маски, созданные с использованием одного положительного и одного отрицательного выбора, эти выборки были случайным образом перемешаны 20 раз.

Таблица 2 – Оценка производительности инструмента аннотации

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| F-мера | >0.60 | >0.65 | >0.70 | >0.75 | >0.80 | >0.85 | >0.90 |
| СOVID-19+ (\*) | 1.54 | 1.82 | 2.18 | 2.73 | 3.77 | 4.20 | 5.00 |
| COVID-19- (\*) | 2.86 | 4.66 | 6.60 | 8.00 | 12.33 | 13.39 | 17.90 |
| % из выборки WSI | 100.0 | 100.0 | 100.0 | 98.89 | 95.56 | 87.78 | 75.56 |

2 Примечания

1 \*– выборка из датасета с наличием патологических изменений, вызванных COVID-19 (СOVID-19+) и их отсутствием (COVID-19-)

В результате была разработана система интегрированная платформа поддержки принятия врачебных решений (ИПППВР), позволяющая создавать эффективные и высококачественные аннотации (рисунок 3) в любом масштабе, производить обмен данными, использовать всю информацию для обучения как моделей глубокого обучения, так и пользователей системы (студентов, исследователей).

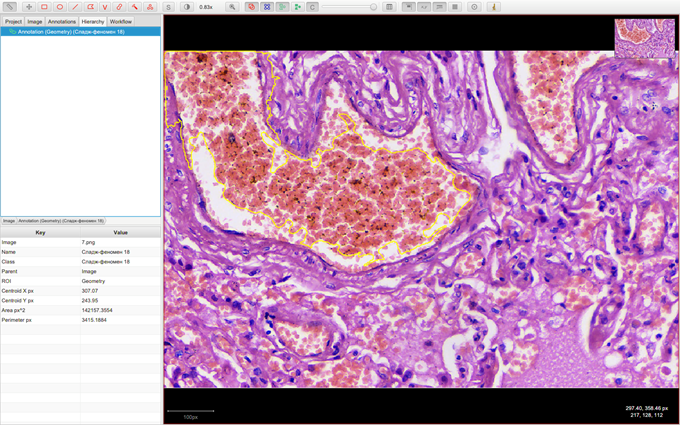


Рисунок 3 – Пример описания объекта для системы принятия решений с помощью встроенных в платформу инструментов аннотации

Так, очевидны преимущества данного подхода. Во-первых, модель классификатора ИПППВР снижает рабочую нагрузку, заменяя трудоемкую задачу первичных аннотаций более простой задачей исправления аннотаций. Во-вторых, модель активного ученика ИПППВР повышает эффективность работы с данными, указывая аннотатору на более информативные области на WSI. Чтобы продемонстрировать эти преимущества, были произведены измерения рабочей нагрузку (доля предложений ИИ, требующих исправления) и эффективность обработки данных (точность проверки по сравнению с размером выборки) для семи патологов, использующих ИПППВР для маркировки морфологических изменений легких при коронавирусной инфекции COVID-19. Использование ИПППВР привело к значительному снижению рабочей нагрузки для всех задач, от 96% для разметки патологических клеток (самая простая задача) до 83% для эозинофилов. Получается, что патологоанатом, вооруженный ИПППВР, создаст примерно в 10 раз больше аннотаций за тот же промежуток времени. В дополнение к этим объективным улучшениям расширение ИИ также позволяет патологоанатомам отслеживать улучшения классификатора и выявлять ошибки на протяжении всего процесса разметки. Это может способствовать более тесному сотрудничеству с инженерами-программистами и способствовать пониманию недостатков модели и распределению ресурсов для маркировки. Подобные системы широко представлены как коммерческие решения и используются в таких областях, как распознавание объектов [6], системы автоматического управления в автомобилях, классификация дерматологических изменений. В итоге нашей работы была создана нейронная сеть с возможностью определения следующих патоморфологических изменений: тромбы в просветах сосудов различных видов (смешанный, организующийся, фибриновый) (рисунок 4), серозный и фибринозно-гнойный экссудат (рисунок 5), утолщение стенок альвеол, геморрагии, ателектазы, дистелектазы, интраальвеолярный отек, гигантоклеточная трансформация альвеолярного эпителия, гиалиновые мембраны на разных стадиях развития, интерстициальная воспалительная инфильтрация, десквамация стенок бронхиол с полиморфноклеточной воспалительной инфильтрацией, микроангиопатия, плоскоклеточная метаплазия альвеолярного эпителия, сладж-феномен, полнокровие крупных сосудов, альвеолярный геморрагический синдром, сидерофаги, фиброзная ткань, скопление нейтрофилов, а также такие структурные элементы легких как: бронхи, вены и венулы, артерии и артериолы, капилляры, хрящи.

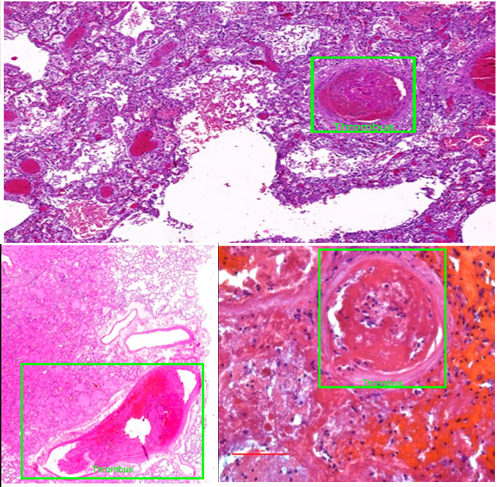


Рисунок 4 – Обнаружение тромбов

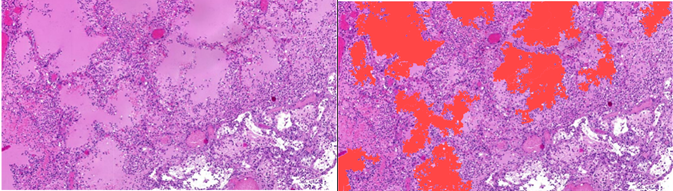


Рисунок 5 – Обнаружение экссудата

Программа способна определять вероятную фазу ДАП и количество встречаемости каждого морфологического изменения с возможностью пространственного картирования. В области иммунного ответа нейронная сеть способна различать клетки иммунной (рисунок 6) и "неиммунной" природы (клетки крови) (рисунок 7).

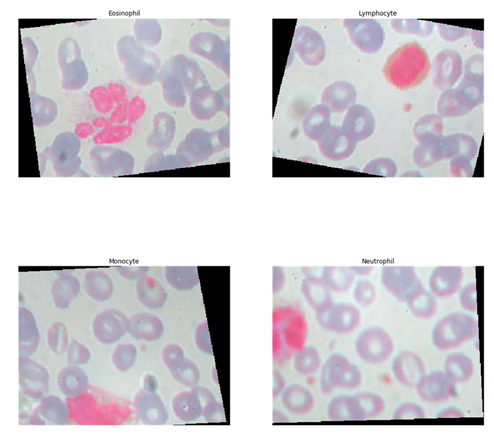


Рисунок 6 – Определение типов иммунных клеток

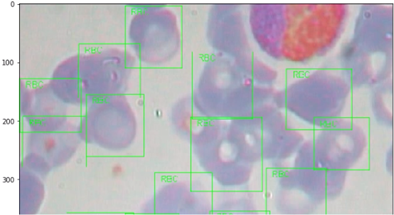
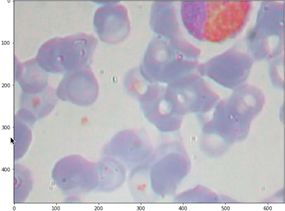


Рисунок 7 – Обнаружение клеток крови

Общая точность классификации составила 83,1% и 79,7% по данным обучения и независимого тестирования датасетов соответственно (рисунок 8).

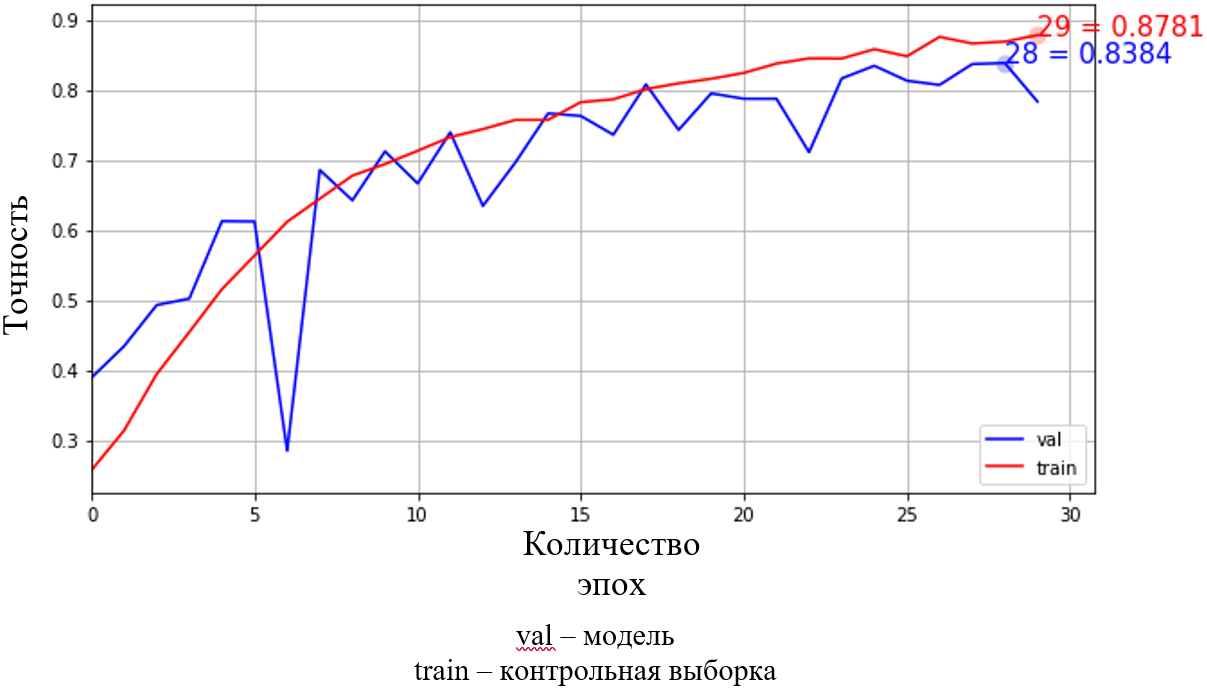


Рисунок 8 – Количественные характеристики модели в процессе обучения

Идентифицируя клетки и классифицируя типы клеток, нейросеть может преобразовать изображение патологии в «пространственную карту» клеток (рисунок 9).

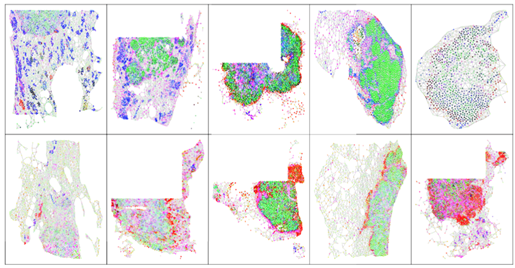


Рисунок 9 – Визуализация кластеров

Из этой пространственной карты мы можем извлечь признаки (рисунок 10), характеризующие микроокружение исследуемых объектов. На основе этих особенностей разработана прогностическая модель классификатора.

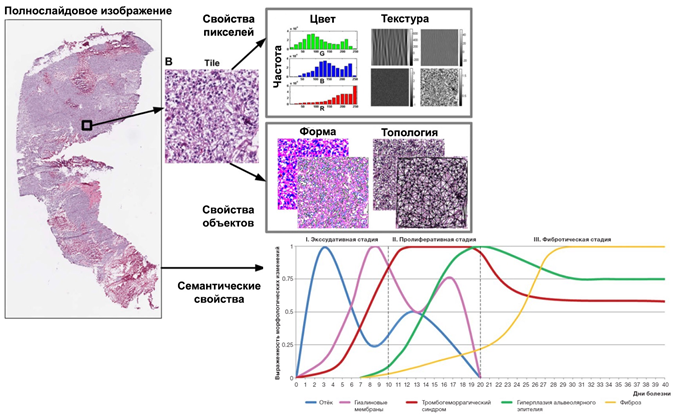


Рисунок 10 – Извлечение свойств нейросетевой моделью

Кроме того, для решения образовательных задач, был создан отдельный элемент ИПППВР – методический раздел, в котором в интерактивном виде собрана основная литература, полезная при изучении патологической анатомии и конкретных патологических изменений. После прочтения соответствующих разделов пользователю платформы предлагается пройти тест или “дообучить” нейросетевую модель, проделав несколько практических упражнений по разметке и аннотации с возможностью контроля и самопроверки. Таким образом в перспективе также могут быть решены врачебные задачи справочно-информационной поддержки, определения степени и тяжести заболевания, генерации тревожных сигналов, ассистирования в диагностике и оптимизации лечения.

# Обсуждение

1. Созданная нейросетевая модель позволяет классифицировать клетки иммунного отве-та и определять ограниченный спектр морфологических изменений в легких.

2. Разработанная образовательно-исследовательская программная платформа позволила оптимизировать аннотацию и разметку, платформа представляет новые возможности обучения как студентов, так и нейросетевых моделей в направлении патологической анатомии.

# Заключение

В ходе работы удалось сформировать оригинальный подход для выделения множества гистологических признаков, основанный на объединении морфологических и математических методов анализа патологических изменений, которые могут быть использованы патологоанатомами с минимальными усилиями для аннотации полнослайдовых изображений. Модель не только эффективно описывает количество и распределение признаков для различных стадий COVID-19, но также результативна в количественном анализе гистологических признаков. Тот же подход может быть применен к другим областям патологии и представляет собой новое направление для объяснительных аналитических моделей. Платформа в свою очередь предоставляет программные компоненты для аннотации, хранения и обработки данных, в том числе нейросетевых моделей, результатов их обучения и образовательных ресурсов. ИПППВР может быть использована как информационная площадка в научно-исследовательских и образовательных целях, обладает неограниченной возможностью пополнения новыми моделями и практическими ресурсами.

Таким образом, созданная нейросетевая модель позволяет классифицировать клетки иммунного ответа и определять ограниченный спектр морфологических изменений в легких, а также значительно оптимизирует рабочие процессы в практике патологоанатомов, в частности аннотацию и разметку, представляет новые возможности обучения как студентов, так и нейросетевых моделей в направлении патологической анатомии.

# Литература

1 ‬Pathology image analysis using segmentation deep learning algorithms

2 Two distinct immunopathological profiles in autopsy lungs of COVID-19 / R. Nienhold, Y. Ciani, V. Koelzer, [et al.] // Nat Commun. – 2020. – No. 11:5086 – P. 4-6.

3 QuPath: open source software for digital pathology image analysis / P. Bankhead, M. Loughrey, J. Fernandez, [et al.] // ‬Sci Rep. – 2017. – No. 7(1):16878. – P. 2.

4 Validation of digital pathology imaging for primary histopathological diagnosis / D. Snead, Y. Tsang, A. Meskiri, [et al.] // Histopathology. – 2016. – No. 68(7):1063. – P. 72.

5 Discriminative unsupervised feature learning with exemplar convolutional neural networks / D. Snead, Y. Tsang, A. Meskiri, [et al.] // IEEE Trans Pattern Anal Mach Intell. – 2016. – No. 38(9):1734. – P. 47.

6 Autopsy in suspected COVID-19 cases / B. Hanley, S.B. Lucas, E. Youd, B. Swift, M. Osborn // J Clin Pathol. – 2020. – No. 73. – P. 239-242.