

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ ПРИКЛАДНОЙ МАТЕМАТИКИ И ИНФОРМАТИКИ

Кафедра дискретной математики и алгоритмики

МАЛЫШЕВ Валерий Дмитриевич

**АЛГОРИТМЫ ИЕРАРХИЧЕСКОГО АНАЛИЗА
ГИСТОЛОГИЧЕСКИХ ИЗОБРАЖЕНИЙ**

Магистерская диссертация
специальность 1-31 80 09 Прикладная математика и информатика

Научный руководитель
Ковалев Василий Алексеевич
ученая степень, ученое звание
кандидат технических наук

Допущена к защите

«___» _____ 2021 г.

Заведующий кафедрой дискретной математики и алгоритмики

_____ В. М. Котов

доктор физ.-мат. наук, профессор

Минск, 2021

ОГЛАВЛЕНИЕ

Введение.....	3
Общая характеристика работы	4
Глава 1 Полнослайдовые изображения и сложности работы с ними.....	7
Глава 2 Обзор методов для анализа полнослайдовых изображений	12
Глава 3 Исходные данные	17
Глава 4 Алгоритмы предобработки полнослайдовых изображений	21
4.1 Выделение ткани на изображении	21
4.2 Нормализация цветового пространства.....	23
Глава 5 Используемые методы глубокого обучения.....	26
Глава 6 Каскадный анализ.....	29
6.1. Стандартный подход.....	29
6.2. Подход с добавлением признаков	34
6.3 Предсказание уверенности в классе центральных тайлов.....	36
Глава 7 Результаты.....	39
Заключение	42
Список использованных источников	43
Приложение А ROC-кривые обученных нейронных сетей.....	46
Приложение Б Гистограммы распределения метрик при тестировании	47

ВВЕДЕНИЕ

Тема научной работы связана с применением новых достижений в области искусственного интеллекта для решения задач анализа полнослайдовых изображений. Эти изображения получаются путем сканирования образцов ткани специализированными аппаратами. В сфере цифровой патологии, полнослайдовые изображения позволяют получить точный диагноз и являются золотым стандартом для диагностики ряда заболеваний, в особенности рака.

Автоматический анализ полнослайдовых изображений является крайне трудоемким процессом ввиду ряда причин, таких как вариативность в зависимости от типа ткани и процесса подготовки, присутствия различных артефактов. Еще одно значительное ограничение связано с разрешением полнослайдовых изображений, которое может достигать 100 000x200 000 пикселей, что делает невозможным прямое применение классических подходов машинного и глубокого обучения. Кроме этого, огромное количество визуальной информации значительно усложняет работу профильных специалистов, что показывает актуальность внедрения методов автоматизированной диагностики в данную сферу.

Целью данной работы является оптимизация применения алгоритмов глубокого обучения для полнослайдовых изображений за счет использования представления полнослайдовых изображений как иерархичной структуры. Во-первых, в данной работе демонстрируется полезность добавления к региону изображения информации об окружающем его контексте. Во-вторых, показано, что использование двухуровневого алгоритма позволяет получать результаты за более короткое время, чем детальный анализ всех регионов изображения, но с сравнительно высокой точностью.

В первой главе описаны особенности полнослайдовых изображений, а также проблемы, возникающие при работе с ними. Во второй главе сделан обзор литературы и существующих решений, включая иерархичные подходы. В третьей главе представлены данные, которые использовались для проверки поставленных гипотез. Используемые методы для предварительной обработки изображений и методы глубокого обучения описаны в четвертой и пятой главах соответственно. Шестая глава описывает разработанный метод применительно к классификации регионов полнослайдовых изображений рака простаты и представляет его сравнение с другими методами. Полученные результаты описаны в седьмой главе для сравнительного анализа разработанных методов.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Магистерская диссертация, 50 с., 23 рис., 9 табл., 2 прил., 30 источн.

ГИСТОЛОГИЯ, РАК, ПОЛНОСЛАЙДОВЫЕ ИЗОБРАЖЕНИЯ, НЕЙРОННЫЕ СЕТИ.

Объект исследования – задача определения типа ткани на регионах полнослайдовых изображений рака простаты с использованием каскадов нейронных сетей.

Цель работы – исследование методов на основе нескольких нейронных сетей, объединенных для улучшения определения рака простаты на полнослайдовых изображениях.

В ходе работы были исследованы существующие подходы к объединению нейронных сетей в каскады для решения поставленной задачи. Был проведен сравнительный анализ методов и рассмотрены способы их усовершенствования. Также были рассмотрены ранее не применявшиеся методы объединения нескольких нейронных сетей для каскадного анализа полнослайдовых изображений.

Актуальность данной работы обусловлена необходимостью разработки точных и эффективных подходов к обнаружению рака на полнослайдовых гистологических изображениях. Существующие методы для автоматического анализа гистологии требуют высокого потребления вычислительных ресурсов в течении долгого времени для обработки одного изображения. Кроме того, существует ряд проблем, порожденных вариативностью гистологических данных.

Диссертация состоит из разделов «Общая характеристика работы», «Введение», основной части, «Заключение», «Список использованных источников» и двух приложений с графическим материалом, поясняющих результаты тестирования описанных алгоритмов. Основная часть включает 7 глав. В первой главе описываются полнослайдовые изображения и их структура. Во второй главе представлен краткий обзор существующих решений для анализа гистологических изображений, включая каскадный анализ. Третья глава дает представление об использованных данных в работе. Четвертая и пятая главы описывают использованные методы для предобработки изображений и методы глубокого обучения соответственно. В шестой главе представлены исследуемые подходы к каскадному анализу полнослайдовых изображений, а в седьмой – результаты тестирования данных подходов.

АГУЛЬНАЯ ХАРАКТЕРЫСТЫКА РАБОТЫ

Магістарская дысертацыя, 50 с., 23 мал., 9 табл., 2 дад., 30 крын.

ГІСТАЛОГІЯ, РАК, ПОЎНАСЛАЙДАВЫЕ МАЛЮНКІ, НЕЙРОННЫЯ СЕТКІ.

Аб'ект даследавання – задача вызначэння тыпу тканіны на рэгіёнах поўнаслайдавых малюнкаў рака прастаты з выкарыстаннем каскадаў нейронавых сетак.

Мэта работы - даследаванне метадаў на аснове некалькіх нейронавых сетак, аб'яднаных для паляпшэння вызначэння рака прастаты на поўнаслайдавых малюнках. У ходзе работы былі даследаваны існуючыя нейрасеткавыя падыходы да аб'яднання нейронавых сетак у каскады для вырашэння пастаўленай задачы. Быў праведзены параўнальны аналіз метадаў і разгледжаны спосабы іх ўдасканалення. Таксама былі разгледжаны раней не выкарастанавыя метады аб'яднання некалькіх нейронавых сетак дзеля каскаднага аналізу поўнаслайдавых малюнкаў.

Актуальнасць магістарскай работы абумоўлена неабходнасцю распрацоўкі дакладных і эфектыўных падыходаў да выяўлення рака на поўнаслайдавых гісталагічных малюнках. Існуючыя метады для аўтаматычнага аналізу гісталагіі патрабуюць высокага спажывання вылічальных рэсурсаў у плыні доўгага часу для апрацоўкі аднаго малюнка. Акрамя таго, існуе шэраг праблем, спароджаных варыятыўнасцю гісталагічных малюнкаў.

Дысертацыя складаецца з раздзелаў «Агульная характарыстыка работы», «Уводзіны», асноўнай часткі «Заклучэнне», «Спіс выкарыстаных крыніц» і двух прыкладанняў з графічным матэрыялам, што тлумачыць вынікі тэставання апісаных алгарытмаў. Асноўная частка ўключае 7 раздзелаў. У першым раздзеле асноўнай часткі апісваюцца агульныя складанасці, якія ўзнікаюць пры працы з поўнаслайдавымі малюнкамі. У другым раздзеле прадстаўлены кароткі агляд існуючых рашэнняў да аналізу гісталагічных малюнкаў, уключаючы каскадны аналіз. Трэці раздзел дае ўяўленне аб выкарыстаных даных у працы. Чацвёрты і пяты раздзелы апісваюць выкарыстаныя метады для папярэдняй апрацоўкі малюнкаў і метады глыбокага навучання адпаведна. У шостым раздзеле прадстаўлены доследныя падыходы да каскаднага аналізу поўнаслайдавых малюнкаў, а ў сёмым - вынікі тэставання дадзеных падыходаў.

ABSTRACT

Master's thesis, 50 pp., 23 figures, 9 tables, 2 appendices, 30 sources.

HISTOLOGY, CANCER, WHOLE-SLIDE IMAGES, NEURAL NETWORKS.

The object of the research is the problem of determining the type of tissue on the regions of whole-slide images of prostate cancer using cascades of neural networks.

The aim of the work is to study methods based on several neural networks combined together to improve the detection of prostate cancer in whole-slide images.

In the course of the work, the existing neural networks and approaches to combining neural networks into cascades were investigated to solve the problem. A comparative study of the methods was carried out and the ways of their improvement were considered. Also, a range of approaches that were not previously used for cascade analysis of whole-slide images were studied.

The need to develop accurate and effective approaches to detecting cancer on whole-slide histological images explains the relevance of the master's thesis. The existing methods for the automatic analysis of histology require a high consumption of computational resources for a long time to process a single image. In addition, there are a number of problems caused by the variability of histological data.

The thesis consists of sections "General characteristics of the work", "Introduction", the main part "Conclusion", "List of sources used" and two appendices with graphic material explaining the results of testing the described algorithms. The main part includes 7 chapters. The first chapter describes the common difficulties that arise when working with whole-slide images. The second chapter provides a brief overview of existing solutions for the analysis of histological images, including cascade analysis. The third chapter gives an idea of the data used in the work. The fourth and fifth chapters describe the methods used for image preprocessing and deep learning methods, respectively. The sixth chapter presents the investigated approaches to the cascade analysis of full-slide images, and the seventh – the results of testing these approaches.

ГЛАВА 1

ПОЛНОСЛАЙДОВЫЕ ИЗОБРАЖЕНИЯ И СЛОЖНОСТИ РАБОТЫ С НИМИ

Полнослайдовые изображения (в дальнейшем ПСИ) получаются путем сканирования образцов ткани специализированным оборудованием. ПСИ являются одним из методов цифровой патологии, направленных на получение точного диагноза путем анализа гистологических изображений. Например, использование ПСИ позволяет определить рак, когда ряд других методов не всегда способен это точно сделать, а также определить тип и агрессивность рака в различных органах.

Однако размер образца ткани достигает нескольких миллиметров, тогда как разрешающая способность позволяет увидеть отдельные клетки. То есть, разрешающая способность может достигать нескольких микрометров (микрон) на пиксел. Следовательно, такие изображения имеют разрешение порядка нескольких гигапикселей. Анализ ПСИ является крайне трудоемким процессом, что показывает актуальность внедрения автоматических методов анализа. Однако, ПСИ обладают рядом особенностей, затрудняющих разработку данных методов. Этими особенностями являются высокая вариативность, большой объем графической информации.

Большинство полнослайдовых изображений получаются, используя подсветку фона под зафиксированным образцом ткани, как, например, образцы, окрашенные гематоксилином и эозином. Такие образцы (см. рисунок 1.1) показывают морфологию ткани, что сделало такой метод наиболее распространённым в сфере цифровой патологии и, соответственно, наиболее привлекательной модальностью для применения методов машинного обучения ввиду большего количества данных. Образцы, окрашенные гематоксилином и эозином,.

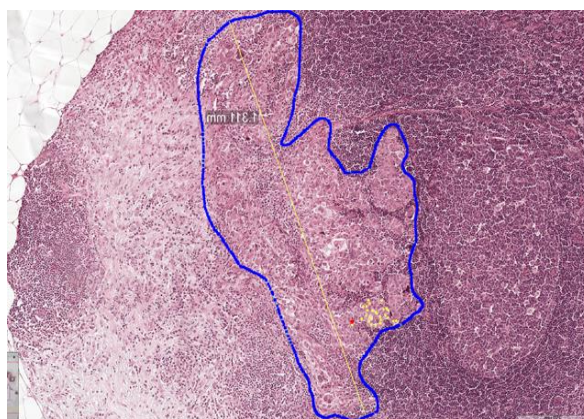


Рисунок 1.1 – Регион полнослайдового изображения с выделенной раковой областью

Тем не менее гематоксилин и эозин не показывают другую информацию, которую можно получить из образца ткани. Например, иммуногистохимия (см. рисунок 1.2) позволяет увидеть степень экспрессии различных белков в клетке, что позволяет определить тип рака. Поэтому иммуногистохимия используется для подтверждения диагноза, полученного при анализе образцов окрашенных гематоксилином и эозином.

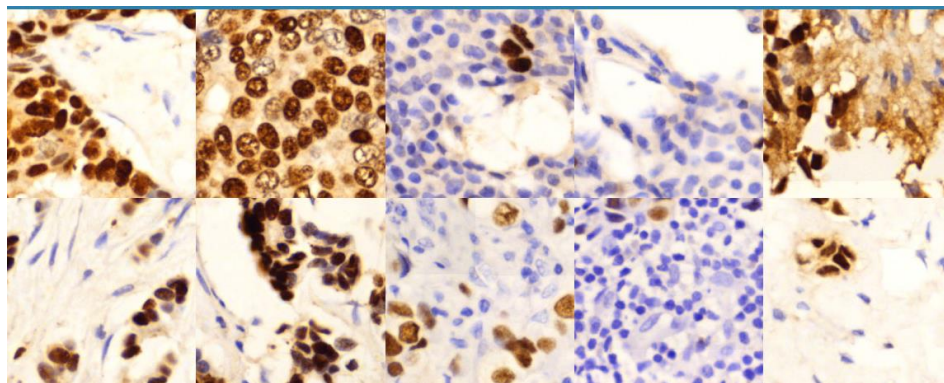


Рисунок 1.2 – Регионы иммуногистохимического изображения

В дополнение к способу получения изображений используя подсветку основы, на которой располагается образец, существует флуоресцентная гистология, которая использует различные маркеры (обычно реагирующие на специфичные белковые соединения для рассматриваемого вида клеток). Однако такие реагенты гораздо дороже что не способствует их распространенности (рисунок 1.3).

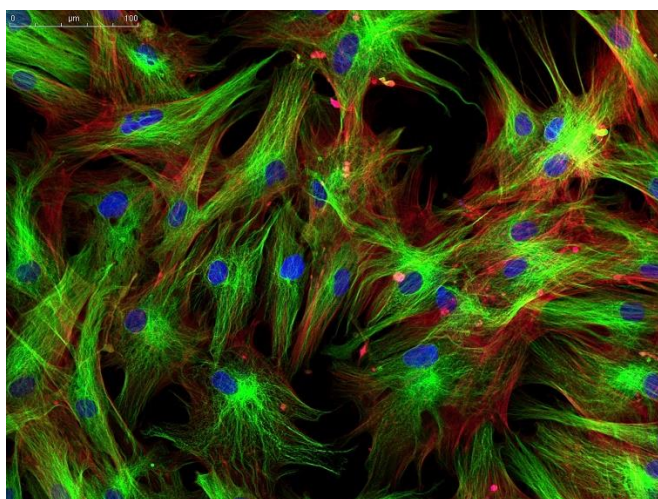


Рисунок 1.3 – Пример флуоресцентной гистологии [1]

На данный момент существует множество производителей гистологических сканеров, способных снимать полнослайдовые изображения разной модальности. Каждый сканер сохраняет и сжимает изображение в разных

форматах, что замедляет разработку алгоритмов, так как различные наборы данных могут иметь изображения, сохранённые в разных форматах. Например, в сфере радиологии эта проблема была решена введением DICOM формата. Тем не менее сейчас ведётся работа для достижения стандартизации форматов данных.

Несмотря на разнообразие форматов, большинство из них обладает свойством иерархичности, то есть изображение сохранено в памяти компьютера 2D блоками, чтобы значительно ускорить доступ к небольшим квадратным участкам. Кроме этого, хранятся также области изображения в более низком разрешении, чтобы не было необходимости загружать из памяти все маленькие блоки для того, чтобы отобразить, например, все изображение (см. рисунок 1.4) [2]. В дальнейшем, в данной работе области одного и того же изображения, но с разным разрешением (увеличением) будут называться «уровнями» изображения, где 0-ой уровень означает максимальное увеличение (что также соответствует максимальному разрешению), а каждый последующий уровень имеет стороны в 2 или 4 раза меньше. Таким образом, иерархическая структура хранения полнослайдовых изображений позволяет получать визуальные данные с фиксированных уровней изображения максимально эффективным способом. Такая структура изображений мотивировала разработку иерархического подхода к анализу ПСИ в данной работе.

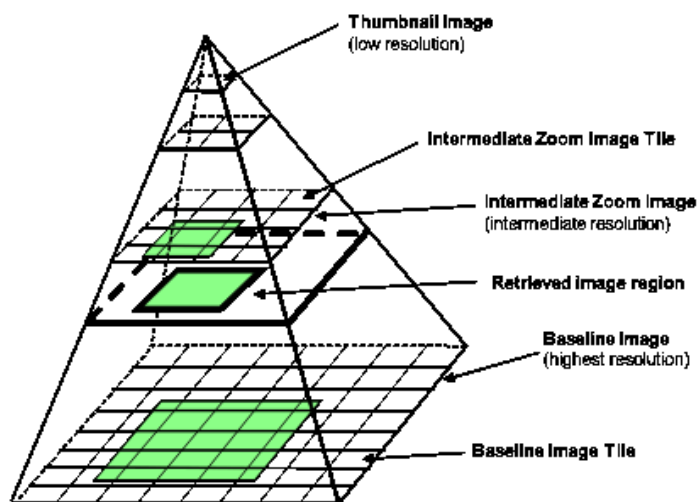


Рисунок 1.4 – Возможная структура полнослайдового изображения в памяти компьютерах [3]

Чтобы иметь достаточную степень обобщаемости для применения в клинической практике, методы машинного обучения должны работать с изображениями, содержащими различные артефакты и подверженных вариативности цветового представления. Артефакты могут появиться как в

процессе подготовки образцов ткани, так и в процессе получения изображений (см. рисунок 1.5). Такие артефакты включают разницу в степени экспрессии гистологических маркеров, разрывы ткани, артефакты сканера. Некоторые артефакты могут проявиться при снятии изображений, так как чтобы получить изображения сканер делает небольшие снимки на максимальном увеличении, а потом собирает из них полнослайдовое изображение. Из-за такого метода получения ПСИ могут проявиться такие артефакты, как разница в освещении, фокусе, проблемы со сдвигом камеры на правильное расстояние. Вместе с высокой вариативностью ткани на изображениях это сильно усложняет применение методов глубокого обучения на таких изображениях. В дополнение к этому высокое разрешение ПСИ делает невозможным применение методов глубокого обучения «напрямую». Таким образом в большинстве существующих решений используется разделение ПСИ на небольшие участки, которые используются в выборке данных (такие участки называются «тайлами»). Краткий список проблем, связанных с исходными данными, которые предстоит решить в ходе научной работы представлен ниже:

- высокое разрешение изображений;
- отсутствие контекста если изображение рассматривается по участкам;
- артефакты на изображениях;
- вариативность цветового распределения, связанная с различиями в освещении и окрашивании образцов;
- вариативность образцов ткани и самой ткани.

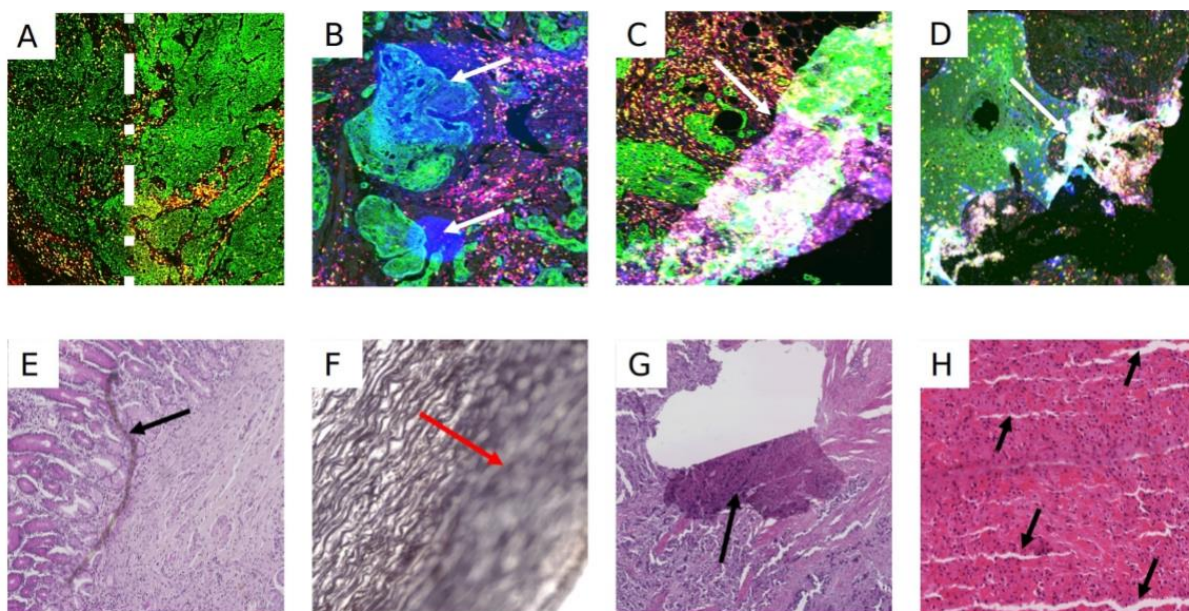


Рисунок 1.5 – Примеры артефактов на гистологических изображениях [2]

Рассмотренные сложности, возникающие при работе с ПСИ, затрудняют использование методов, показывающих хорошие результаты на других задачах. Кроме того, анализ полнослайдовых изображений требует дополнительных алгоритмов для разделения изображений на регионы и агрегации результатов. Но структура полнослайдовых изображений также позволяет применять иные подходы, с помощью которых можно получить лучшие алгоритмы анализа ПСИ, как например используемые в данной работе подход, направленный на использование иерархической структуры изображений.

ГЛАВА 2

ОБЗОР МЕТОДОВ ДЛЯ АНАЛИЗА ПОЛНОСЛАЙДОВЫХ ИЗОБРАЖЕНИЙ

На сегодняшний день существует множество алгоритмов для анализа полнослайдовых изображений, так как большая часть задач компьютерного зрения присутствует в анализе полнослайдовых изображений. Стоит иметь в виду, что в большинстве случаев основной задачей, стоящей перед автоматическими системами, является получение так называемого второго мнения для одного полнослайдового изображения. Однако, ввиду причин, указанных в предыдущей главе получить набор характеристик изображения возможно только через разбиение изображения на участки, решение подзадач и агрегацию результатов. В данной работе, в основном, рассматриваются методы для улучшения второго этапа, а именно решения некоторой подзадачи на участках полнослайдового изображения.

В ряде работ [4-6] применяется следующий подход. Первым шагом изображения разбиваются на небольшие квадратные участки, которые именуются тайлами. Далее каждому тайлу назначается класс, в зависимости от типа той ткани, которая занимает большую часть данного тайла. Следующим шагом решается задача классификации тайлов. В итоге, при анализе полнослайдового изображения можно получить класс каждого тайла изображения. Данная информация может быть представлена специалисту для помощи в принятии решения, либо она может быть агрегирована в конечное решение автоматической системы.

Другим возможным подходом к анализу полнослайдовых изображений является сегментация ткани на определённые классы. В данном случае существуют различные подходы в зависимости от алгоритма агрегации результатов. Например, в одной работе [7] использовалось деление на раковые и не раковые области, тогда как в другой задаче [8] использовалась расширенная классификации ткани на типы и несколько стадий агрессивности рака, чтобы в итоге получить более детальный анализ (см. рисунок 2.1).

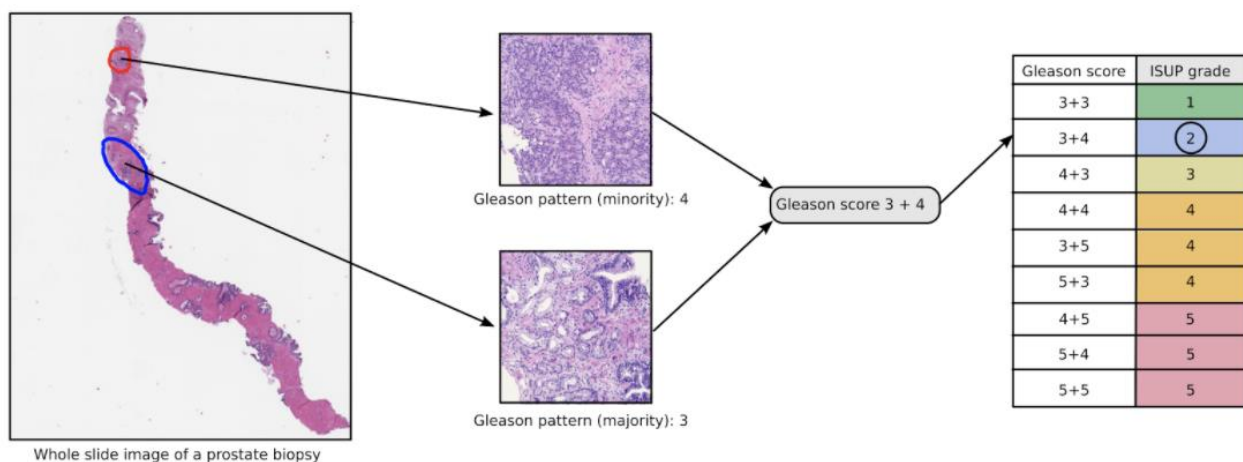


Рисунок 2.1 – Пример сложного процесса агрегации результатов для диагностики рака простаты [8]

Наконец, стоит отметить широкую область задач по локализации и сегментации клеток и возможном делении их на классы. В данной области лидирует нейронная сеть HoVer-Net [9] (см. рисунок 2.2), которая позволяет получать множества объектов небольшого размера и решает ряд недостатков присущих классическим подходам из компьютерного зрения [10, 11] при использовании их для решения задач на гистологических изображениях.

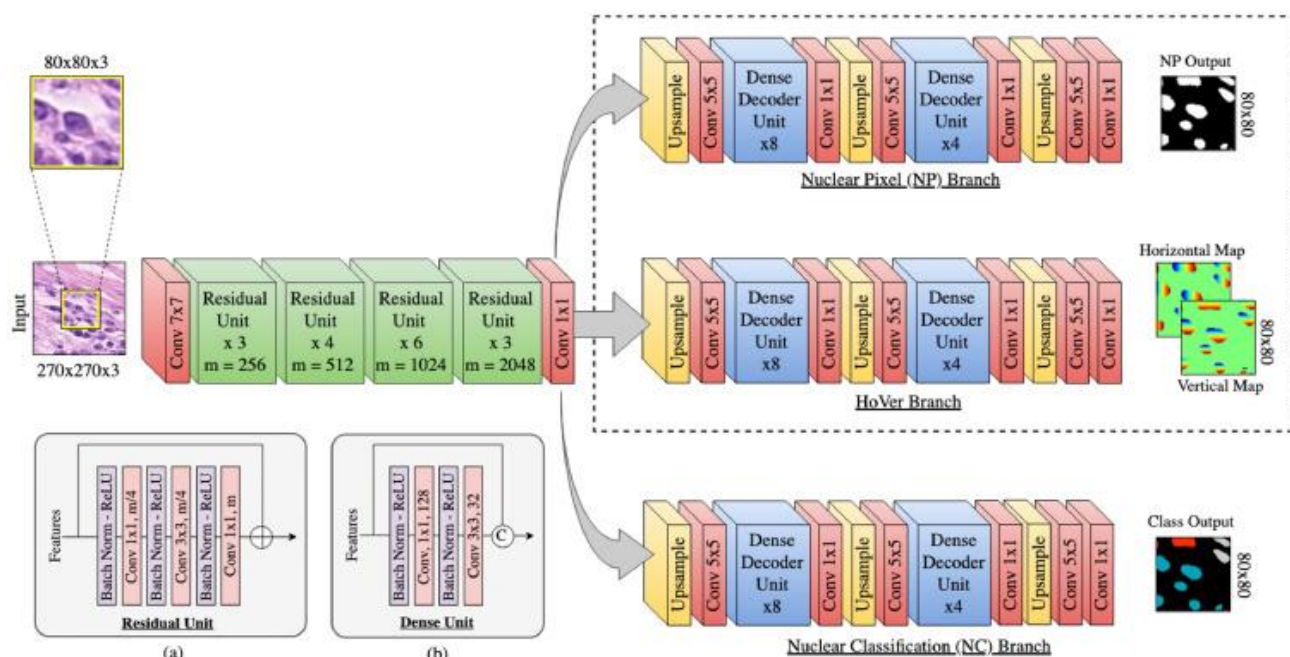


Рисунок 2.2 – Архитектура HoVer-Net

Стоит отметить, что при подготовке данных для обучения нейронных сетей или иных алгоритмов машинного обучения несколько специалистов аннотируют данные. Среди специалистов существует определенный уровень несогласованности при принятии решений, поэтому при оценке качества

полученных алгоритмов требуется учитывать, что присутствует значительный фактор неопределённости в истинных значениях меток для данных. Более того, алгоритмы глубокого обучения могут достигать более высоких значений согласованности результатов чем группа специалистов (например, 66.6% для обученной нейронной сети против 62.7% согласия между специалистами [12]).

При анализе полнослайдовых изображений может возникнуть вопрос с выбором уровня, с которого стоит выбирать тайлы. В данном случае можно говорить о том, какая информация более важна для конкретной задачи. Например, при поиске ядер клеток на ПСИ лучше брать более высокое разрешение (0-ой или 1-ый уровень) чем при сегментации крупных структур в ткани (2-ой уровень или выше). Примеры тайлов с разных уровней одного и того же изображения показаны на рисунке 2.3.

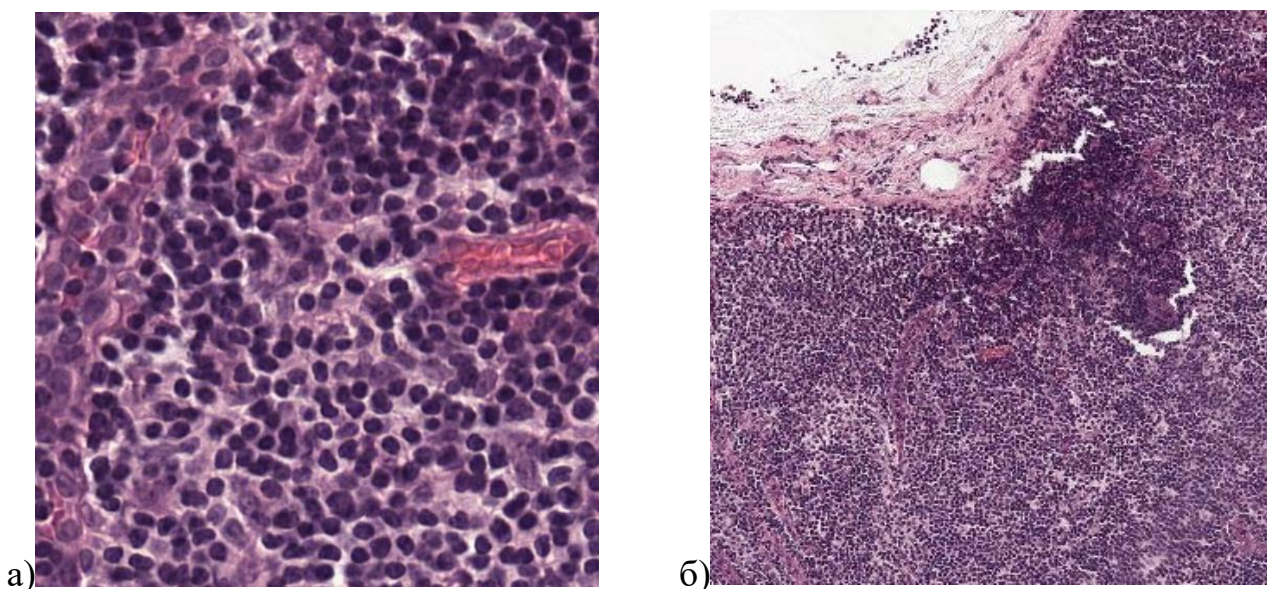


Рисунок 2.3 – Пример а) детального и б) общего вида одного и того же региона полнослайдового изображения

В данной работе рассматривается задача классификации тайлов полнослайдовых изображений рака простаты на 5 классов, из которых два представляют из себя здоровую ткань, а три – разные стадии развития опухоли. Опухолевые области отличаются от нормальной ткани по ряду признаков, таких как размер и форма клеток, организация и делимость области и другие. Часть из этих признаков проявляется на клеточном уровне, а часть на более общем виде изображения. Таким образом, показана одна из основных проблем, существующих в потайловой обработке полнослайдовых изображений, а именно недостаток информации об окружении при анализе маленького региона изображения. Другими, словами при классификации региона изображения значительная доля информации может быть получена не только из того, что

находиться в пределах данного региона, но и за его пределами, в окружающей его ткани, так называемом контексте тайла.

В данной сфере уже получили распространение некоторые методы иерархичного анализа полнослайдовых изображений, которые используют тайлы разного уровня, но изображающие одну область в целях повышения точности решения поставленной задачи. Один из способов заключается в увеличении глубины нейронной сети для увеличения объема визуальной информации, учитываемой при решении задачи на тайлах, таким образом увеличивая эффективный размер тайла [13]. В другом случае предложен алгоритм, состоящий из двух нейронных сетей, в которых одна предсказывает грубое положение объекта, а вторая, используя более детализированные исходные данные, получает точную сегментацию [14]. Такой подход применялся, чтобы компенсировать проблему низкой кучности раковых областей на исходных изображениях. В другой работе [15] применялся альтернативный подход по сегментации клеток на микроскопии. Данный подход состоял из ряда нейронных сетей, каждая из которых получала более детальные маски клеток, чем предшествующая ей (см. рисунок 2.4).

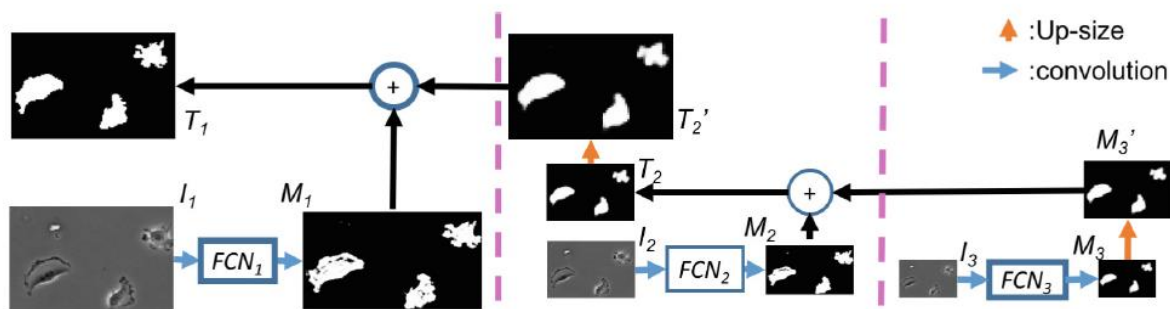


Рисунок 2.4 – Схема сегментации клеток [15]

Другой подход вместо использования нескольких нейронных сетей использует одну U-Net [16] образную сеть, однако, вместо получения признаков и уменьшения изображения с помощью слоев субдискретизации, авторы используют разные уровни ПСИ и несколько сверточных слоев, чтобы получить признаки данного уровня и дальше используют их в увеличивающей части U-net [17] (см. рисунок 2.5).

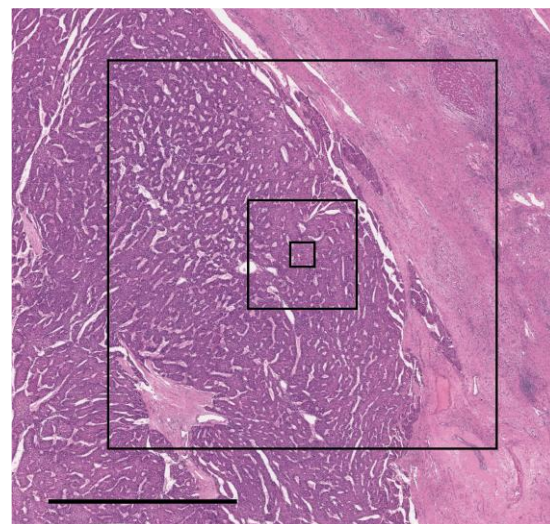
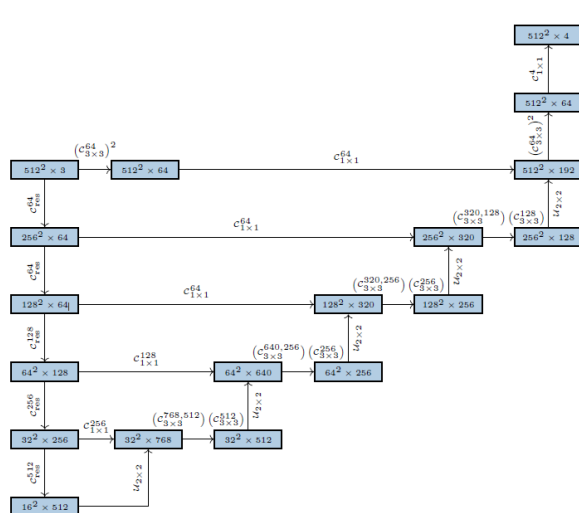


Рисунок 2.5 – Архитектура и получение исходных данных в Multi-Scale CNN [17]

Таким, образом в представленных работах, авторы получают более высокие показатели точности благодаря использованию каскадного подхода к анализу полнослайдовых изображений. Однако, стоит отметить, что анализ одного полнослайдового изображения на максимальном увеличении (x40) может занимать около 8 часов при использовании нескольких GPU, а дополнительное увеличение сложности алгоритма за счет дополнительных сетей/слоев будет увеличивать время обработки полнослайдовых изображений. Поэтому в данной работе этот факт принимается во внимание и показаны несколько оптимизаций позволяющих получить более эффективный и точный каскадный алгоритм чем продемонстрированные выше. Такой результат достигается за счет нескольких средств:

- при обработке на низком разрешении получают признаки для использования более чем для одного тайла высокого разрешения;
- анализ на высоком разрешении не всегда необходим ввиду того, что информации полученной на низком разрешении может быть достаточно, при уверенности в ее точности.

Кроме того, в данной работе решается задача классификации, а не сегментации, как в показанных работах, что позволяет использовать несколько иные подходы к анализу тайлов на нескольких уровнях чем в представленных работах.

ГЛАВА 3

ИСХОДНЫЕ ДАННЫЕ

В качестве исходных данных использовались 10 616 ПСИ с размером большей стороны до 100 000 пикселей, которые были разбиты на регионы размером 256x256 пикселей (тайлы). Данные регионы используются в качестве входных данных для алгоритмов, представленных в данной работе. Изображения были получены из соревнования PANDA (Prostate cANcer graDe Assessment), целью которого была оценка агрессивности рака на изображениях. Однако, в целях данной работы была поставлена иная задача классификации отдельных регионов изображения по типу ткани, представленной на них.

Рак простаты является одним из распространенных видов онкологических заболеваний у мужчин. Каждый год в мире диагностируется более 1 млн новых случаев заболевания и умирает более 350 тыс. человек [18]. Одним из общепризнанных способов уменьшения смертности от рака простаты является его раннее обнаружение и правильная оценка степени агрессивности.

В качестве способа количественной оценки заболевания используется его рейтинг по так называемой шкале Глисона (Gleason Score, [19]), предложенной Дональдом Глисоном. Показатель Глисона определяется визуально по гистологическим изображениям высокого разрешения, которые снимаются с образцов ткани, получаемых в результате биопсии. Существующие варианты данной шкалы незначительно отличаются друг от друга в зависимости от исторически сложившихся предпочтений и некоторых других причин. Однако общей основой оценки является базовая шкала с целыми значениями {0, 3, 4, 5}. При этом уровень 0 соответствует здоровой ткани, уровень 3 назначается участку биопсии с минимальной злокачественностью и положительным прогнозом выживаемости при соответствующем лечении, а уровень 5 обозначает максимальную агрессивность рака с вероятным (иногда уверенным) отрицательным исходом. Соответственно, уровень 4 является промежуточным между 3 и 5. Кроме оценки полнослайдового изображения данный показатель может быть выделен по небольшим регионам ткани, что и было сделано в данном наборе данных и представлено на масках классов для изображений.

К настоящему времени уже был выполнен ряд работ по проблеме компьютеризированной диагностики и оценки степени злокачественности рака простаты [20-23]. Так, в работах [20, 21] с помощью методов глубокого обучения были получены хорошие результаты при решении задач бинарной классификации ткани простаты (рак против нормы, ранняя стадия рака против поздней). В то же время уровень определения конкретного показателя Глисона был значительно хуже при применении компьютеризированных систем не

только в качестве самостоятельного предсказателя ([20, 23], каппа 0,62 и 0,70 соответственно), но и в качестве вспомогательного модуля в процессе поддержки принятия решений гистопатологами ([22], каппа 0,733). При этом во всех рассмотренных работах количество полнослайдовых изображений (ПСИ), использованных для обучения нейронных сетей, было существенно меньше, чем в настоящем исследовании. Тем не менее среди них следует отметить работу [21], в которой исследование проводилось на значительном количестве ПСИ. Однако большинство изображений (11 429) использовалось все же только для тестирования.

10 616 ПСИ с соответствующими показателями в шкале Глисона и масками классов были получены из двух источников. Часть изображений была предоставлена Университетом Неймегена, Голландия (увеличение $\times 20$, пиксел 0,24 мкм, второй уровень с пикселом 0,48 мкм), вторая – Каролинским институтом, Швеция (увеличение $\times 20$, пиксел 0,48 мкм).

Как было указано ранее, для каждого ПСИ имелись маски, на которых указаны опухолевые области и их показатели Глисона. На основе данных масок, из этого набора данных были исключены 671 ПСИ, которые содержали артефакты или у которых информация на маске и показатели не соответствовали друг другу.

Маски классов, представленные для изображений с обоих источников, отличались по классам, представленных на них. Разметка из Университета Неймегена была детализированной и показывала разделение на 5 классов, а именно нормальная ткань, нормальный эпителий и раковый эпителий с 3-5 рейтингами Глисона. С другой стороны, данные из Каролинского Института имели только общие очертания опухолевых областей и не имели разбиения на разные степени агрессивности опухоли по шкале Глисона. Ввиду таких отличий между данными и большей сложностью решения задачи классификации участков изображений на отдельные рейтинги Глисона, использовались только данные из университета Неймегена. Примеры ПСИ и тайла представлены на рисунке 3.1.

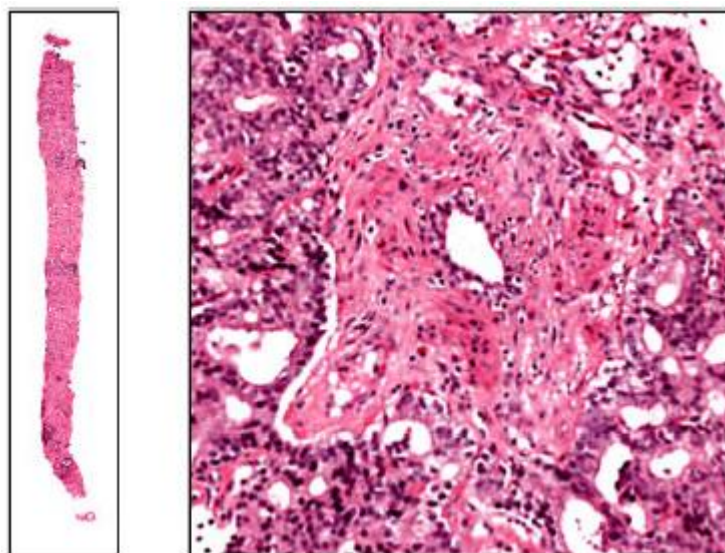


Рисунок 3.1 – Полнослайдовое изображение и тайл

Для проверки гипотез, поставленных в данной работе, использовались 5583 ПСИ из университета Неймегена. Данные изображения разбивались на квадратные регионы, именующиеся тайлами, которые использовались как входные данные для разрабатываемых алгоритмов. Тайлы имели разрешение 256x256 пикселей и выбирались с двух уровней увеличения изображения. С уровня максимального увеличения (x20, в дальнейшем – 0-ой уровень) было выбрано в общей сложности 994 087 тайлов. С следующего уровня увеличения (x5, в дальнейшем, 1-ый уровень) было получено 147 834 тайлов с учетом перекрытия в половину размера тайла, чтобы получить более полный набор данных.

При получении тайлов использовался ряд методов для избегания использования нерелевантных регионов таких как фон и артефакты при получении изображений. Данные методы описаны в следующей главе, посвященной алгоритмам предобработки изображений, как и детальное описание алгоритма получения тайлов, который позволил собирать наборы данных за разумное время.

Каждому тайлу был назначен класс в соответствии с максимальным классом, представленным на маске, соответствующей данному тайлу. Таким образом, поставлена задача классификации тайлов на основании типа ткани (норма, эпителий, опухоль с рейтингами Глисона от 3 до 5). Распределение по классам тайлов из обоих итоговых наборов данных представлено в таблице 3.1. Кроме того, стоит отметить, что выбор максимального класса не гарантировал, что на тайле не будут присутствовать другие классы или что большая часть тайла занята представленным классом.

Таблица 3.1 – Распределения тайлов по классам

Уровень	Общее кол-во	Норма	Эпителий	3 рейтинг Глисона	4 рейтинг Глисона	5 рейтинг Глисона
0(x40)	994 087	225 574 (22%)	172 400 (17%)	187 093 (18%)	322 477 (32%)	86 543 (8%)
1(x5)	147 834	3 215 (2%)	31 007 (20%)	28 268 (19%),	62 233 (42%),	23 111 (15%)

В дальнейшем, при проведении экспериментов с использованием признаков разных уровней для предсказания искомого класса использовалось меньшее количество данных. Это обусловлено тем, что для получения признаков для различных уровней изображения и построения соответствия между ними требовалось сопоставить между собой тайлы с разных уровней. Так как множество тайлов с 0-го уровня не имели соседних тайлов, то надежность результатов, получаемых для них могла быть довольно низкой, ввиду того, что большинство таких тайлов находилось на границе ткани и фона, что могло вызвать дополнительные ошибки в используемых алгоритмах. Таким образом общее количество тайлов 0-го уровня в экспериментах с объединением признаков составило 552 823.

ГЛАВА 4

АЛГОРИТМЫ ПРЕДОБРАБОТКИ ПОЛНОСЛАЙДОВЫХ ИЗОБРАЖЕНИЙ

Перед непосредственно анализом изображений требовалось определить регионы изображения, для выделения только тех тайлов, которые представляли интерес для рассматриваемой задачи классификации. Другой задачей, которую требовалось решить перед использованием алгоритмов глубокого обучения являлась задача нормализации изображений. Несмотря на то, что стандартные подходы к нормализации, используемые сегодня могут дать определённый результат, в гистологических изображениях существует высокая вариативность, вызванная различиями в методах подготовки образцов для сканирования и некоторыми другими факторами. Для компенсации эффектов, вызванных до получения изображения, использовался алгоритм нормализации цветового пространства.

4.1 Выделение ткани на изображении

Для выделения ткани на изображении существует ряд методов [24, 25], притом большинство из них основано на использовании порогов для выделения ткани. Несмотря на эффективность такого метода ему свойственно допускать ошибки в данных, для которых данный порог не совсем приспособлен. Кроме того, если речь идет об адаптивных порогах, то объём ткани на изображении может сильно варьироваться особенно если требуется обработка как изображений, полученных при анализе удаленной опухоли (ткань занимает большую часть изображения) так и результатов биопсии (тонкие полосы ткани на изображении).

В данной работе для отделения фона от интересующей части изображения использовалось два метода. Первый пороговый метод использовался для подготовки данных для обучения нейронных сетей. Суть этого метода заключается в том, что представить каждый пиксель как вектор в трехмерном пространстве RGB. Далее вычислялось расстояние между пикселями и «белым» цветом. Данное расстояние усреднялось по квадратным блокам, соответствующим тайлам, таким образом можно было получить количество фона на каждом тайле и, в зависимости от этого, принять решение использовать этот тайл в обучающей выборке или нет. Одним из плюсов данного алгоритма является возможность использовать верхние уровни изображения (низкое увеличение) для того, чтобы отделять фон от ткани. Например, на изображении с обычным разрешением порядка 100 000 пикселей на каждую сторону

использование 4-го уровня с шагом 4x между уровнями приведет к изображению с 6 250 пикселей на каждую сторону. И каждому тайлу 0-го уровня будет соответствовать область размером 4x4 пикселя.

С другой стороны, порог, использованный в данном примере, был подобран таким образом, чтобы исключить наибольшее количество пустых тайлов и исключить их влияние при обучении нейронных сетей. Поэтому при тестировании разработанных алгоритмов использовался другой метод, основанный на поиске небольших однообразных регионов на изображении и их последующим объединении.

В данном алгоритме сначала берется область полнослайдового изображения порядка нескольких мегапикселей, покрывающая все ПСИ, и к ней применяется слабый эффект размытия для того, чтобы уменьшить шум в выборе регионов, получаемых с помощью алгоритма заливки методом «наводнение». Такой способ содержит множество параметров, изменение которых позволяет регулировать количество областей. Эти параметры включают размер минимальной области, связность, максимально допустимая разница между интенсивностями пикселей в пределах одного региона и так далее. Кроме этого важно какой канал изображения используется, в данный момент, в основном, используется S канал HSV представления изображения, показывающий степень насыщенности цвета в каждом пикселе, что отлично подходит для поиска фона изображения. Алгоритм последовательно заливает регионы из свободных, не залитых точек изображений, при этом каждый пиксель может принадлежать только одному региону. Далее вычисляются различные дескрипторы для найденных областей. Дескрипторы включают в себя гистограммы каналов изображения в различных представлениях (RGB, HSV, разложения на маркеры), а также норму отличия от направления единичного вектора, например, в RGB это значение говорит о том, насколько серый цвет записан в данном пикселе. Для того, чтобы оптимизировать время, затрачиваемое на алгоритм кластеризации, вместо самого дескриптора используется результат его преобразования с помощью метода главных компонент. Это позволяет сократить количество элементов дескриптора до 16 и значительно ускорить вычисления. Применяя алгоритм K-средних для кластеризации полученных регионов с помощью вычисленных дескрипторов, получаются 3 группы регионов. Далее регионы одной группы объединяются и вычисляются средние значения S канала пикселей этих регионов. Эти значения используются для того, чтобы определить какие из этих трех регионов принадлежат фону, а какие нет. В итоге получен алгоритм, достаточно надежный для применения на изображениях, используемых в данной работе, включая покрашенные с помощью гематоксилина и эозина. Пример работы алгоритма представлен на рисунке 4.1.

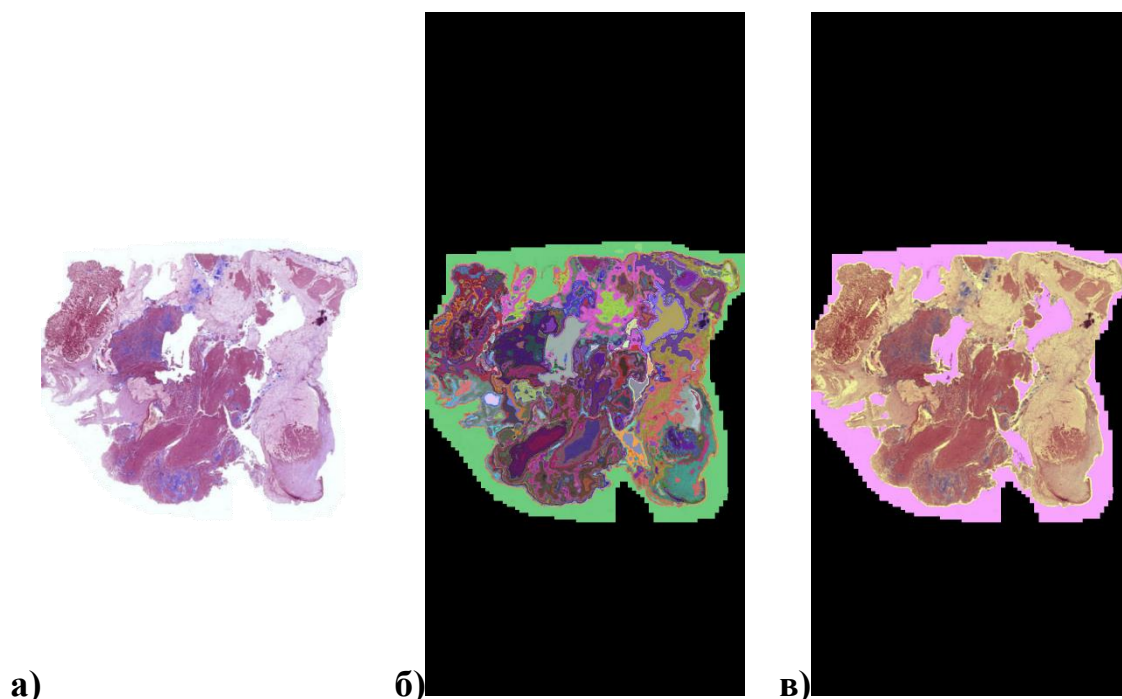


Рисунок 4.1 – а) Оригинальное изображение, б) найденные регионы, в) объединенные регионы с выделенной областью с образцом ткани

4.2 Нормализация цветового пространства

Нормализация цвета на гистологических полнослайдовых изображениях необходима в виду того, что изображения могли быть получены в различных условиях, таких, как отличающиеся характеристики сканеров для получения полнослайдовых изображений, различная степень освещенности, разный объем химических маркеров, использующихся для покраски образца ткани. Основным методом нормализации цвета на полнослайдовых изображениях является следующий обобщенный алгоритм:

- исключение неподходящих для определения цветовой схемы областей (например, фона на котором отсутствует образец ткани);
- перевод RGB пространства в другое пространство на основе преобладающих цветов;
- определение опорных векторов, линейная комбинация которых определяет данное пространство;
- обратное преобразование этих векторов в RGB пространство для того, чтобы получить основные цвета комбинации которых присутствуют на изображении (рисунок 4.2);
- перевод изображения в цвета, определенные выше (рисунок 4.3);

- замена основных цветов на эталонные заданные до начала работы алгоритма;
- обратное преобразование из значений концентрации основных цветов и RGB значений эталонных цветов в RGB пространство.



Рисунок 4.2 – Полученные цвета компонент области полнослайдового гистологического изображения (гематоксилин, эозин, остаточная компонента)

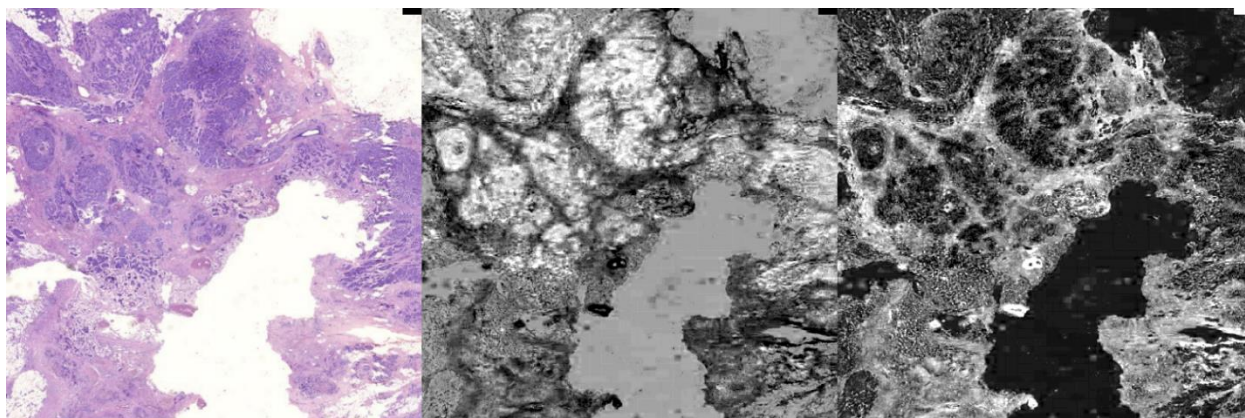


Рисунок 4.3 – Разложение области полнослайдового изображения на компоненты, соответствующие химическим маркерам. Слева направо: оригинальное изображение, компонент гематоксилина, компонент эозина

В данной работе было опробовано два подхода к нормализации цветового распределения, в обоих использовалось преобразование RGB значение пикселей изображения в значения оптической плотности по формуле [26]

$$OD = -\log_{10}(I)$$

где I – это интенсивность пикселя на черно-белом изображении.

На изображении не рассматривались пиксели с слишком низкой оптической плотностью, так как это сигнализирует о том, что рассматривается фон, который на рассматриваемых изображениях был белого цвета. В первом алгоритме используется SVD разложение, а после вектора нормализуются по длине векторов. Дальнейшие операции происходят с углами между векторами и направлениями SVD разложения [26]. Во втором алгоритме вместо SVD разложения используется ковариационная матрица RGB каналов изображения и

вектора получаются из собственных векторов этой матрицы [27]. Углы считаются как углы между векторами, определяемыми координатами в новом пространстве, и полученными собственными векторами. В качестве основных векторов брались 1-ый и 99-ый перцентили полученного распределения углов векторов, относительно найденных базисных векторов.

Таким образом цветовая схема всех изображений соответствует одинаковым базовым цветам химических маркеров. Далее применялись алгоритмы нормализации изображений, а именно эквализация гистограммы яркости изображения и удаление 1-го и 99-го перцентилей интенсивностей пикселей изображения, для того, чтобы частично избежать влияния различных артефактов изображений. Стоит отметить, что эти два алгоритма нормализации применялись к каналу Y в $YCbCr$ представлении изображений, так как нормализуется только контрастность изображения, а цветовая схема остается такой же.

Рассмотренные методы нормализации ПСИ позволяют уменьшить влияние вариативности цветового представления изображений на процесс обучения моделей глубокого обучения, а также алгоритм выделения ткани значительно уменьшает время, требуемое на анализ одного ПСИ, так как анализу подвергается только доля изображения, содержащая релевантную визуальную информацию.

ГЛАВА 5

ИСПОЛЬЗУЕМЫЕ МЕТОДЫ ГЛУБОКОГО ОБУЧЕНИЯ

В данной главе рассматриваются алгоритмы и методы, используемые в процессе обучения и тестирования нейронных сетей для каскадного анализа гистологических изображений.

В качестве основной модели использовалась нейронная сеть EfficientNet-B7 [28]. Данный выбор был обусловлен показателями точности данной модели в сравнении с другими для задачи классификации изображений. С другой стороны, имеются более эффективные модели и методы для классификации изображений основанные на EfficientNet [29], однако, ввиду ограничений вычислительных мощностей, результаты данных работ не использовались. Изменению подвергался только выходной слой нейронной сети. Для предсказания «уверенности» в том, что ткань на рассматриваемом тайле принадлежит одному классу, а также для использования признаков других сетей, в модель вводились дополнительные ветви, которые по своим характеристикам полностью соответствовали аналогичным частям оригинальной модели.

В качестве признаков, которые получались в процессе предсказания одной сетью, а использовались другой, брались признаки с последнего сверточного блока EfficientNet-B7. Для входного изображения разрешением 256x256 данные признаки имели размер 8x8x2560 (см. рисунок 5.1). В данной работе не использовались предобученные веса, за исключением полученных в ходе работы.

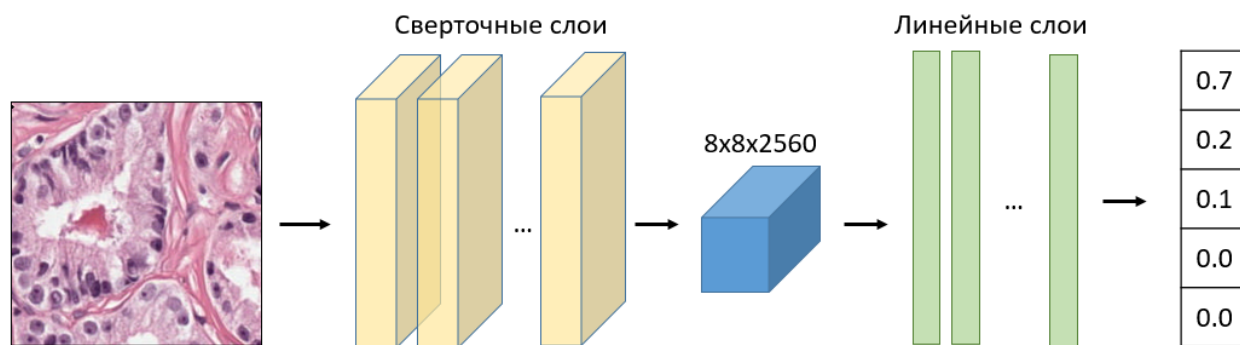


Рисунок 5.1 – Стандартная схема работы сверточной нейронной сети

Так как рассматривается задача классификации, в которой, несмотря на разбалансировку данных по классам, имеются тысячи изображений каждого класса, то использовалась перекрестная энтропия в качестве функции потерь. Однако, разбалансировка классов принималась во внимание при выборе метрик для оценки получаемых моделей. В качестве основной метрики в процессе

обучения модели использовалась Area-under-Curve (AUC), в дополнение к ней вычислялась F1 метрика. Для «уверенности» (см. главу 6) использовались такие же метрики, несмотря на более бинарный случай и более сбалансированное разбиение (46.4% тайлов содержали разные классы ткани). В финальном эксперименте по предсказанию полной потайловой карты классов, для всех имеющихся ПСИ, использовалась только F1 метрика, так как она лучше отражает финальную точность, на которую может рассчитывать конечный пользователь.

Для оптимизации функции потерь использовалась модифицированная версия распространённого оптимизатора Adam – AdamP [30]. Основное отличие заключается в том, что при использовании методов нормализации (Batch Normalization в данной работе), большая часть параметров становится инвариантной к сдвигу и масштабу. С другой стороны, в Adam момент постоянно увеличивает значения параметров из-за неучтенного эффекта нормализации. Поэтому, AdamP при использовании вместе с моделью, в которой присутствует нормализация, позволяет значительно ускорить обучение и уменьшить значения весов, что в итоге приводит к более высоким показателям тестовых метрик (рисунок 5.2). В процессе обучения нейронных сетей в данной работе использовалась скорость обучения 5×10^{-5} .

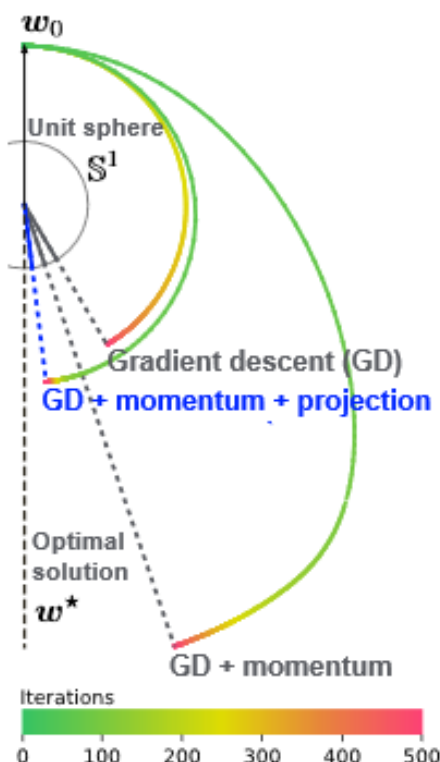


Рисунок 5.2 – Сравнение скорости процесса обучения с использованием градиентного спуска, градиентного спуска с моментом, и AdamP [30]

Для улучшения качества классификации также использовался ряд аугментаций, которые применялись к тренировочным данным на каждом шаге процесса обучения нейронной сети. Данные аугментации включали в себя гауссовский шум, повороты и отражения изображений и аффинное преобразование с небольшой амплитудой. Данные аугментации отражают возможные изменения, которые могут претерпевать изображения в процессе подготовки и сканирования. На рисунке 5.3 представлены примеры аугментированных изображений

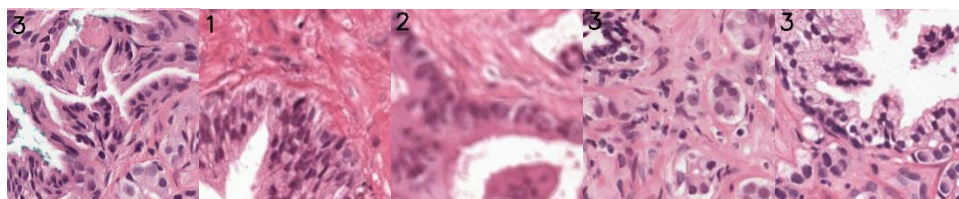


Рисунок 5.3 – Аугментированные тайлы

ГЛАВА 6

КАСКАДНЫЙ АНАЛИЗ

Для проверки гипотезы о том, что каскадный анализ полнослайдовых изображений улучшает как качество распознавания раковых регионов, так и вычислительную эффективность были применены следующие методы, которые в дальнейшем используются для сравнительного анализа. Первым шагом задача классификации решалась стандартным методом через прямое применение нейронных сетей. То есть на обоих наборах данных с 0-го и 1-го уровней тренировались нейронные сети, результаты работы которых стали отправной точкой для дальнейшей работы.

Далее был применен первый подход к каскадному анализу, а именно был построен набор данных из изначального набора 0-го уровня, в который были добавлены признаки извлеченные из натренированной сети 1-го уровня. Далее, используя веса сверточных слоев уже натренированной модели 0-го уровня, тренировалась сеть, входными данными которой являлись как тайлы 0-го уровня, так и признаки из сети 1-го уровня.

Следующим шагом была обучена нейронная сеть 1-го уровня аналогичная описанной выше с тем изменением, что кроме класса тайла она предсказывала «уверенность» в том, что на данном регионе центральные тайлы принадлежат одному и тому же классу. В дальнейшем предсказанная уверенность используется для того, чтобы во время тестирования алгоритмы проверять стоит ли использовать сеть 0-го уровня для уточнения предсказаний или нет.

Детальное описание и результаты работы данных алгоритмов на соответствующих им наборах данных представлены далее в подразделах данной главы, тогда как результаты тестирования для сравнительного анализа на уровне полнослайдовых изображений представлены в следующей главе.

6.1. Стандартный подход

Первые два полученных решения представляли собой прямое использование сети EfficientNet-B7 [28] для классификации тайлов. Данный эксперимент проводился на наборах данных 0-го и 1-го уровней (см. рисунок 6.1), чтобы была возможность сравнить качество классификации при детальном изображении и более «контекстным» 1-ым уровнем.

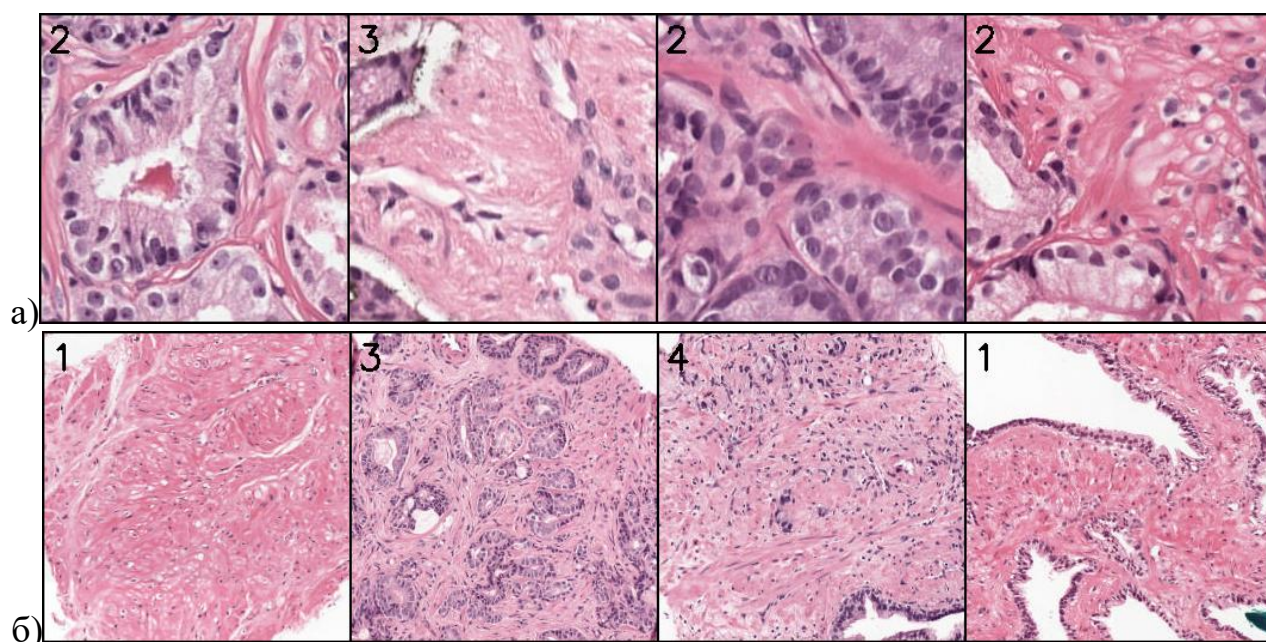


Рисунок 6.1 – Тренировочные примеры с а) 0-го уровня и б) 1-го уровня

Количество эпох в ходе которых обучались нейронные сети выбиралось исходя из того, что каждая из них должна была обучаться на примерно одинаковом количестве примеров, так как процесс сходимости являлся довольно гладким и обучение могло продолжаться довольно значительное время с небольшими улучшениями после основного процесса, который покрывался заданным количеством эпох. Таким образом, для 0-го уровня сеть обучалась 10 эпох, а на 1-ом уровне 70 эпох, что в обоих случаях составило около 10 миллионов примеров, которые видела нейронная сеть. В предыдущей главе описываются использованные метрики, оптимизатор и некоторые иные параметры. Кроме того, сравнение между моделями глубокого обучения проводится на основании общих результатов классификации, то есть по метрике точности, AUC и F1, чтобы избежать переобучения на определённую метрику и иметь большую уверенность в получаемых результатах сравнения.

Так как процесс обучения довольно гладкий (см. рисунок 6.2) и из-за значительного объёма данных различие в значениях метрик между тренировочным, валидационным и тестовым наборах в основном не превышало 0.005, то использовалось разбиение на валидационный и тестовый наборы данных вместо кросс-валидации. Для подтверждения в данной работе будут представлены значения как тестовых метрик, так и валидационных. Ещё стоит отметить, что разбиение на тренировочный, валидационный и тестовый набор проводилось с учетом того факта, что разные тайлы могли быть получены из одного изображения, то есть в разных наборах не могли присутствовать тайлы из одного изображения. Данное разбиение было фиксированным для каждого

набора данных и составляло 80% всех данных в тренировочном наборе и 10% данных в валидационном и тестовом наборах.

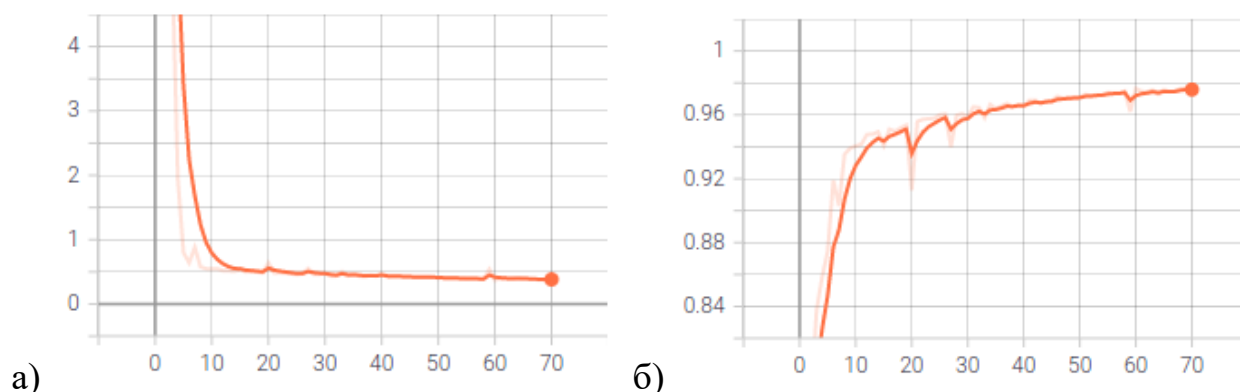


Рисунок 6.2 – Значения метрик в процессе обучения на валидационном наборе данных: а) функция потерь, б) средняя F1 метрика

В результате были получены начальные результаты для предсказания тайлов на 0-ом и 1-ом уровнях (в дальнейшем данные нейронные сети будут называться сеть 0-го уровня и сеть 1-го уровня). Результаты представлены в таблицах 6.1 и 6.2.

Таблица 6.1 – Значения метрик сети, обученной на 0-ом уровне

Метрика		Валидация	Тест
AUC	Норма	0.981	0.965
	Эпителий	0.947	0.949
	Глисон 3	0.935	0.884
	Глисон 4	0.927	0.928
	Глисон 5	0.946	0.949
	Среднее	0.947	0.935
F1	Норма	0.885	0.854
	Эпителий	0.725	0.728
	Глисон 3	0.708	0.627
	Глисон 4	0.768	0.779
	Глисон 5	0.649	0.652
	Среднее	0.746	0.753
Точность		0.767	0.773

Таблица 6.2 – Значения метрик сети, обученной на 1-ом уровне

Метрика		Валидация	Тест
AUC	Норма	0.986	0.986
	Эпителий	0.981	0.980
	Глисон 3	0.973	0.975
	Глисон 4	0.963	0.961
	Глисон 5	0.986	0.984
	Среднее	0.978	0.977
F1	Норма	0.694	0.682
	Эпителий	0.849	0.852
	Глисон 3	0.870	0.833
	Глисон 4	0.901	0.876
	Глисон 5	0.762	0.832
	Среднее	0.818	0.815
Точность		0.852	0.851

Как можно видеть, точность классификации, а также остальные метрики выше при использовании данных с 1-го уровня, однако стоит иметь ввиду, что эти результаты подразумевают, что в качестве входных данных используются тайлы 1-го уровня. Другими словами, можно получить указанную точность если нас не интересует более детальная классификация на 0-ом уровне, которая является тестовой задачей при тестировании методов в следующей главе. Дополнительно к таблицам 6.1 и 6.2 в приложении А представлены ROC-кривые для обоих начальных нейронных сетей.

Как видно из таблиц 6.1 и 6.2 существует явное различие между AUC метрикой, которая приближается к значению 1.0 и F1 с значением 0.753 для 0-го уровня и 0.815 для 1-го уровня. Исходя из предположения о том, что данный эффект мог быть вызван присутствием на одном тайле не только классов нормы и одного класса из всех остальных, но и других классов. Для проверки данного предположения была обучена нейронная сеть для классификации на несколько классов с использованием бинарной энтропии в качестве функции потерь на наборе тайлов 1-го уровня. Разбиение набора данных по классам с учетом того, что один тайл может содержать области разных классов представлен в таблице 6.3.

Таблица 6.3 – Разбиение по классам набора данных для мульти-классовой классификации

Уровень	Общее кол-во	Без класса	Эпителий	3 рейтинг Глисона	4 рейтинг Глисона	5 рейтинг Глисона
1(x5)	147 834	14735 (10%)	41369 (28%)	43971 (30%)	67943 (46%),	18549 (13%)

При подготовке считалось, что класс представлен на тайле если как минимум 0.02 площади тайла заняты этим классом. Полное отсутствие классов считалось нормой, поэтому в данной задаче имелись только 4 класса: Эпителий и три показателя Глисона 3-5. Стоит отметить, что даже на тайлах опухолевых областей большая часть была занята нормой, тогда как небольшие участки с опухолевыми клетками были выделены маской. Особенно это касается класса эпителия. На рисунке 6.3 представлено распределение площади занимаемой разными классами на тайлах, которым назначен тот или иной класс.

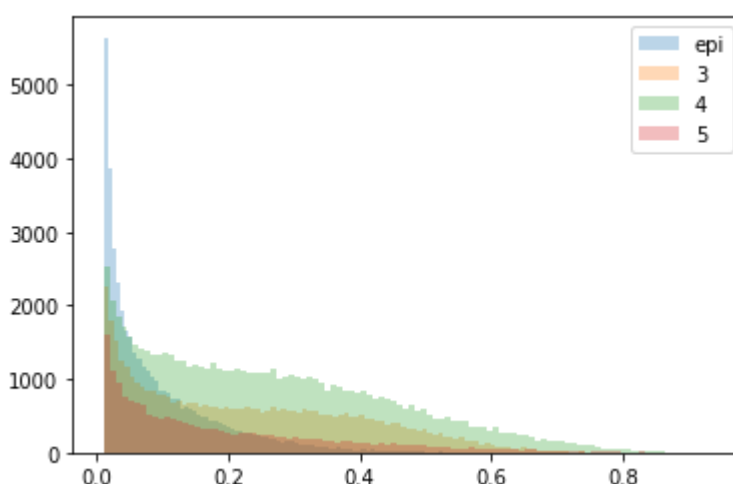


Рисунок 6.3 – Распределение доли классов на тайлах

Результаты представлены в таблице 6.4 в сравнении с результатами, полученными при обычной классификации из таблицы 6.2, а также ROC-кривая в приложении А. Как видно из таблицы 6.4 в основном классификация стала немногим лучше, так как AUC метрика не сильно возросла в сравнении с обычной классификацией на 1-ом уровне. Однако, F1 метрика сильно возросла за счет того, что ошибки, допускаемые сетью с простой схемой классификации в данном случае, означали, что требовалось предсказать два класса одному тайлу, с чем обученная нейронная сеть успешно справлялась. Учитывая, что качество классификации лучше при использовании мультиклассового подхода, можно убедиться в том, что данный метод лучше подходит для анализа тайлов, на которых может присутствовать несколько классов.

Таблица 6.4 – Результат мульти-классификации на 1-ом уровне в сравнении с обычной классификацией

Метрика		Валидация	Тест
AUC	Эпителий	0.983 (+0.002)	0.980 (0.000)
	Глисон 3	0.982 (+0.009)	0.980 (+0.005)
	Глисон 4	0.983 (+0.020)	0.982 (+0.021)
	Глисон 5	0.990 (+0.004)	0.990 (+0.006)
	Среднее	0.985 (+0.007)	0.983 (+0.006)
F1	Эпителий	0.890 (+0.041)	0.880 (+0.028)
	Глисон 3	0.893 (+0.023)	0.885 (+0.052)
	Глисон 4	0.925 (+0.024)	0.926 (+0.050)
	Глисон 5	0.854 (+0.092)	0.865 (+0.033)
	Среднее	0.890 (+0.072)	0.889 (+0.074)
Точность		0.942 (+0.090)	0.941 (+0.090)

Несмотря на полученное улучшение использование данного подхода имеет определенное ограничение, связанное с тем, что, в общем случае, достаточно детальная разметка данных для получения данного эффекта встречается довольно редко. Таким образом, в дальнейшем для анализа разрабатываемых подходов используется обычная классификация.

6.2. Подход с добавлением признаков

В следующем подходе к анализу тайлов ПСИ использовались результаты предыдущих двух нейронных сетей. Для получения каскадной схемы анализа тайлов в качестве основной нейронной сети использовалась сеть 0-го уровня, веса которой использовались для инициализации сверточной части в данном подходе.

Как указывалось, ранее после сверточной части EfficientNet-B7 с используемым разрешением тайлов получается массив выделенных признаков тайла размером 8x8x2560, после чего из каждого канала выбирается максимальный признак для получения вектора размером 2560, который в свою очередь обрабатывается рядом линейных слоев.

Для исключения влияния граничных эффектов и принимая во внимание то, что разница в разрешении между 0-ым и 1-ым уровнями составляла 4 раза (то есть один тайл 1-го уровня вмещал 16 тайлов 0-го уровня) брались только 4 центральных тайла. Если считать, что на 4x4 тайла приходились 8x8x2560 признаков, то получается, что одному тайлу 0-го уровня соответствовал массив признаков сети 1-го уровня размером 2x2x2560, которые и использовались для передачи информации о «контексте» вокруг тайла 0-го уровня.

Для добавления данных признаков в сеть 0-го уровня эти признаки сначала проходили через глобальную субдискретизацию по максимуму, а далее через несколько линейных слоев, чтобы подготовить их к объединению с признаками, выделенными сверточными слоями. Соответственно, после эти признаки соединились с признаками из сверточных слоев образуя вектор размером 5120, по которому уже предсказывался конечный класс тайла. Схема работы данной сети представлена на рисунке 6.4.

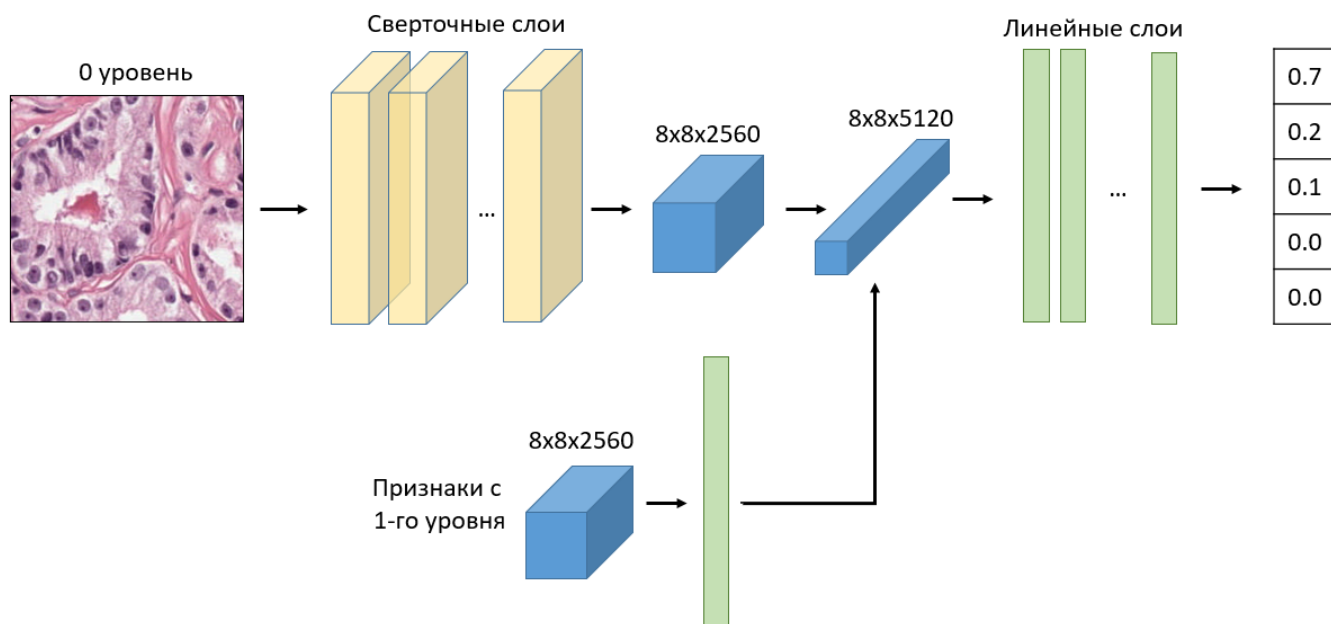


Рисунок 6.4 – Схема работы нейронной сети с использованием признаков верхнего уровня

В процессе подготовки данных использовались уже существующие наборы данных для 0-го и 1-х уровней, но для части тайлов 0-го уровня отсутствовали соответствующие им тайлы 1-го уровня, так они находились в окружении фона. По этой причине набор данных для обучения данной нейронной сети состоял из 552 823 с сохранением процентного разбиения по классам, как и в наборе тайлов 0-го уровня. Учитывая уменьшения количества данных обучение проходило в течении 20 эпох. Результаты представлены в таблице 6.5 и ROC-кривая, полученная на тестовом наборе данных – в приложении А.

Таблица 6.5 – Результаты обучения каскадной сети в сравнении с результатами на 0-ом уровне

Метрика		Валидация	Тест
AUC	Норма	0.987 (+0.006)	0.987 (+0.022)
	Эпителий	0.972 (+0.025)	0.970 (+0.021)
	Глисон 3	0.971 (+0.036)	0.970 (+0.086)
	Глисон 4	0.965 (+0.038)	0.965 (+0.037)
	Глисон 5	0.983 (+0.037)	0.983 (+0.034)
	Среднее	0.975 (+0.027)	0.975 (+0.040)
F1	Норма	0.894 (+0.009)	0.889 (+0.035)
	Эпителий	0.766 (+0.041)	0.765 (+0.037)
	Глисон 3	0.811 (+0.103)	0.815 (+0.188)
	Глисон 4	0.864 (+0.096)	0.863 (+0.084)
	Глисон 5	0.801 (+0.152)	0.800 (+0.148)
	Среднее	0.827 (+0.081)	0.826 (+0.073)
Точность		0.842 (+0.075)	0.841 (+0.068)

Как видно из таблицы 6.5 использование признаков сети 1-го уровня для передачи контекстной информации 0-му уровню привело к повышению точности классификации, до уровня чуть ниже показателей сети 1-го уровня. Таким образом показано, что использование «контекстной» информации при классификации тайлов 0-го уровня позволяет получить лучшие результаты, чем при использовании только тайла 0-го уровня.

6.3 Предсказание уверенности в классе центральных тайлов

Для построения каскадного алгоритма анализа изображений требовался алгоритм, определяющий стоит ли анализировать изображения на более детальном уровне или только что рассмотренного уровня вполне достаточно. Для достижения данной цели был применен следующий подход. Так как в качестве нижнего уровня использовались 4 центральных тайла, то в качестве меры «уверенности» в результатах 1-го уровня использовался тот факт, что все 4 центральных тайла 0-го уровня имеют один и тот же класс. Другими словами, требовалось предсказать значение уверенности 1.0 если предсказание точно определяет центральные тайлы иначе значение уверенности должно быть 0.0. При построении набора данных было обнаружено что 46.4% всех тайлов 1-го уровня содержат центральные тайлы 0-го уровня одинакового класса.

Уверенность предсказывается с помощью линейных слоев имеющих почти такую же конфигурацию, как и основная часть для классификации и использующих признаки, получаемые из сверточной части сети. Схема работы полученной нейронной сети представлена на рисунке 6.5. Для оценки точности

определения уверенности использовались те же метрики, что и для классификации: AUC и F1. Значения данных метрик представлены в таблице 6.6. При этом значения метрик для классификации не изменились или уменьшились для некоторых классов на значения не превышающие 0.01.

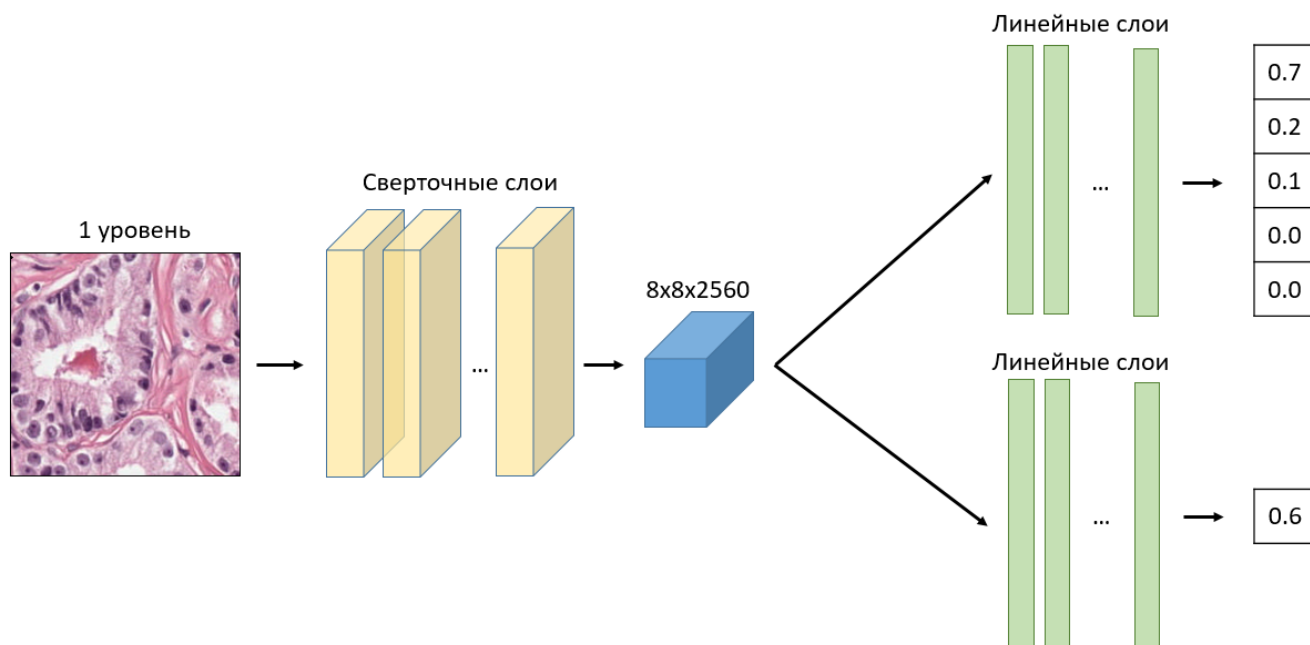


Рисунок 6.5 – Схема предсказания уверенности для тайлов 1-го уровня

Таблица 6.6 – Точность предсказания уверенности

Метрика		Валидация	Тест
Уверенность	AUC	0.914	0.912
	F1	0.761	0.760
	Точность	0.845	0.844

Так как при обучении данной нейронной сети не использовались веса, полученные на предыдущих этапах и, следовательно, в данной сети выделяемые признаки на сверточном слое отличались от признаков из сети 1-го уровня без предсказания уверенности, то требовалось переобучить сеть 0-го уровня с использованием признаков 1-го уровня полученных из сети с предсказанием уверенности. Интересно заметить, что признаки данной сети улучшили качество классификации на 0-ом уровне. Предположительно, добавление второй задачи, которую требовалась решать изначальной сети привело к получению более общих признаков, которые несли больше информации. Результаты представлены в таблице 6.7 в сравнении с результатами для сети 0-го уровня с использованием признаков. ROC-кривые полученные на тестовом наборе данных показаны в приложении А

Таблица 6.7 – Значение классификационных метрик для сети 0-го уровня с признаками из сети 1-го уровня предсказывающий класс и уверенность

Метрика		Валидация	Тест
AUC	Норма	0.989 (+0.002)	0.989 (+0.002)
	Эпителий	0.979 (+0.007)	0.978 (+0.007)
	Глисон 3	0.979 (+0.008)	0.977 (+0.007)
	Глисон 4	0.977 (+0.012)	0.976 (+0.011)
	Глисон 5	0.985 (+0.002)	0.986 (+0.003)
	Среднее	0.982 (+0.007)	0.981 (+0.006)
F1	Норма	0.904 (+0.010)	0.898 (+0.009)
	Эпителий	0.804 (+0.038)	0.805 (+0.040)
	Глисон 3	0.839 (+0.028)	0.832 (+0.017)
	Глисон 4	0.887 (+0.023)	0.885 (+0.022)
	Глисон 5	0.815 (+0.014)	0.819 (+0.019)
	Среднее	0.850 (+0.023)	0.848 (+0.022)
Точность		0.863 (+0.021)	0.860 (+0.019)

Отдельно, стоит отметить, что значения метрик, полученные в данном эксперименте, превышают значения метрик всех остальных нейронных сетей, используемых в данной работе, то есть точность сети выше чем у сети 1-го уровня и сети 0-го уровня. Данный факт показывает эффективность использования каскадного подхода к анализу тайлов 0-го уровня изображений.

ГЛАВА 7

РЕЗУЛЬТАТЫ

Учитывая рассуждения и результаты, показанные в предыдущей главе, было проведено дополнительное тестирование для определения эффективности каскадного подхода. Для этого применялась следующая методика тестирования.

Для всех 5583 ПСИ строилась маска на основе классов тайлов 0-го уровня (см. рисунок 7.1). Эффективность предсказания нейронных сетей основывалась на том, как точно они предсказывают класс данных тайлов. Отдельно стоит отметить, что для сети 1-го уровня также использовались тайлы 0-го уровня, таким образом появилась возможность оценить какую погрешность вносит использование меньшего увеличения. Кроме того, измерялось время, затрачиваемое на каждое изображение в целях сравнения эффективности подходов.

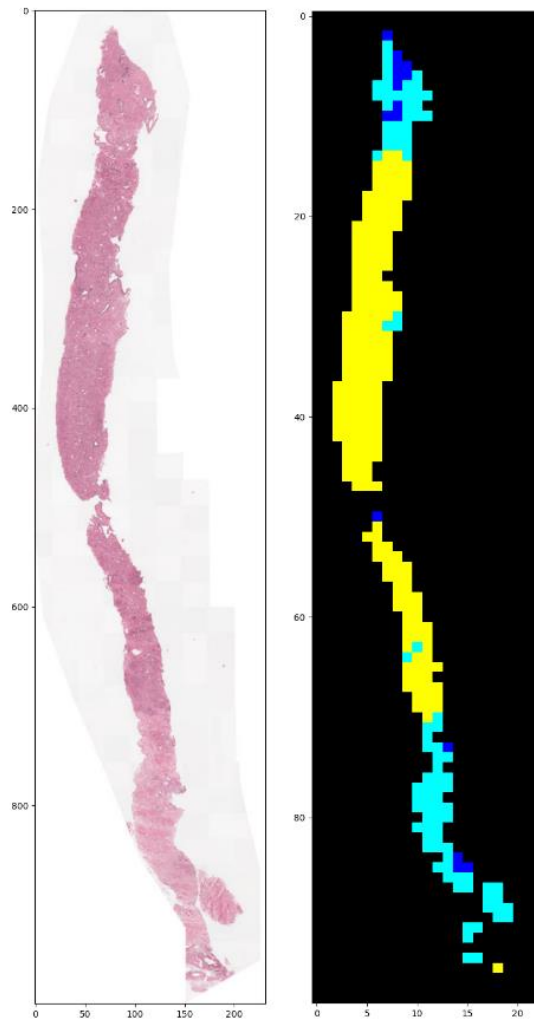


Рисунок 7.1 – Изображение и маска классов тайлов 0-го уровня

Применение сети 0-го уровня к тайлам того же уровня не требует дальнейших пояснений, но при применении сети 1-го уровня использовался сдвиг равный половине тайла, чтобы иметь возможность усреднять получаемые результаты. Также при выделении ткани на изображении также использовались тайлы 1-го уровня, что привносило определенную неточность в анализ изображений данным подходом.

При анализе изображений с помощью сети 0-го уровня с использованием признаков имелось выбирались тайлы 0-го уровня, которые должны были быть предсказаны, далее они группировались по группам рядом-стоящих тайлов, чтобы использовать сеть 1-го уровня для получения признаков. Таким образом получилось избежать проблем, связанных с грубым приближением контура ткани на изображении.

Наконец, алгоритм применение пары сетей с использованием предсказанной уверенности для определения стоит ли анализировать более детально или нет сначала использовал сеть 1-го уровня для классификации и получения координат и признаков для тайлов, которые должны быть обработаны сетью 0-го уровня. Учитывая двухшаговое предсказание, порог для определения тайла как содержащего ткань, а не фон, был уменьшен, так как данная сортировка позже могла быть проведена на 0-ом уровне, тогда как данное изменение позволило избежать некоторой части ошибок при определении тайлов, содержащих ткань, как фон.

Результаты, полученные с использованием представленной процедуры тестирования, представлены в таблице 7.1. В дополнение к этому в приложении Б показаны гистограммы распределения значений различных метрик по изображениям для каждого из представленных методов.

Таблица 7.1 – Результаты тестирования на всем наборе данных

Эксперимент		Уровень 1	Уровень 0	Уровень 0 (признаки)	Уровень 0 (признаки, уверенность)
Время, с		1.645	10.213	14.768	9.55
Точность		0.865	0.962	0.965	0.906
F1	Норма	0.963	0.998	0.997	0.971
	Эпителий	0.160	0.726	0.712	0.453
	Глисон 3	0.238	0.485	0.546	0.261
	Глисон 4	0.478	0.618	0.681	0.502
	Глисон 5	0.543	0.690	0.751	0.506
	Среднее	0.4764	0.7034	0.7374	0.5386

Как видно из полученных результатов наилучшие значения метрик получены при использовании признаков при предсказании классов нейронной сетью 0-го уровня с использованием признаков. С другой стороны, использование сети 1-го уровня позволяет получить результат за гораздо более короткий промежуток времени, сравнимый с масштабированием между обоими уровнями (1:16), но при этом результат будет значительно отставать от того, что возможно получить с использованием 0-го уровня. Подход с использованием уверенности для определения предсказывать класс на 0-ом уровне или оставить предсказание с 1-го уровня достиг компромиссного значения, получая результат выше чем на 1-ом уровне, но все еще недостаточный для конкуренции с полным анализом на 0-ом уровне.

ЗАКЛЮЧЕНИЕ

В данной работе рассмотрен метод каскадного анализа полнослайдовых изображений на примере изображений рака простаты. Данный метод подразумевает использование нескольких уровней увеличения изображения для получения более полной информации о рассматриваемом регионе. Метод основан на обучении нейронной сети с использованием регионов меньшего разрешения, чтобы в дальнейшем использовать признаки, извлекаемые из данной сети, вместе с регионами высокого увеличения, чтобы получать более точные предсказания. Показано, что такой подход работает лучше, чем обычная классификация регионов изображения. Кроме того, предложен способ по более эффективному анализу изображений в случае ограниченного времени или для получения приближенных значений. В таком случае, нейронная сеть низкого увеличения предсказывает некоторое значение, по которому можно определить необходимо ли проводить дальнейший детальный анализ с помощью другой нейронной сети. Данная работа показывает, что таким образом можно получить результаты лучше, чем при стандартном подходе и за время, меньшее, чем при предсказании на высоком разрешении, однако стоит иметь ввиду, что в таком случае точность предсказания значительно меньше точности полного анализа на высоком разрешении.

В дальнейшем могут быть рассмотрены другие способы оценки различий внутри регионов полнослайдовых изображений в целях улучшения определения случаев, когда требуется более детальный анализ региона. Также, более широкое тестирование с использованием других наборов гистологических данных может точнее показать преимущества использования рассматриваемых методов.

Результаты данной работы могут использоваться в широком спектре задач автоматического анализа полнослайдовых изображений для улучшение общего качества разрабатываемого программного обеспечения, направленного на предоставление дополнительной информации специалистам, работающим с данными изображениями.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

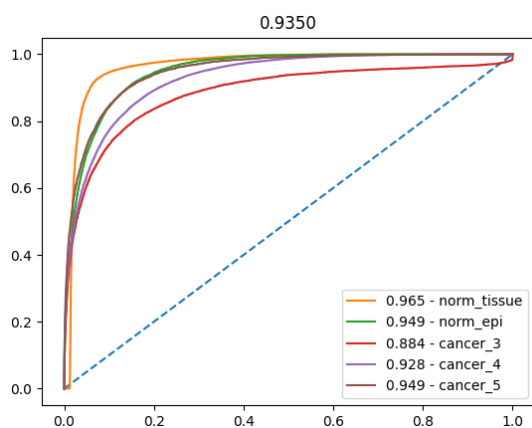
1. Biocompare, производство антител // [Электронный ресурс]. – 2019. Режим доступа: <https://www.biocompare.com/Antibody-Manufacturing/354958-CST-Immuno-fluorescence-Validated-Antibodies/>. Дата доступа: 2.12.2019.
2. Deep Learning for Whole Slide Image Analysis: An Overview / N. Dimitriou, O. Arandjelovic, P. D. Caie // *Frontiers in Medicine*. – 2019. – Vol. 6 – P. 264.
3. DICOM Whole Slide Imaging // [Электронный ресурс]. – 2019. Режим доступа: <http://dicom.nema.org/Dicom/DICOMWSI/>. Дата доступа: 26.02.2021.
4. Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks / A. Cruz-Roa [et al.] // *Proceedings of Medical Imaging 2014: Digital Pathology* – 2014.
5. Patch-based convolutional neural network for whole slide tissue image classification / L. Hou [et al.] // *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. – 2016. – P. 2424.
6. WSISA: Making survival prediction from whole slide histopathological images / X. Zhu [et al.] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* – 2017. – P. 7234-7242.
7. Using deep convolutional neural networks to identify and classify tumor-associated stroma in diagnostic breast biopsies / B. E. Bejnordi [et al.] // *Modern Pathology*. – 2018. – Vol. 31, iss. 10. – P. 1502-1512.
8. Prostate Cancer Grade Assessment (PANDA) Challenge // [Электронный ресурс]. – 2020. Режим доступа: <https://www.kaggle.com/c/prostate-cancer-grade-assessment>. Дата доступа: 18.02.2021.
9. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images / S. Graham [et al.] // *Medical Image Analysis*. – 2019. – Vol. 58. – P.101563.
10. Face detection with the faster R-CNN / H. Jiang, E. Learned-Miller // *In 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition* – May 2017. – P. 650-657.
11. Apple detection during different growth stages in orchards using the improved YOLO-V3 model / Y. Tian [et al.] // *Computers and electronics in agriculture*. – 2019. – Vol. 157. – P. 417-426.
12. Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks / J.W. Wei [et al.] // *Scientific reports* – 2019. – Vol. 9, iss.1. – P. 1-8.
13. Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images / B.E. Bejnordi // *Journal of Medical Imaging*. – 2017. – Vol. 4, iss. 4. – P. 044504.

14. CNN Cascades for Segmenting Whole Slide Images of the Kidney / M. Gadermayr [et al.] // arXiv preprint arXiv:1708.00251. – 2017.
15. Pyramid-based fully convolutional networks for cell segmentation / T. Zhao and Z. Yin // In International Conference on Medical Image Computing and Computer-Assisted Intervention. – September 2018. – P. 677-685.
16. U-net: Convolutional networks for biomedical image segmentation / O. Ronneberger, P. Fischer and T. Brox // In International Conference on Medical image computing and computer-assisted intervention. – October 2015. – P. 234-241.
17. Multi-scale fully convolutional neural networks for histopathology image segmentation: from nuclear aberrations to the global tissue architecture / R. Schmitz [et al.] // Medical image analysis. – 2021. – P. 101996.
18. Epidemiology of prostate cancer / P. Rawla // World Journal of Oncology. – 2019. – Vol. 10, iss. 2. – P. 63–89.
19. ISUP Grading Committee. The 2005 international society of urological pathology (ISUP) consensus conference on gleason grading of prostatic carcinoma. / J. I. Epstein [et al.] // The American Journal of Surgical Pathology. – 2005. – Vol. 29, iss. 9. – P. 1228–1242.
20. Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study / P. Ström. [et al.] // The Lancet Oncology. – 2020. – Vol. 21, iss. 2. – P. 222–232.
21. An artificial intelligence algorithm for prostate cancer diagnosis in whole slide images of core needle biopsies: a blinded clinical validation and deployment study / L. Pantanowitz [et al.] // The Lancet Digital Health. – 2020. – Vol. 2, iss. 8. – P. e407–e416.
22. Artificial intelligence assistance significantly improves Gleason grading of prostate biopsies by pathologists / W. Bulten [et al.] // Modern Pathology. – 2020.
23. Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer / K. Nagpal [et al.] // Nature Partner Journal Digital Medicine. – 2019. – Vol. 2, iss. 48. – P. 1–10.
24. A novel method for tissue segmentation in high-resolution H&E-stained histopathological whole-slide images / P. Kleczek, J. Jaworek-Korjakowska and M. Gorgon // Computerized Medical Imaging and Graphics – 2020. – Vol. 79. – P. 101686.
25. Resolution-agnostic tissue segmentation in whole-slide histopathology images with convolutional neural networks / P. Bándi [et al.] // PeerJ. – 2019. – Vol. 7. – P. e8242.
26. A method for normalizing histology slides for quantitative analysis / M. Macenko [et al.] // In IEEE International Symposium on Biomedical Imaging: From Nano to Macro – 2009. – P.1107-1110.

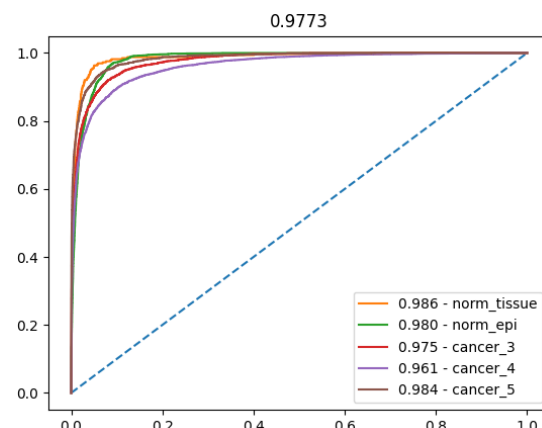
27. QuPath: Open source software for digital pathology image analysis / P. Bankhead [et al.] // Scientific Reports – 2017. – Vol. 7, iss. 1. – P. 16878.
28. Efficientnet: Rethinking model scaling for convolutional neural networks / M. Tan, Q. Le // In International Conference on Machine Learning. – May 2019. – P. 6105-6114.
29. Sharpness-Aware Minimization for Efficiently Improving Generalization / P. Foret [et al.] // arXiv preprint arXiv:2010.01412. – 2020.
30. AdamP: Slowing down the slowdown for momentum optimizers on scale-invariant weights / B. Heo [et al.] // arXiv: 2006.08217. – 2020.

ПРИЛОЖЕНИЕ А

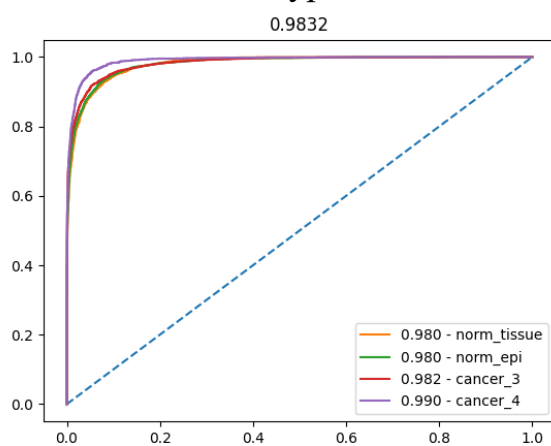
ROC-кривые обученных нейронных сетей



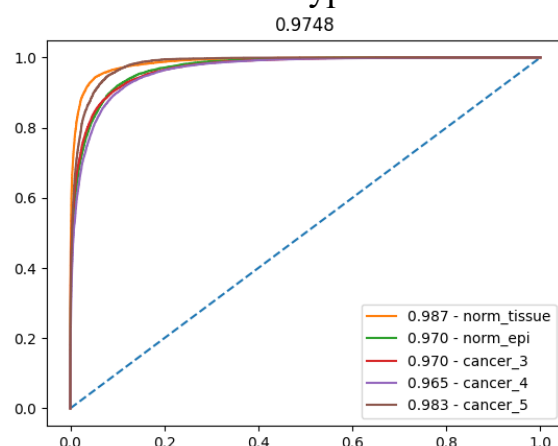
Сеть 0-го уровня



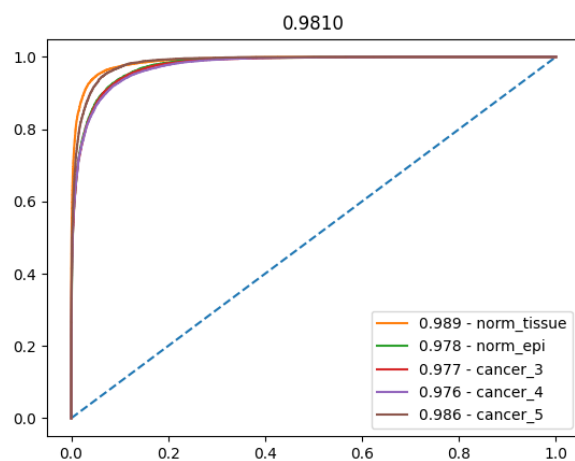
Сеть 1-го уровня



Мульти-классовой нейронная сеть



Сеть 0-го уровня с признаками из сети 1-го уровня



Сеть 0-го уровня с признаками из сети 1-го уровня с предсказанием уверенности

ПРИЛОЖЕНИЕ Б

Гистограммы распределения метрик при тестировании

