


ПРАВИТЕЛЬСТВО МОСКВЫ
ДЕПАРТАМЕНТ ЗДРАВООХРАНЕНИЯ ГОРОДА МОСКВЫ

СОГЛАСОВАНО

Главный внештатный специалист
по лучевой и инструментальной
диагностике
Департамента здравоохранения
города Москвы


Ю. А. Васильев
«11» декабря 2024 г.

РЕКОМЕНДОВАНО

Экспертным советом по науке
Департамента здравоохранения
города Москвы № 19



«11» декабря 2024 г.

РАДИОМИЧЕСКИЙ АНАЛИЗ
ПОСЛЕ ПОЛУЧЕНИЯ МАСКИ ОБЛАСТИ ИНТЕРЕСА:
РУКОВОДСТВО ДЛЯ ВРАЧА-РЕНТГЕНОЛОГА

Методические рекомендации № 80

Москва
2024

УДК 615.84+616-073.75

ББК 53.6

P15

Серия «Лучшие практики лучевой и инструментальной диагностики»

Основана в 2017 году

Организация-разработчик:

Государственное бюджетное учреждение здравоохранения города Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы»

Авторы-составители:

Васильев Ю. А. – канд. мед. наук, главный внештатный специалист по лучевой и инструментальной диагностике ДЗМ, директор ГБУЗ «НПКЦ ДиТ ДЗМ»

Нанова О. Г. – канд. биол. наук, ведущий научный сотрудник отдела научных медицинских исследований ГБУЗ «НПКЦ ДиТ ДЗМ»

Романенко М. О. – младший научный сотрудник отдела научных медицинских исследований ГБУЗ «НПКЦ ДиТ ДЗМ»

Гележе П. Б. – канд. мед. наук, старший научный сотрудник отдела научных медицинских исследований ГБУЗ «НПКЦ ДиТ ДЗМ»

Гончар А. П. – канд. мед. наук, научный сотрудник отдела научных медицинских исследований ГБУЗ «НПКЦ ДиТ ДЗМ», врач-рентгенолог отделения лучевой диагностики Онкологического центра № 1 ГБУЗ «ГКБ им. С. С. Юдина ДЗМ»

Мынко О. И. – инженер сектора исследований в лучевой диагностике ГБУЗ «НПКЦ ДиТ ДЗМ»

Блохин И. А. – канд. мед. наук, начальник сектора исследований в лучевой диагностике отдела научных медицинских исследований ГБУЗ «НПКЦ ДиТ ДЗМ»

Решетников Р. В. – канд. физ.-мат. наук, руководитель отдела научных медицинских исследований ГБУЗ «НПКЦ ДиТ ДЗМ»

P15 Радиомический анализ после получения маски области интереса: руководство для врача-рентгенолога: методические рекомендации / Ю. А. Васильев, О. Г. Нанова, М. О. Романенко [и др.] // Серия «Лучшие практики лучевой и инструментальной диагностики». – Вып. 145. – М.: ГБУЗ «НПКЦ ДиТ ДЗМ», 2024. – 68 с.

Рецензенты:

Нуднов Николай Васильевич – д-р мед. наук, профессор, заместитель директора по научной работе, заведующий НИО комплексной диагностики заболеваний и радиотерапии ФГБУ «РНЦРР» Минздрава России

Буренчев Дмитрий Владимирович – д-р мед. наук, профессор, заведующий отделением рентгенодиагностических и радионуклидных методов исследования ГБУЗ «ГКБ им. А. К. Ерамишанцева ДЗМ»

Методические рекомендации предназначены для использования в повседневной практике врачами-рентгенологами для извлечения радиомических признаков.

Данные методические рекомендации разработаны в ходе выполнения научно-исследовательской работы «Научное обоснование методов лучевой диагностики опухолевых заболеваний с использованием радиомического анализа» (№ ЕГИСУ: 123031500005-2) в соответствии с приказом от 21.12.2022 № 1196 «Об утверждении государственных заданий, финансовое обеспечение которых осуществляется за счет средств бюджета города Москвы, государственным бюджетным

(автономным) учреждениям, подведомственным Департаменту здравоохранения города
Москвы, на 2023 год и плановый период 2024 и 2025 годов»
Департамента здравоохранения города Москвы

*Данный документ является собственностью Департамента здравоохранения
города Москвы, не подлежит тиражированию и распространению
без соответствующего разрешения*

© Департамент здравоохранения города
Москвы, 2024

© ГБУЗ «НПКЦ ДиТ ДЗМ», 2024

© Коллектив авторов, 2024

СОДЕРЖАНИЕ

НОРМАТИВНЫЕ ССЫЛКИ	5
ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ.....	8
ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ	9
ВВЕДЕНИЕ	10
ИЗВЛЕЧЕНИЕ И АНАЛИЗ РАДИОМИЧЕСКИХ ПРИЗНАКОВ	13
Программы для разметки изображения.....	13
Программы для извлечения радиомических признаков	14
Общая характеристика радиомических признаков	16
Использование фильтров постобработки	18
Установка необходимого программного обеспечения	18
Загрузка изображений.....	23
Извлечение радиомических признаков.....	27
ЗАКЛЮЧЕНИЕ	38
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	39
ПРИЛОЖЕНИЕ А	43
ПРИЛОЖЕНИЕ Б.....	59

НОРМАТИВНЫЕ ССЫЛКИ

В настоящем документе использованы ссылки на следующие нормативные документы (стандарты):

1. Указ Президента Российской Федерации от 07.05.2018 № 204 «О национальных целях и стратегических задачах развития Российской Федерации на период до 2024 года».

2. Федеральный закон от 21.11.2011 № 323-ФЗ «Об основах охраны здоровья граждан в Российской Федерации».

3. Федеральный закон от 25.12.2018 № 489-ФЗ «О внесении изменений в статью 40 Федерального закона „Об обязательном медицинском страховании в Российской Федерации” и Федеральный закон „Об основах охраны здоровья граждан в Российской Федерации” по вопросам клинических рекомендаций».

4. Постановление Правительства Российской Федерации от 12.04.2018 № 447 «Об утверждении Правил взаимодействия иных информационных систем, предназначенных для сбора, хранения, обработки и предоставления информации, касающейся деятельности медицинских организаций и предоставляемых ими услуг, с информационными системами в сфере здравоохранения и медицинскими организациями».

5. Постановление Правительства Российской Федерации от 05.05.2018 № 555 «О единой государственной информационной системе в сфере здравоохранения».

6. Постановление Правительства Российской Федерации от 31.10.2018 № 1288 «Об организации проектной деятельности в Правительстве Российской Федерации».

7. Постановление Правительства Москвы от 27.12.2018 № 1703-1111 «О Территориальной программе государственных гарантий бесплатного оказания гражданам медицинской помощи в городе Москве на 2019 год и на плановый период 2020 и 2021 годов».

8. Приказ Министерства здравоохранения Российской Федерации от 14.09.2001 № 360 «Об утверждении перечня лучевых методов исследования».

9. Приказ Министерства здравоохранения и социального развития Российской Федерации от 05.05.2012 № 502н «Об утверждении порядка создания и деятельности врачебной комиссии медицинской организации» (в ред. от 02.12.2013).

10. Приказ Министерства здравоохранения Российской Федерации от 15.11.2012 № 915н «Об утверждении порядка медицинской помощи детям по профилю „Онкология”» (в редакциях).

11. Приказ Министерства здравоохранения Российской Федерации от 20.12.2012 № 1177н «Об утверждении порядка дачи информированного добровольного согласия на медицинское вмешательство и отказа от медицинского вмешательства в отношении определенных видов медицинских вмешательств, форм информированного добровольного согласия на медицинское вмешательство и форм отказа от медицинского вмешательства» (в ред. от 10.08.2015 и 17.07.2019).

12. Приказ Министерства здравоохранения Российской Федерации от 13.10.2017 № 804н «Об утверждении номенклатуры медицинских услуг» (в ред. от 16.04.2019 и 05.03.2020).

13. Приказ Министерства здравоохранения Российской Федерации от 30.11.2017 № 965н «Об утверждении порядка организации и оказания медицинской помощи с применением телемедицинских технологий».

14. Приказ Министерства здравоохранения Российской Федерации от 04.05.2018 № 201н «Об утверждении показателей, характеризующих общие критерии оценки качества условий оказания услуг медицинскими организациями, в отношении которых проводится независимая оценка».

15. Приказ Министерства здравоохранения Российской Федерации от 25.03.2019 №158н «О внесении изменений в Правила проведения

патологоанатомических исследований, утвержденные приказом Министерства здравоохранения от 24.03.2016 № 179н».

16. Приказ Министерства здравоохранения Российской Федерации от 07.06.2019 № 381н «Об утверждении Требований к организации и проведению внутреннего контроля качества и безопасности медицинской деятельности».

17. Приказ Департамента здравоохранения города Москвы от 15.03.2018 № 183 «Об утверждении регламента организации оказания медицинской помощи по профилям „Рентгенология” и „Радиология” с применением телемедицинских технологий».

ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ

В настоящем документе применены следующие термины с соответствующими определениями:

Воксель – наименьший логический элемент трехмерного цифрового изображения в растровой графике. Воксели являются аналогами двумерных пикселей для трехмерного пространства.

Область интереса (Region of interest, ROI) – определенная область изображения или набора данных, которая является объектом анализа.

Пиксель (англ. pixel – сокращение от pictures element) – наименьший логический элемент двумерного цифрового изображения в растровой графике. Пиксель представляет собой неделимый объект прямоугольной или круглой формы, характеризуемый определенным цветом. Растровое компьютерное изображение состоит из пикселей, расположенных по строкам и столбцам.

Радиомика – это метод, который извлекает большое количество признаков из медицинских изображений с использованием специализированных алгоритмов характеристики данных. Метод лежит на стыке лучевой диагностики, компьютерных наук и математической статистики. В качестве синонима может использоваться термин «текстурный анализ».

Радиомический признак – количественная характеристика изображения, получаемая в результате радиомического анализа.

Уровни серого (оттенки серого, градации серого) – цветовой режим изображений, реализованный в виде таблицы (шкала серого, англ. Grayscale) с указанием яркости белого цвета для каждой точки.

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

ГК – главная компонента

GLCM – Gray Level Co-occurrence Matrix (матрица совпадения уровней серого)

GLDM – Gray Level Dependence Matrix (матрица зависимости уровней серого)

GLRLM – Gray Level Run Length Matrix (матрица длин серий уровней серого)

GLSZM – Gray Level Size Zone Matrix (матрица размера зон уровня серого)

NGTDM – Neighbouring Gray Tone Difference Matrix (матрица различий соседних оттенков серого)

ROI – Region of interest (область интереса)

RMSE – root-mean-square error (среднеквадратичная ошибка)

ВВЕДЕНИЕ

Радиомика – перспективный метод современной медицины [1], направленный на повышение качества диагностики и планирования терапии с использованием продвинутых методик анализа медицинских изображений. Радиомический анализ проводится на основе извлечения и обработки параметров медицинских изображений, невидимых человеческому глазу – радиомических признаков [2]. Радиомика позволяет количественно анализировать текстурную информацию, получаемую посредством математического извлечения пространственного распределения интенсивностей сигналов и взаимосвязей пикселей. Кроме того, с помощью радиомики можно количественно оценить визуально заметные различия в интенсивности, форме или текстуре изображения, что позволяет уменьшить субъективный характер интерпретации медицинских изображений (компьютерной томографии, магнитно-резонансной томографии, ультразвукового исследования, позитронно-эмиссионной томографии).

Перспектива метода – использование его как вспомогательного инструмента для прогнозирования и принятия решений по тактике лечения [3, 4]. Наиболее широкое применение радиомика нашла в онкологии [5].

Радиомика не предполагает автоматизации диагностических процессов, а обеспечивает получение дополнительных данных. Радиомический анализ может быть полезен для разных медицинских задач – диагностики, стадирования, оценки ответа на терапию. В последние годы активно проводятся исследования по взаимосвязи радиомических признаков с экспрессией генов (радиогеномика [6, 7]) и иммуногистохимическими признаками [8, 9], что в будущем поможет улучшить прогнозы ответа на лечение, основанные на радиомике. Дельта-радиомика [10] вносит временной компонент и включает в себя извлечение количественных признаков из наборов изображений, полученных в ходе

курса лечения, в котором представлена информация об эволюции значений признаков.

Радиомический анализ имеет ряд ограничений. Основным из них является отсутствие стандартизированных методик и, как следствие, низкая воспроизводимость результатов [11, 12]. Фактически в каждом исследовании отбираются свои радиомические признаки и строится модель. Невоспроизводимость значимых радиомических признаков между исследованиями связана с отсутствием стандартизации в получении и обработке изображений, с различными способами получения и отбора признаков. В настоящий момент делаются попытки преодолеть это ограничение за счет создания стандартизированной базы радиомических признаков. Например, одно из новейших решений – проект IBSI [13].

Основные этапы текстурного анализа включают в себя получение медицинских изображений, сегментацию области интереса на полученных медицинских изображениях, извлечение радиомических признаков, создание прогностической модели на основе извлеченных и отобранных радиомических признаков, валидацию модели на внутренних и внешних данных, оценку предсказательной силы модели и дальнейшее ее совершенствование с использованием специально разработанных для этого инструментов. Специализированные навыки врачей-рентгенологов требуются на этапах получения изображений, их разметки и извлечения радиомических признаков. На этапах создания и тестирования прогностической модели необходимы специализированные навыки медицинских статистиков. Тем не менее мы рекомендуем врачам-рентгенологам – конечным пользователям разработанных моделей – участвовать во всех этапах создания прогностических моделей для облегчения интерпретации получаемых результатов.

Целью настоящих рекомендаций является описание основных этапов и изложение ключевых понятий радиомического анализа. Для этого дана подробная характеристика радиомических признаков, пошагово описана

процедура их извлечения и приведен пример статистического анализа полученных данных (приложение А).

Несмотря на обилие литературы в области радиомики, данная работа является первой методической разработкой по радиомическому анализу для русскоязычной аудитории. Мы предполагаем, что наши рекомендации помогут внедрить этот новейший и перспективный метод в рутинную практику врачей-рентгенологов.

ИЗВЛЕЧЕНИЕ И АНАЛИЗ РАДИОМИЧЕСКИХ ПРИЗНАКОВ

Программы для разметки изображения

Перед началом радиомического анализа необходимо произвести разметку изображений – выделить интересующие области. Для ознакомления с современными программами разметки медицинских изображений рекомендуем статью «Обзор современных средств разметки цифровых диагностических изображений» [14]. Также для удобства здесь мы приводим сводную информацию об общих характеристиках современных программ для разметки цифровых диагностических изображений (таблица 1).

Таблица 1 – Общие характеристики, работа с файлами в программном обеспечении для аннотации цифровых диагностических изображений [14]

Функция/характеристика		3D Slicer	ITK-SNAP	MITK	Medseg	Supervisely	CVAT
Общие характеристики							
Open source		да	да	да	нет	нет	да
Веб-приложение		нет	нет	нет	да	да	да
Возможность установить локально		да	да	да	нет	в расширенной версии	да
Добавление пользовательских модулей		да	да	да	нет	да	да
Сложность установки		низкая	низкая	низкая	—	—	высокая
Системные требования	RAM (минимум/рекомендовано), количество ядер	4/8 Gb	2/4 Gb	—	—	8/32 Gb CPU 4/8 ядер	—
	ОС (операционная система)	Windows, macOS, Linux	Windows, macOS, Linux	Windows, macOS, Linux	—	Linux	Windows, macOS, Linux
	Разрешение экрана	1024×768/ 1280×1024	1024×768/ 1280×1024	—	—	—	—
	Дополнительно (рекомендовано)	наличие GPU	наличие GPU	—	Google Chrome, наличие GPU	Диск 200 Gb SSD наличие GPU	наличие GPU

Функция/характеристика	3D Slicer	ITK-SNAP	MITK	Medseg	Supervisely	CVAT
Поддержка форматов медицинских изображений, загрузка файлов						
Поддержка основных форматов медицинских изображений	+	+	+	–	–	–
Возможность загружать несколько изображений и сегментаций	+	+	+	–	+/-	–
Одновременная работа с несколькими сегментациями для одного изображения	+	+/-	+	–	–	–
<i>Примечание: «—» — неприменимо или отсутствует информация; «+» — наличие данной функции в программном обеспечении; «-» — отсутствие данной функции; «+/-» — функция реализована частично.</i>						

Таким образом, из программного обеспечения, имеющегося в свободном доступе, 3D Slicer обладает наиболее полным функционалом для разметки изображений. Центром диагностики и телемедицины подготовлены методические рекомендации по основному функционалу 3D Slicer для разметки медицинских изображений на примере сегментации артериальных сосудов [15].

Программы для извлечения радиомических признаков

Мы провели сравнительный анализ (таблица 2) популярных бесплатных программ для извлечения радиомических признаков: PyRadiomics [16], LIFEx [17], MaZda [18], IBEX [19].

Таблица 2 – Общие характеристики программ для извлечения радиомических признаков

Характеристика	PyRadiomics	LIFEx	MaZda	IBEX
Бесплатная программа	да	да	да	Да
Открытый доступ	да	нет (требуется регистрация через почту)	нет (требуется подписание лицензионного соглашения онлайн)	нет (требуется аккаунт на Dropbox)
Дата последнего обновления / периодичность обновлений	2023 (1 раз в 1-2 года)	2024 (1-2 раза в год)	2009 (не обновляется)	2017 (не обновляется)

Характеристика	PyRadiomics	LIFEx	MaZda	IBEX
Популярность: число цитирований в Google Академия с момента публикации	4157	794	743	373
Общее число рассчитываемых радиомических признаков	107	44	352 (для 2D), 892 (для 3D)	91
Основные группы рассчитываемых радиомических признаков	Признаки первого порядка, второго порядка, признаки формы			
ЗадOCUMENTИРОВАННОСТЬ математических уравнений для радиомических признаков	да	да	нет	нет
Согласованность радиомических признаков с IBSI-платформой	да	да	нет	нет
Число градаций уровней серого	200	400	–	256
Поддержка основных форматов медицинских изображений	NIFTI-1, JPG, TIFF, PNG, BMP	DICOM, а также NIFTI-1, JPG, TIFF, PNG, BMP	DICOM	DICOM, BMP
Возможность установить локально	Да			
Язык	Python	Java	C++	Matlab
Возможность работы с 3D Slicer	да (расширение SlicerRadiomics)	нет	нет	нет
Операционная система	Windows, MacOS, Linux	Windows, MacOS, Linux	Windows	Windows

PyRadiomics – наиболее популярная для решения этой задачи программа. В отличие от других бесплатных программ она находится в свободном доступе и не требует регистрации или подписания лицензионного соглашения.

Несомненным плюсом PyRadiomics является расчет радиомических признаков, согласованных с инициативой IBSI [20], что важно для их стандартизации, а также работа с широким спектром форматов медицинских изображений [21, 22, 23].

PyRadiomics может быть рекомендован как наиболее предпочтительная программа для извлечения радиомических признаков.

PyRadiomics возможно использовать в двух вариантах: через консоль и графический интерфейс. В первом случае он работает как пакет Python (версии 2.7, 3.4 и 3.5), во втором – распространяется в виде расширения для 3D Slicer, имеющего название SlicerRadiomics. В этом варианте пользователям предоставляется возможность работать с интуитивно понятным графическим интерфейсом. Именно этот наиболее удобный для широкого круга пользователей вариант мы рассмотрим в настоящих методических рекомендациях.

Рекомендуемые параметры операционной системы для 3D Slicer:

- Windows 10 или 11 с установленными обновлениями;
- MacOS Big Sur 11 или более поздние версии;
- Linux Ubuntu 18.04 или более поздние версии.

Общая характеристика радиомических признаков

В этом разделе мы приводим краткие характеристики основных групп радиомических признаков, извлекаемых с помощью PyRadiomics, для того чтобы исследователь имел представление, с какими характеристиками изображений он работает. Традиционно выделяют три основные группы радиомических признаков: признаки первого порядка, признаки второго порядка и признаки формы. Число и состав извлекаемых признаков в этих группах может варьировать в разных программах [24].

При необходимости с подробным англоязычным описанием каждого признака с формулами расчета можно ознакомиться на сайте PyRadiomics (<https://pyradiomics.readthedocs.io/en/latest/features.html>). Более детальное описание отдельных признаков также приведено в приложении Б настоящего руководства.

Текстурный анализ использует математические методы для оценки интенсивности серого цвета и расположения пикселей в изображении.

Анализ радиомических **признаков первого порядка**, иначе известный как гистограммный анализ, извлекает значения интенсивности пикселей в интересующей области, которые затем отображаются графически (рисунок 1).

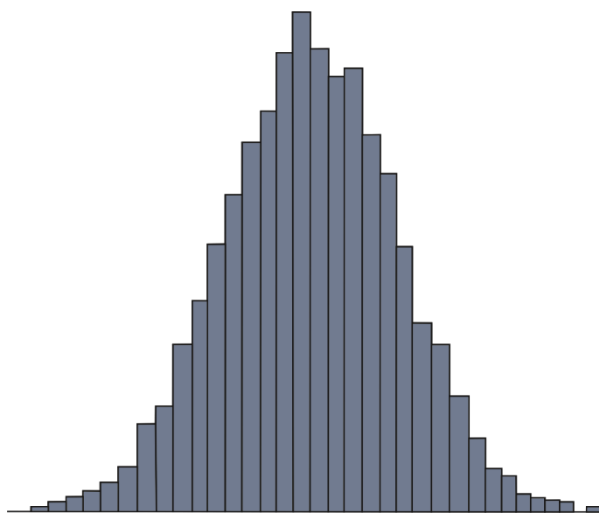


Рисунок 1 – Гистограмма распределения интенсивностей серого, признаки первого порядка

Упрощенный текстурный анализ включает в себя первоначальную настройку изображения путем применения к нему различных фильтров, которые позволяют извлекать и количественно определять невидимые невооруженным глазом особенности изображения с точки зрения неравномерности и яркости. На основе гистограммы вычисляют метрики, включающие равномерность, разброс, симметрию и случайность значений интенсивности пикселей в пределах области интереса.

Наиболее распространенными характеристиками гистограммы, которые приводятся в опубликованных источниках, являются среднее (mean), стандартное отклонение (standard deviation), асимметрия (skewness), куртозис (kurtosis), энтропия (entropy) и энергия (energy).

Более сложный радиомический анализ особенностей изображения исследует связи между пикселями в пределах области интереса. Показатели, полученные в ходе такого анализа, называют **признаками второго порядка**. Оценивая их, можно получить дополнительную информацию

об изменчивости интенсивности сигнала пикселей в более сглаженных, однородных областях, имеющих меньшую текстурную изменчивость, или же более разнородных областях, имеющих бóльшую текстурную изменчивость.

Наконец, **признаки формы** зависят не от интенсивности серого, а от геометрических особенностей области интереса.

Использование фильтров постобработки

Радиомические признаки можно извлекать как из оригинальных изображений, так и из обработанных с помощью фильтров. PyRadiomics позволяет использовать вейвлет-преобразование (wavelet-based features). Эти фильтры постобработки применяются для увеличения числа извлекаемых текстурных показателей и удаления шумов из изображений.

Вейвлет использует набор специальных фильтров, называемых квадратурными зеркальными фильтрами, для рекурсивного разложения исходного изображения на низкочастотные и высокочастотные поддиапазоны вдоль осей x , y и z . Всего из одного изображения можно получить восемь новых изображений того же размера, что и исходное.

Установка необходимого программного обеспечения

Для установки 3D Slicer требуются права администратора на компьютере. Так как расширение SlicerRadiomics регулярно обновляется, для работы с ним рекомендовано использование последней версии (nightly release) 3D Slicer. Загрузите последнюю версию 3D Slicer для своей операционной системы с сайта <http://download.slicer.org> и установите ее (рисунки 2, 3).

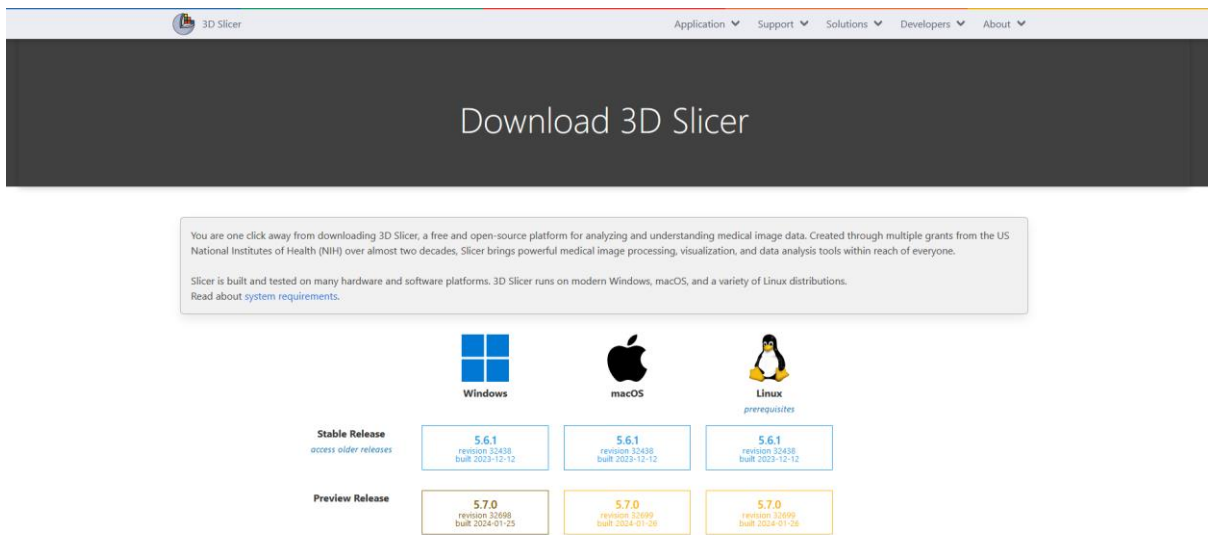


Рисунок 2 – Панель для загрузки 3D Slicer

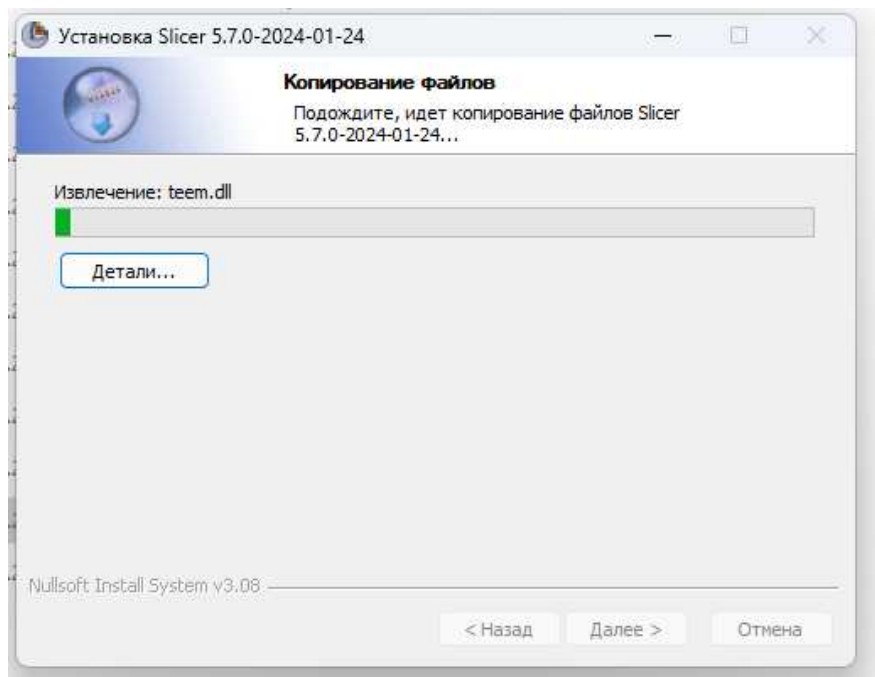



Рисунок 3 – Процесс установки 3D Slicer

3D Slicer, помимо своих базовых функций, использует расширения – дополнительные программы со специфическими функциями, требующие установки. Всего в 3D Slicer на данный момент доступно 179 разных расширений.

Для установки интересующего расширения используется менеджер расширений Extension Manager (рисунок 4). Нажатием кнопки  открывается список расширений (рисунок 5).

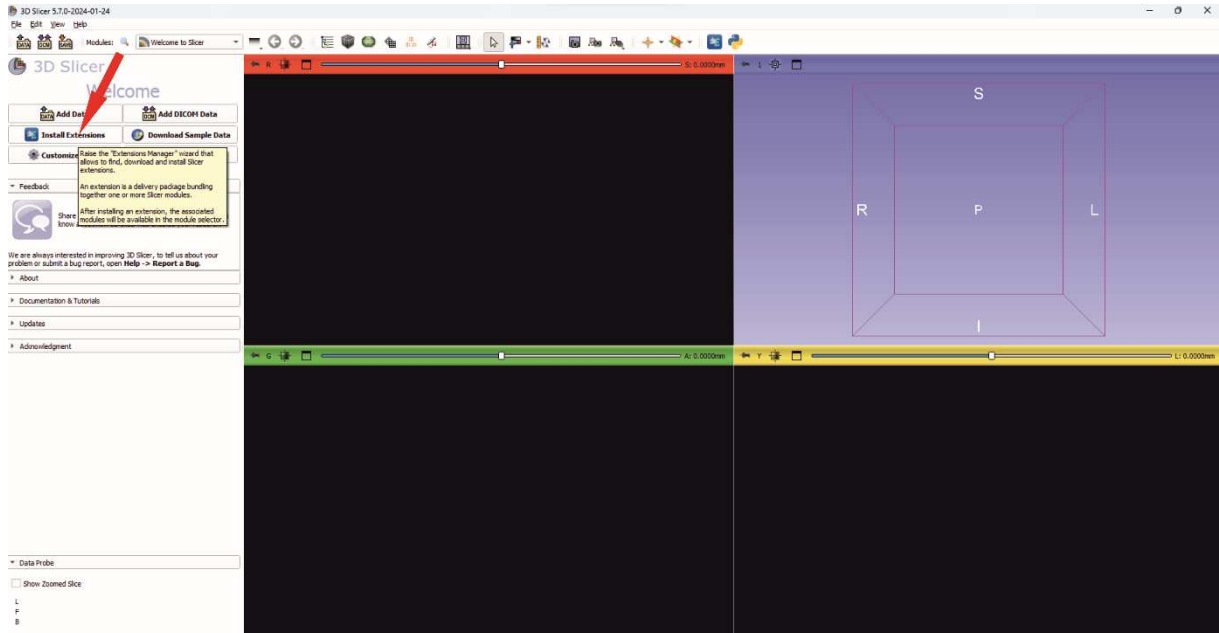


Рисунок 4 – Процесс открытия Extension Manager. Красной стрелкой указано расположение кнопки Install Extensions

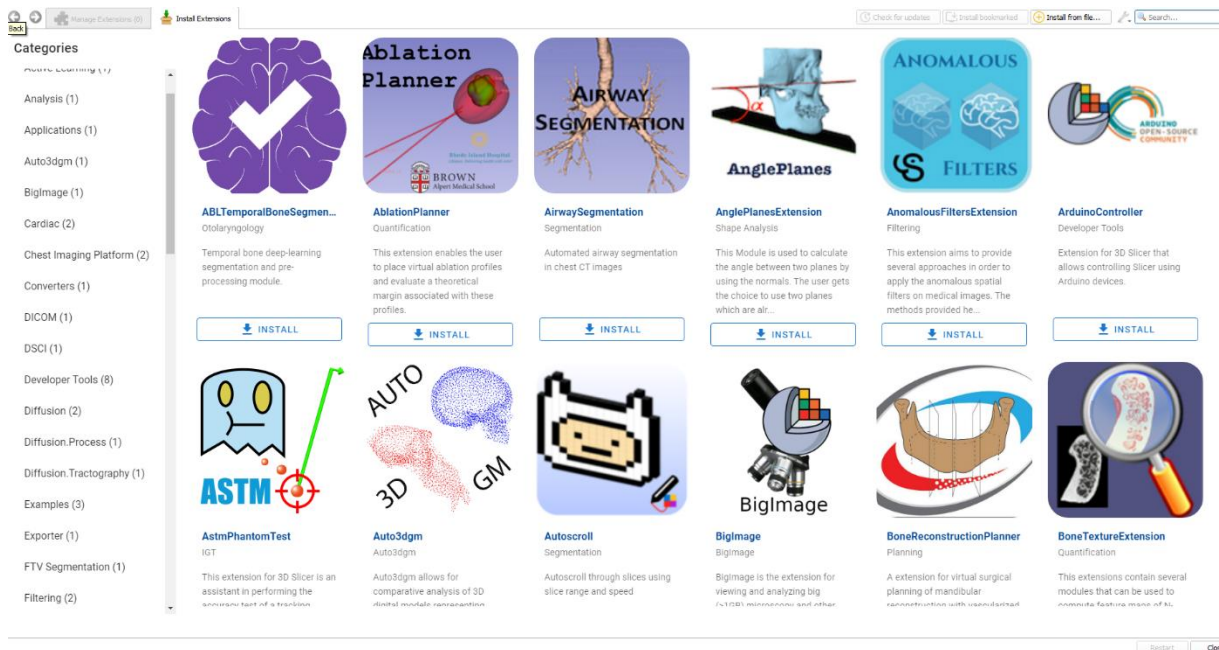



Рисунок 5 – Общий вид списка расширений

В нашем случае понадобится установить два расширения.

Расширение SlicerRadiomics

Нужно открыть менеджер расширений Extension Manager и развернуть список нажатием кнопки Install Extensions. В списке нужно найти расширение SlicerRadiomics (рисунок 6) и установить его, нажав кнопку . Это расширение располагается в папке Informatics, либо его можно найти через поисковую строку в правом верхнем углу. После установки нужно перезагрузить 3D Slicer кнопкой Restart в правом нижнем углу.

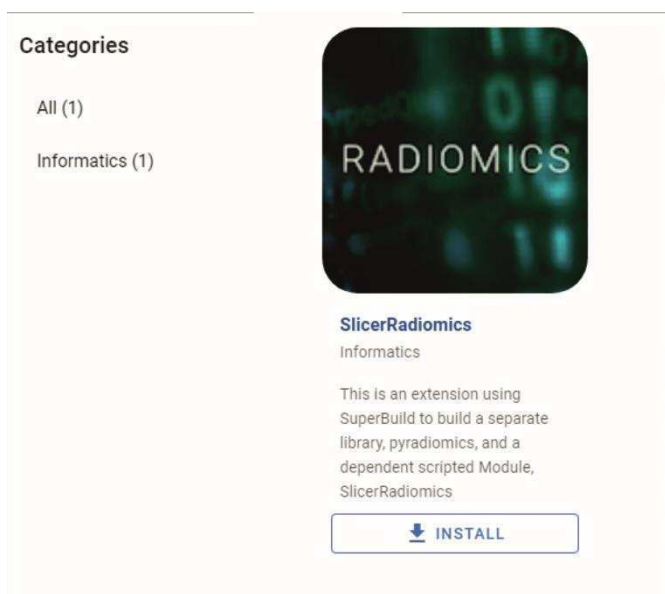


Рисунок 6 – Вид иконки расширения SlicerRadiomics

Если установка прошла успешно, то расширение появится в списке установленных расширений 3D Slicer (рисунок 7).

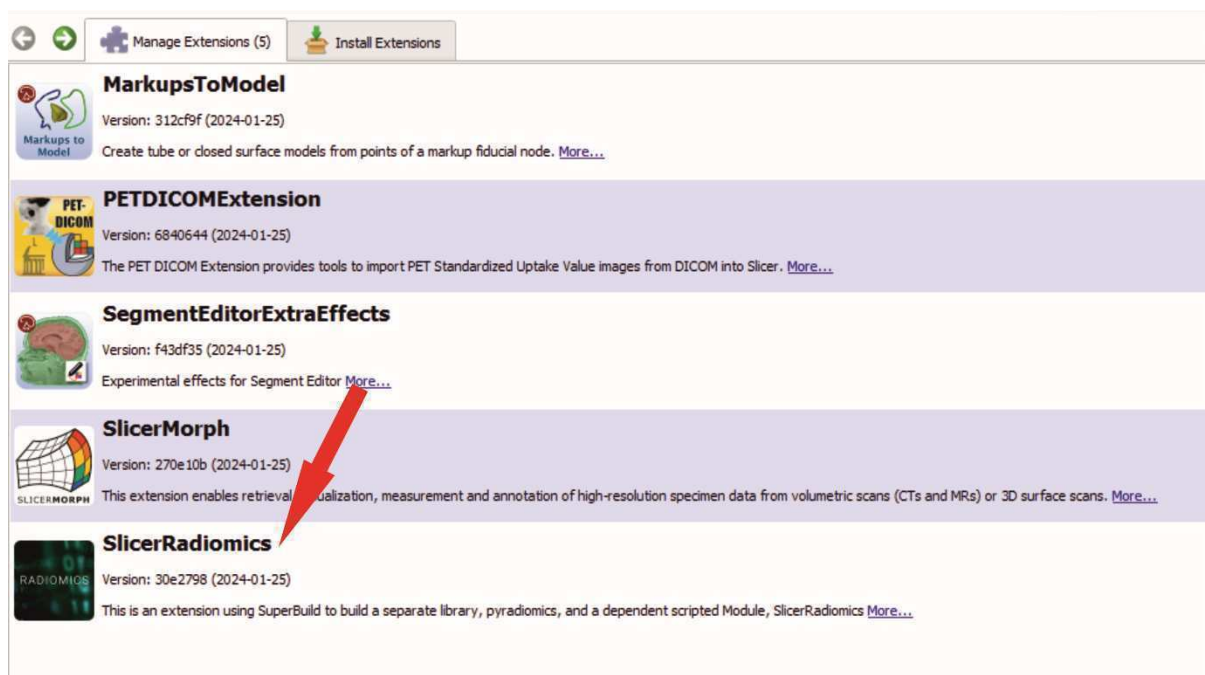


Рисунок 7 – Список установленных расширений


Расширение PETDICOMExtension

3D Slicer и PyRadiomics работают с большим числом форматов. В настоящих методических рекомендациях мы в качестве иллюстративного материала используем медицинские изображения ПЭТ-КТ (КТ-серия и ПЭТ-серия) как наиболее информативные для диагностики онкологических заболеваний и разметку новообразования в формате NifTI (расширение файла .nii). Радиомические признаки можно извлечь как из КТ-серии, так и из ПЭТ-серии.

Для загрузки ПЭТ-изображений в 3D Slicer понадобится расширение PETDICOMExtension, которое нужно установить тем же образом, что и SlicerRadiomics – через Extension Manager и дальнейший выбор нужного расширения (см. описание выше).

Загрузка изображений

Чтобы загрузить изображение и маску, нужно последовательно выполнить следующие действия:

 **Add Data** → **Choose File(s) to Add** → указание пути к нужным файлам → выбор файлов (рисунок 8), хранящихся в одной папке (например, DICOMDIR и Segmentation.nii).

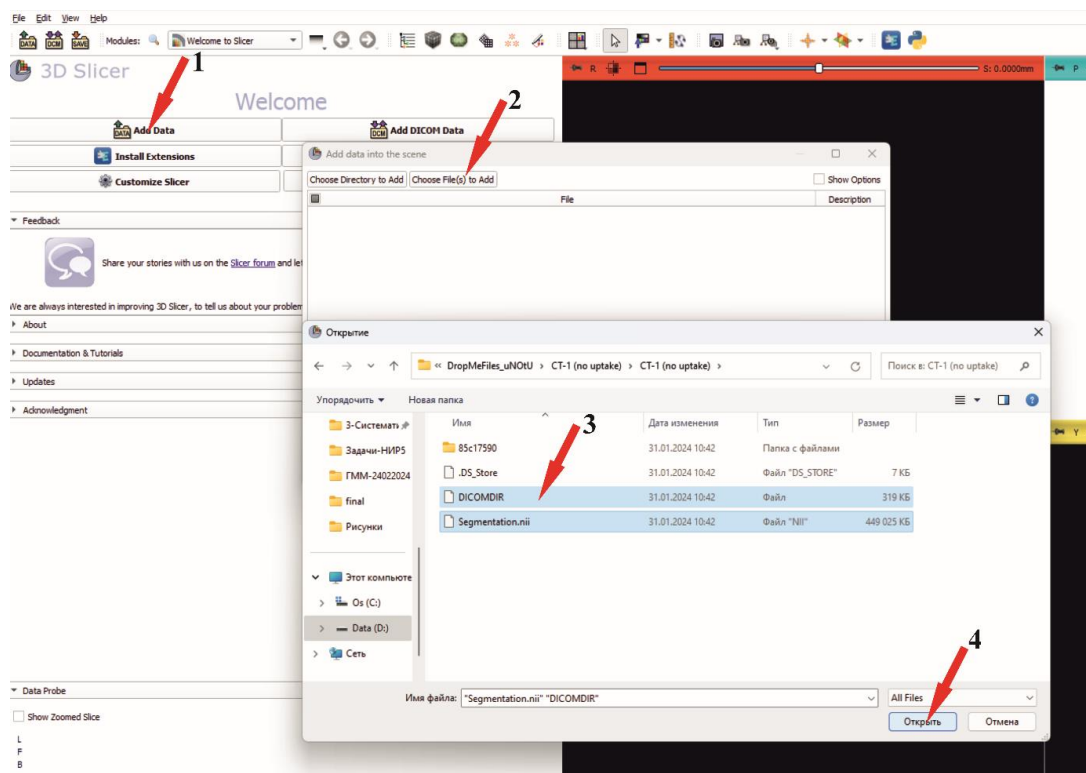


Рисунок 8 – Стрелками указана последовательность действий, необходимая для загрузки медицинского изображения и разметки к нему

Анализируемую маску необходимо загружать совместно с исходным медицинским изображением. При загрузке исходное медицинское изображение маркируем как DICOM import, маску – как Segmentation (по умолчанию программа предлагает маркировать маску как Volume – нужно изменить, чтобы маска считывалась как сегментация (рисунок 9)).

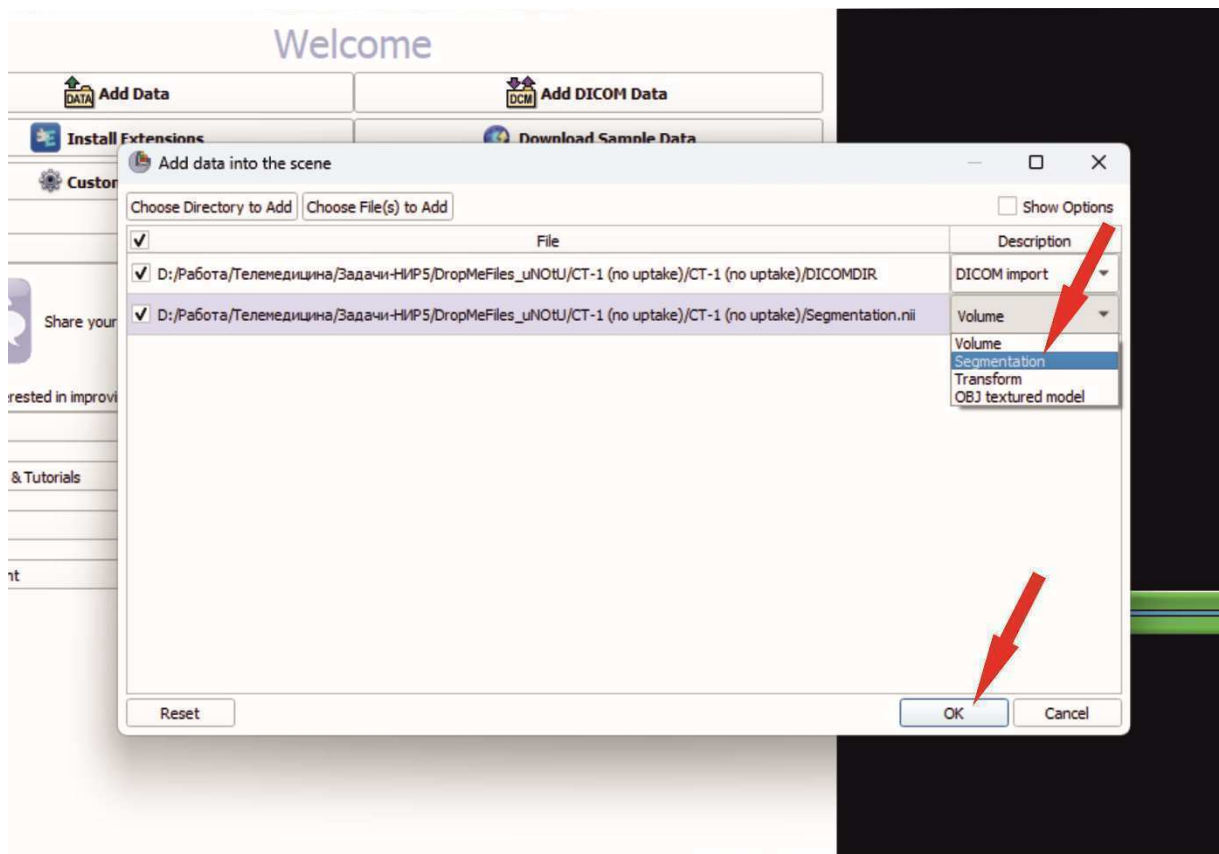
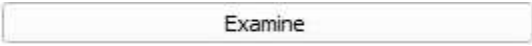
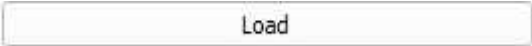


Рисунок 9 – Последовательность выбора типа файлов при загрузке

Для продолжения загрузки данных в окне **DICOM database** нужно выделить левой кнопкой мыши загружаемый файл (рисунок 10) и последовательно нажать кнопки  и .

DICOM database

Patients: Studies: Series:

Patient name	Patient ID	Birth date	Sex
1004562743	1004562743	1944-07-19	

1

Study date	Study ID	Study description
20181113		PET^Wholebody_FM_AV (Adult)

Series #	Series description
3	PET WB
4	CT WB 2.0 HD_FoV

DICOM Data Reader Warnings

2 **3**

Uncheck All Examine Load

Рисунок 10 – Последовательность проверки содержимого и загрузка медицинского изображения

Если загрузка файлов прошла правильно, на дереве данных должны отобразиться файлы медицинского изображения и файл с сегментацией (рисунок 11), помеченный цветом, соответствующим цвету узлов на срезах.

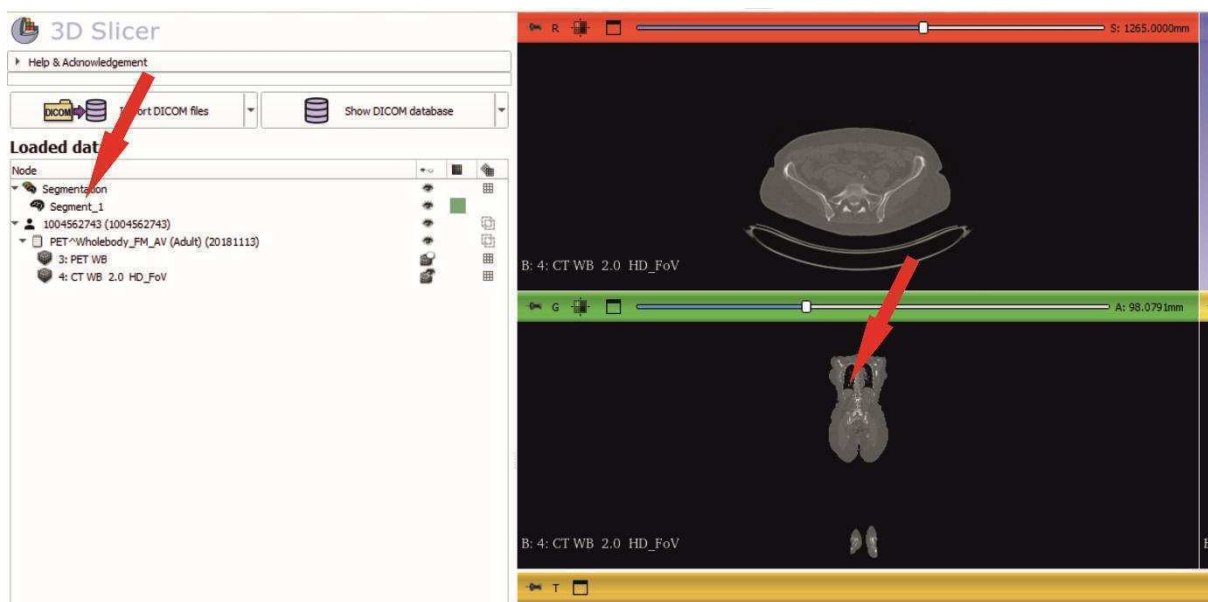



Рисунок 11 – Общий вид после загрузки файлов. Стрелками обозначена разметка на дереве файлов и на срезе

Управлять файлами на дереве данных можно, выделив файл левой кнопкой мыши, а затем нажав правую. Например, таким образом можно задавать удобные имена файлов (Rename). Эта опция может понадобиться при работе сразу с несколькими наборами данных.

Для получения трехмерной модели области разметки нужно открыть вкладку  Segment Editor, далее выбрать интересующую разметку и нажать клавишу Show 3D. Трехмерная модель сегментированной области появится в правом верхнем окне (рисунок 12), при этом модель может лежать вне области ограничительной рамки.

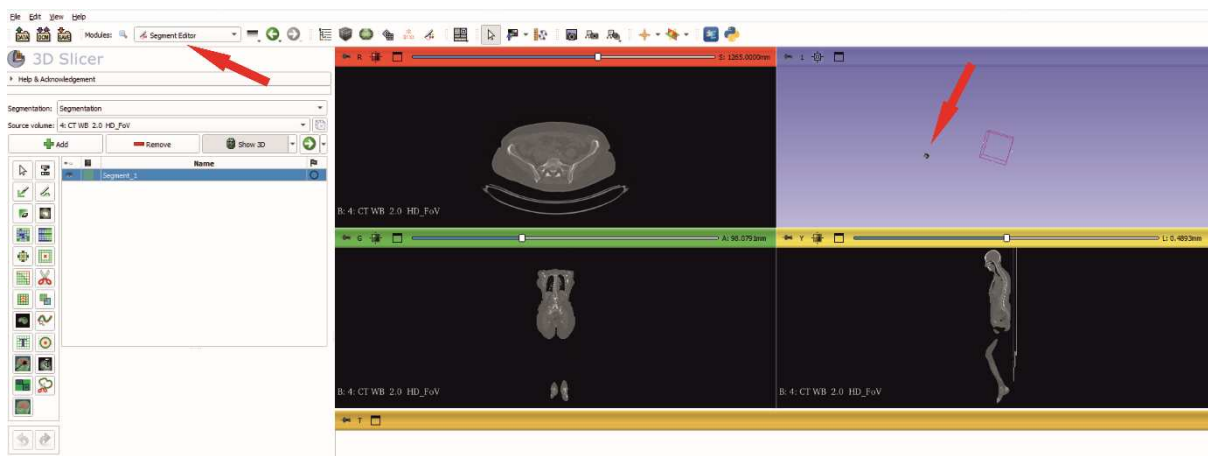



Рисунок 12 – Реконструкция трехмерной модели разметки. Стрелками указаны вкладка Segment Editor и трехмерная реконструкция разметки

Извлечение радиомических признаков

Общая последовательность для извлечения радиомических признаков в SlicerRadiomics для одного исследования:

- 1) открыть вкладку Informatics;
- 2) выбрать расширение Radiomics;
- 3) выбрать анализируемую серию (Input Image Volume);
- 4) выбрать маску изображения (Input regions);
- 5) выбрать извлекаемые радиомические признаки (по умолчанию All Features);
- 6) указать ширину столбцов диаграммы (binWidth, по умолчанию 25);
- 7) запустить процесс извлечения (кнопка Apply);
- 8) экспортировать таблицу в csv;
- 9) распределить данные по колонкам;
- 10) перейти к следующему исследованию.

Следует обратить внимание, что одна таблица с радиомическими признаками соответствует одной маске изображения, но при этом, в зависимости от методологии исследования, маска может содержать одну или несколько областей интереса.

Для извлечения радиомических признаков нужно выбрать вкладку Informatics, а в ней  (рисунок 13). В результате откроется окно радиомики (рисунок 14).

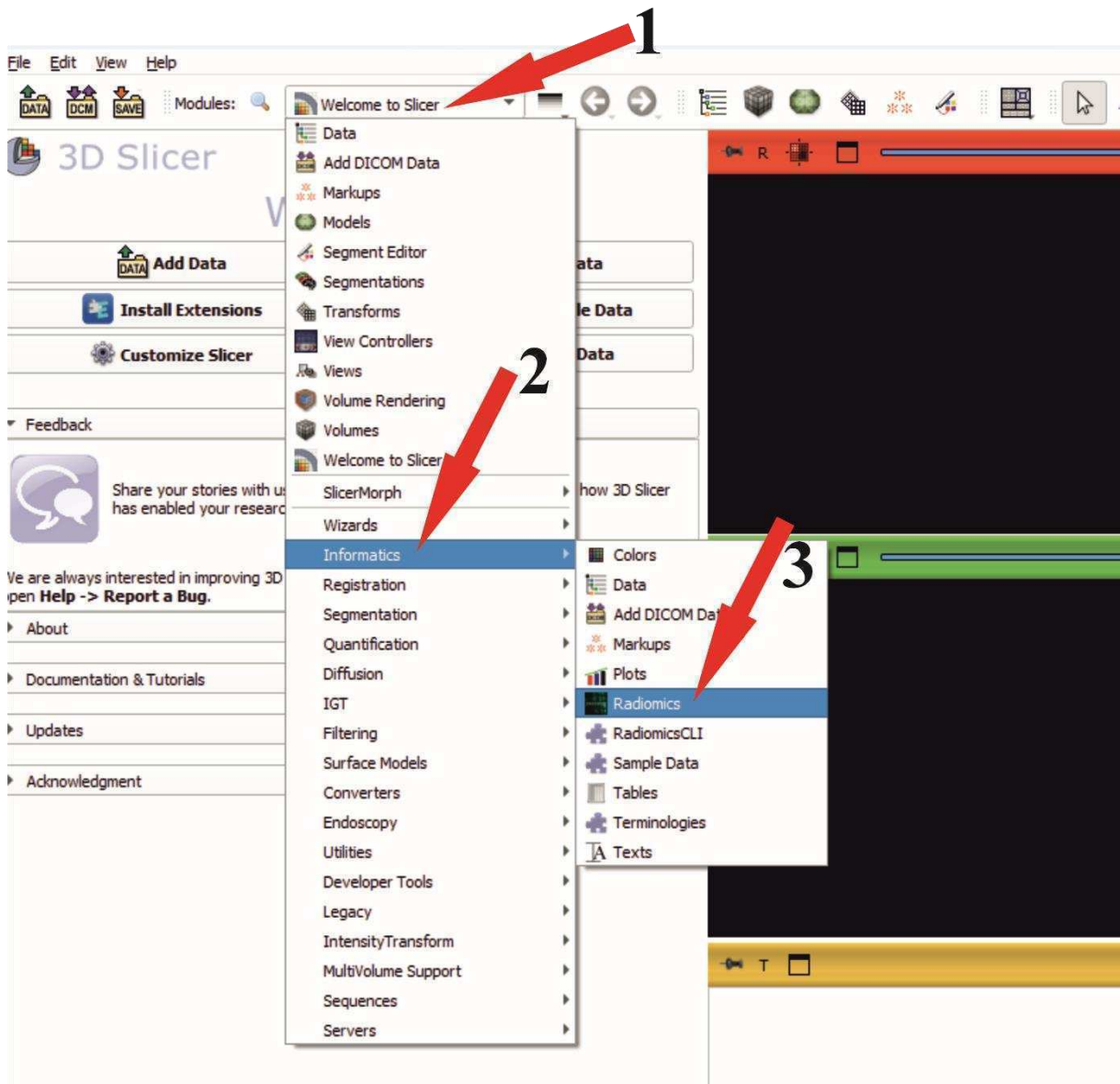


Рисунок 13 – Путь запуска расширения SlicerRadiomics. Стрелками указаны необходимые шаги, цифрами – их очередность

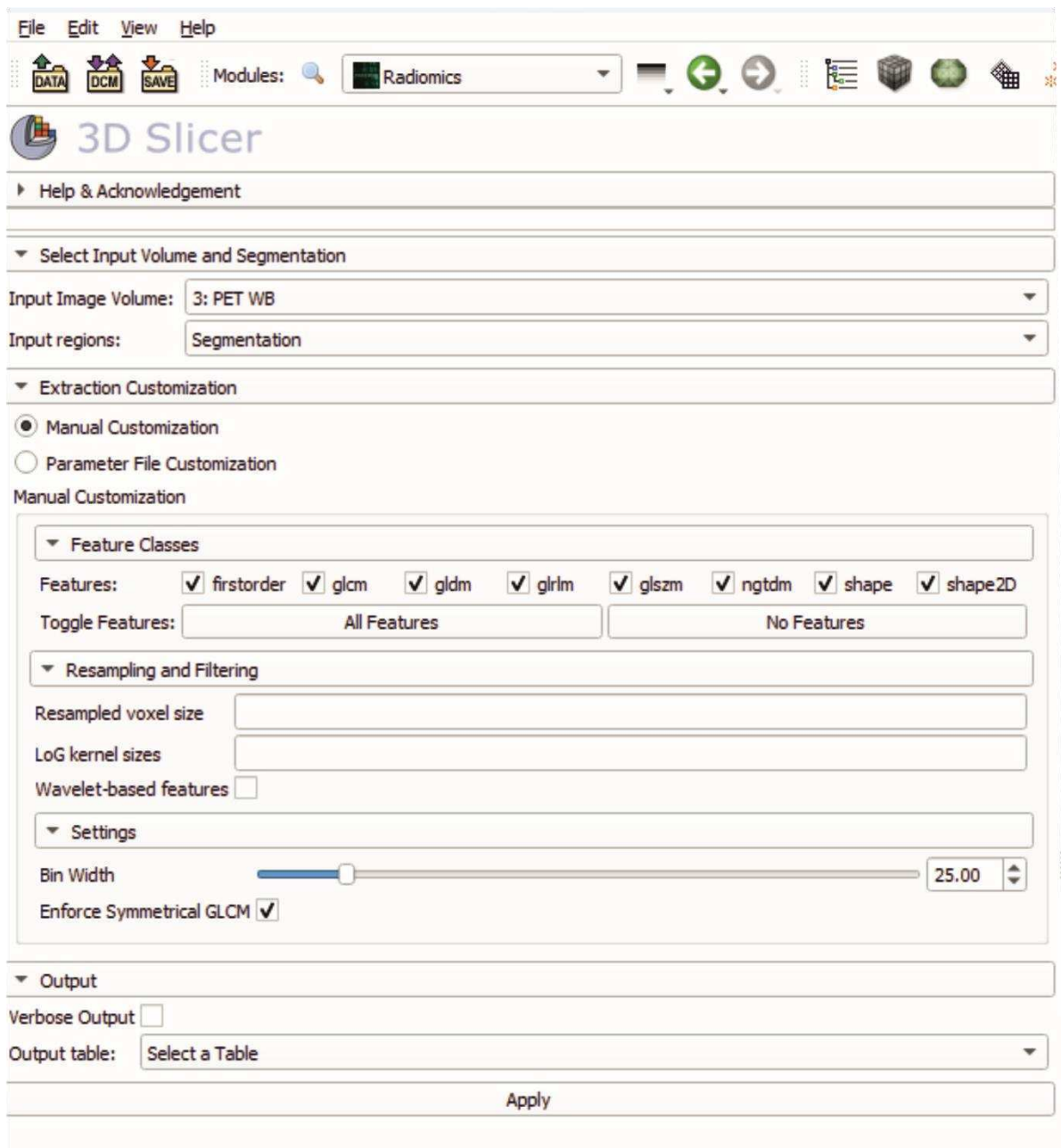


Рисунок 14 – Рабочее окно SlicerRadiomics

Далее во вкладке **Select Input Volume and Segmentation** необходимо выбрать интересующее изображение (в нашем случае это может быть PET или CT) и соответствующую этому изображению сегментацию.

Во вкладке **Extraction Customization** нужно выбрать радиомические признаки, которые будут извлечены (классы firstorder, gldm, glszm, ngtdm, shape, shape 2D). Выбрать все признаки сразу можно, нажав кнопку

All Features

Также можно добавить Wavelet-based features, поставив соответствующую галочку (этот фильтр описан при рассмотрении радиомических признаков).

Используя настройку параметра binWidth, можно указать ширину отдельных столбцов гистограммы, по которой рассчитываются радиомические показатели первого порядка (см. рисунок 1). По умолчанию в программе этот параметр равен 25, в литературе на сегодняшний день не существует единого мнения о его оптимальном значении. Для врача-рентгенолога принципиально важно выбрать единое значение при проведении собственных исследований, чтобы вычисления для разных пациентов проводились с идентичными настройками. Если же перед вами стоит задача сравнить собственные результаты с ранее опубликованными, значение этого параметра должно совпадать со значением в проверяемых данных. Наглядный пример влияния параметра binWidth на характеристики получаемых изображений приведен на рисунке 15.

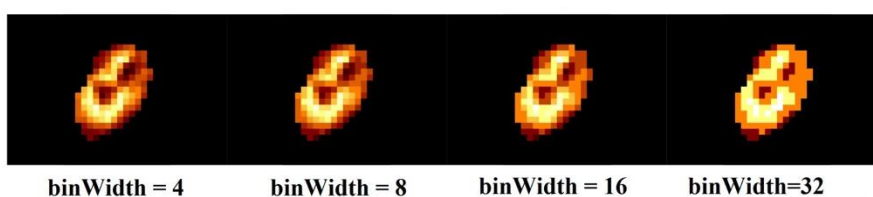


Рисунок 15 – Пример изображений, полученных с разными значениями параметра binWidth (IBSI <https://ibsi.readthedocs.io/>)

Результаты вычисления радиомических признаков программа записывает в таблицу, по умолчанию присваивая ей порядковый номер. Воспользовавшись командой Create New Table as во вкладке Output Table, вы можете присвоить таблице результатов нужное имя.

Нажмите кнопку **Apply** для запуска вычислений радиомических признаков.

Результирующая таблица, содержащая радиомические признаки размеченной области, появится в нижнем окне, а также на дереве данных (рисунок 16).

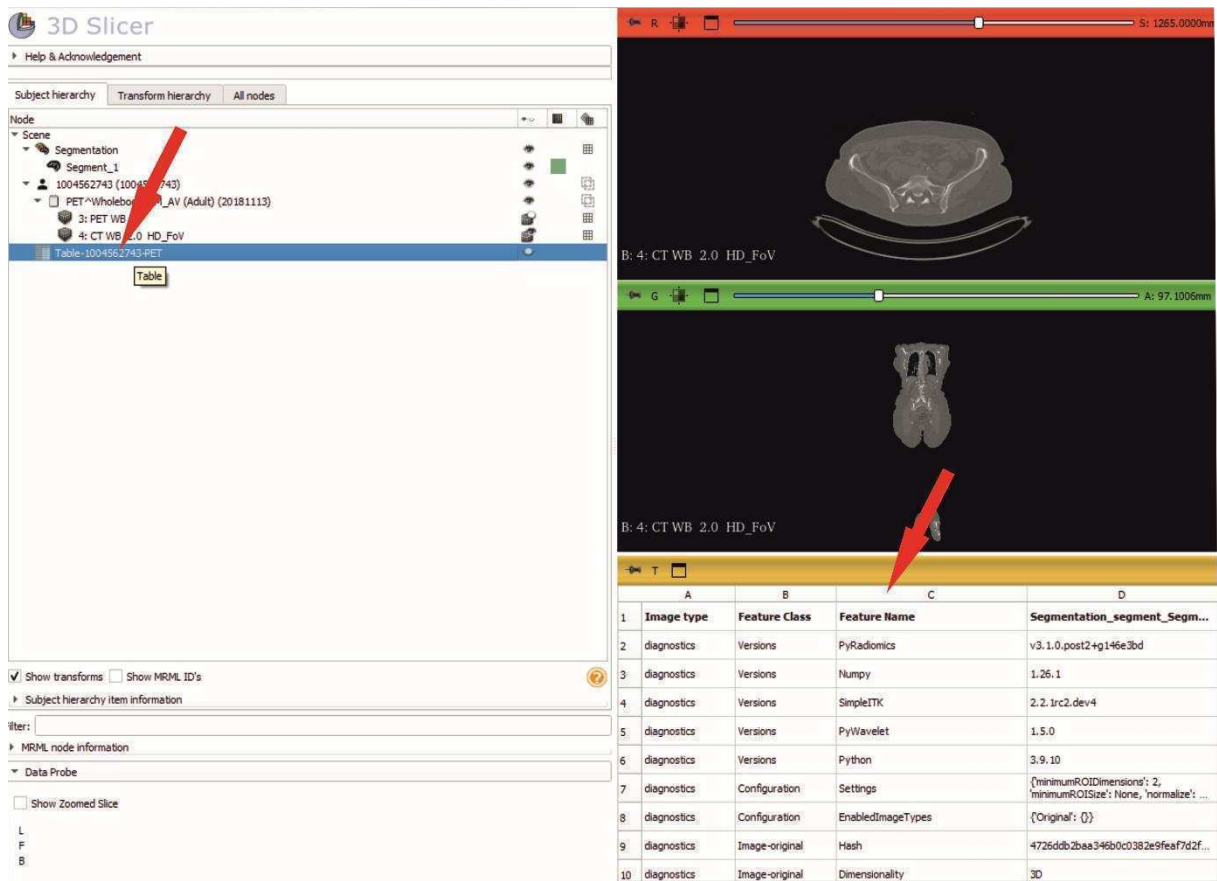


Рисунок 16 – Окно радиомики с рассчитанными радиомическими признаками. Стрелками указана созданная таблица с радиомическими признаками

Для дальнейшего анализа полученную таблицу нужно экспортировать в csv-файл (рисунки 17, 18). Для этого на дереве данных нужно выделить таблицу левой кнопкой мыши → нажать правую кнопку мыши → в появившемся списке команд выбрать Export to file → задать путь и выбрать нужное расширение сохраняемой таблицы (.csv).

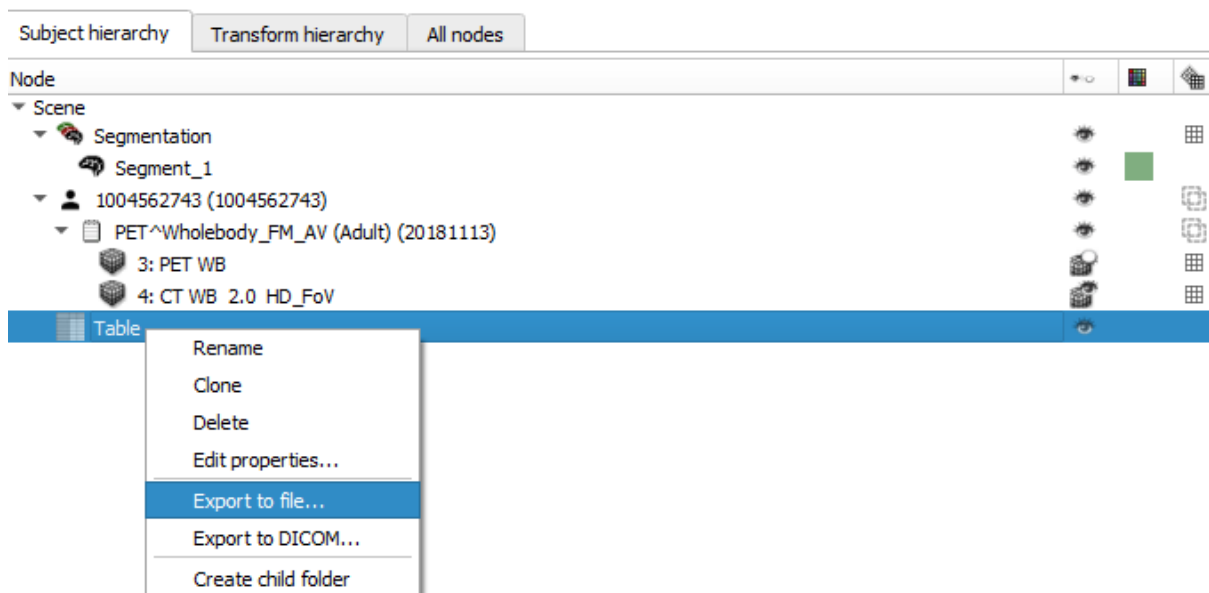


Рисунок 17 – Сохранение таблицы с радиомическими признаками: выбор команды Export to file

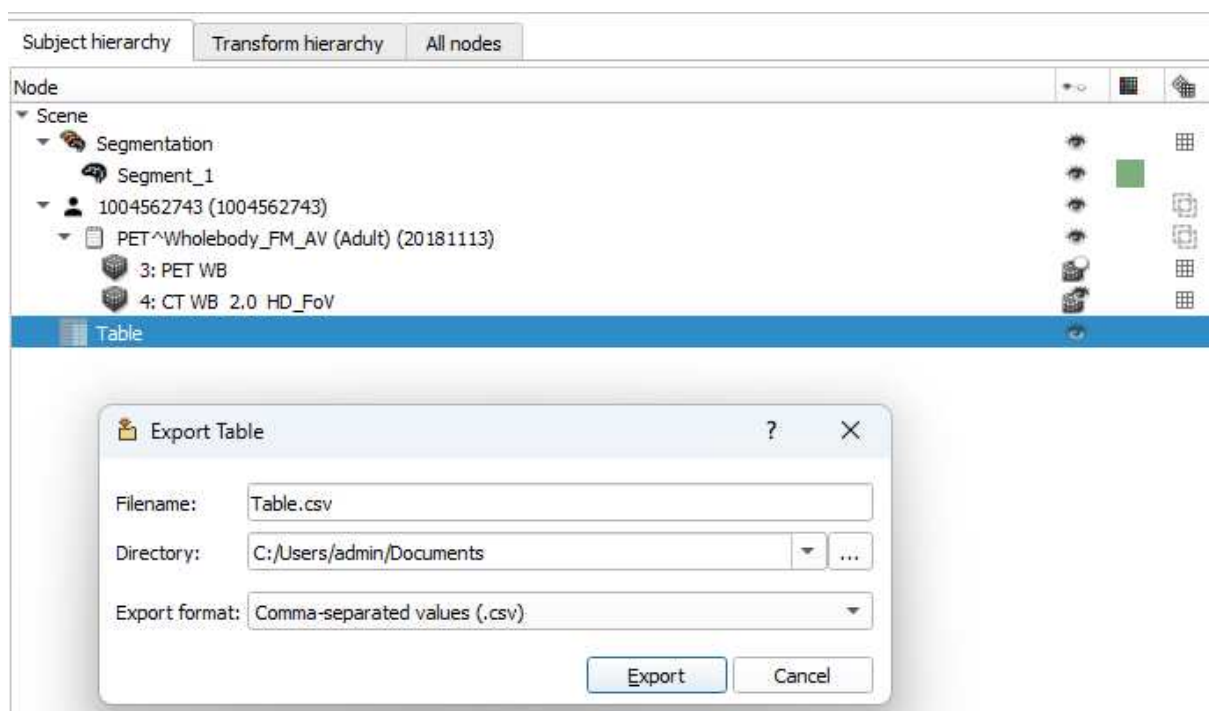


Рисунок 18 – Сохранение таблицы с радиомическими признаками: выбор формата файла для экспорта

Полученный csv-файл нужно открыть с помощью программы Excel. При открытии файла вы увидите, что все данные находятся в одной колонке (рисунок 19).

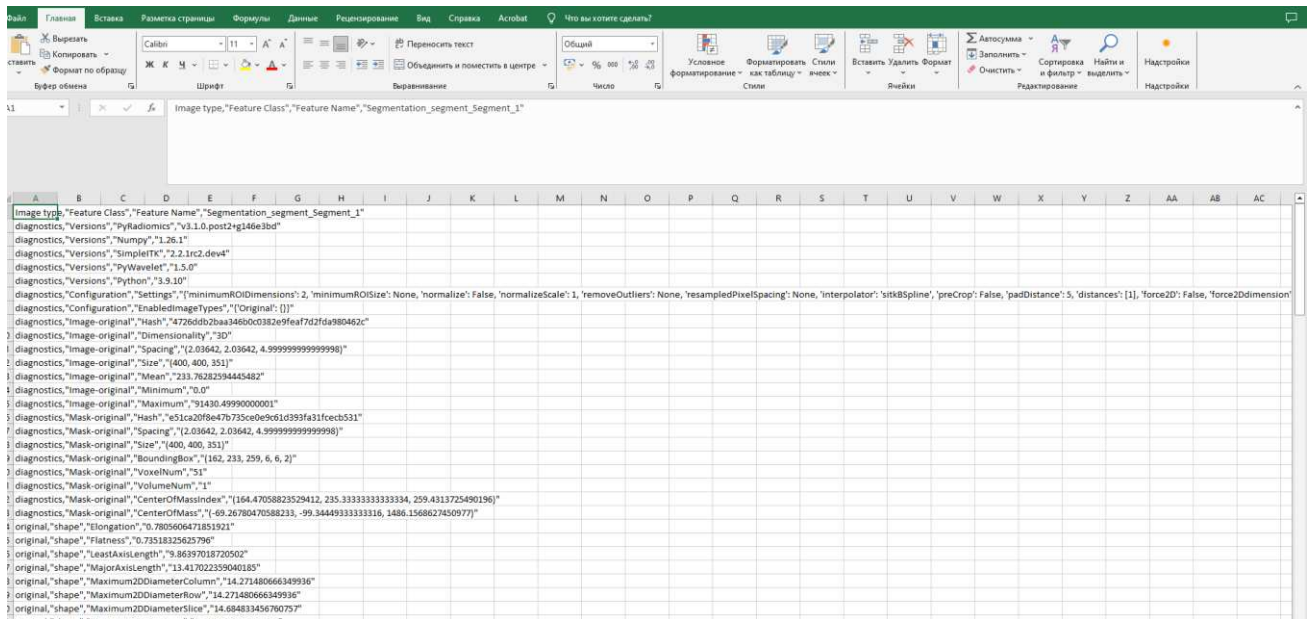


Рисунок 19 – Первоначальный вид файл с радиомическими признаками в Excel

Полученные данные нужно правильно распределить по колонкам (рисунок 20). Для этого необходимо выполнить следующие действия: выделить данные для разбора → «Данные» → «Текст по столбцам».

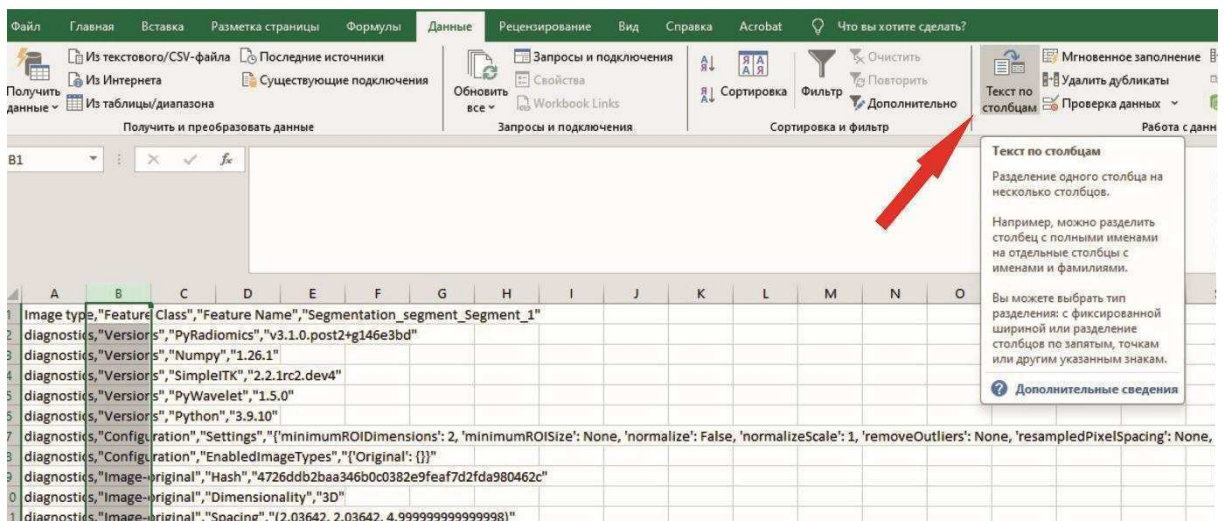


Рисунок 20 – Распределение данных по столбцам: выбор функции «Текст по столбцам» (указано стрелкой)

После перехода в «Мастер распределения текста по столбцам» необходимо выбрать нужные характеристики текста и разделителей (рисунки 21, 22).

Шаг 1: «с разделителями» → «Далее».

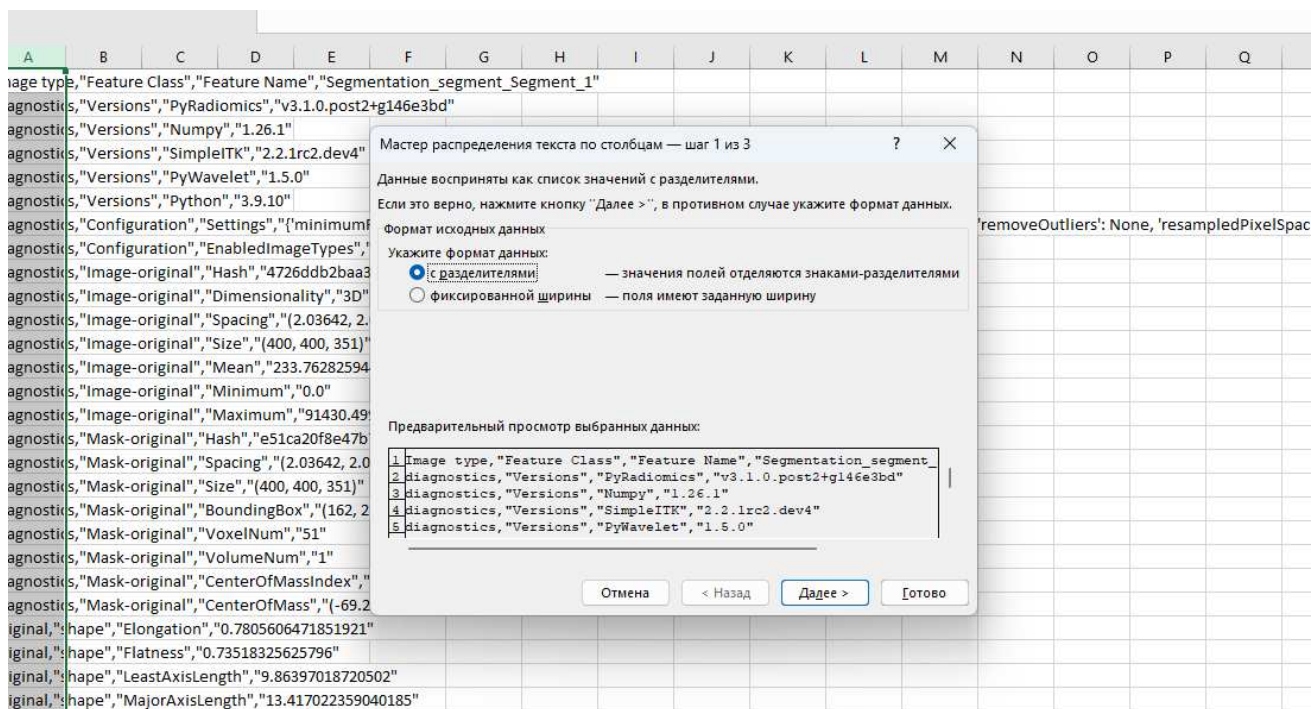


Рисунок 21 – Распределение данных по столбцам: шаг 1

Шаг 2: выбор нужного разделителя. В нашем примере это «запятая». Проверяем предлагаемое программой распределение данных по столбцам. Для завершения сортировки нужно нажать кнопку «Готово».

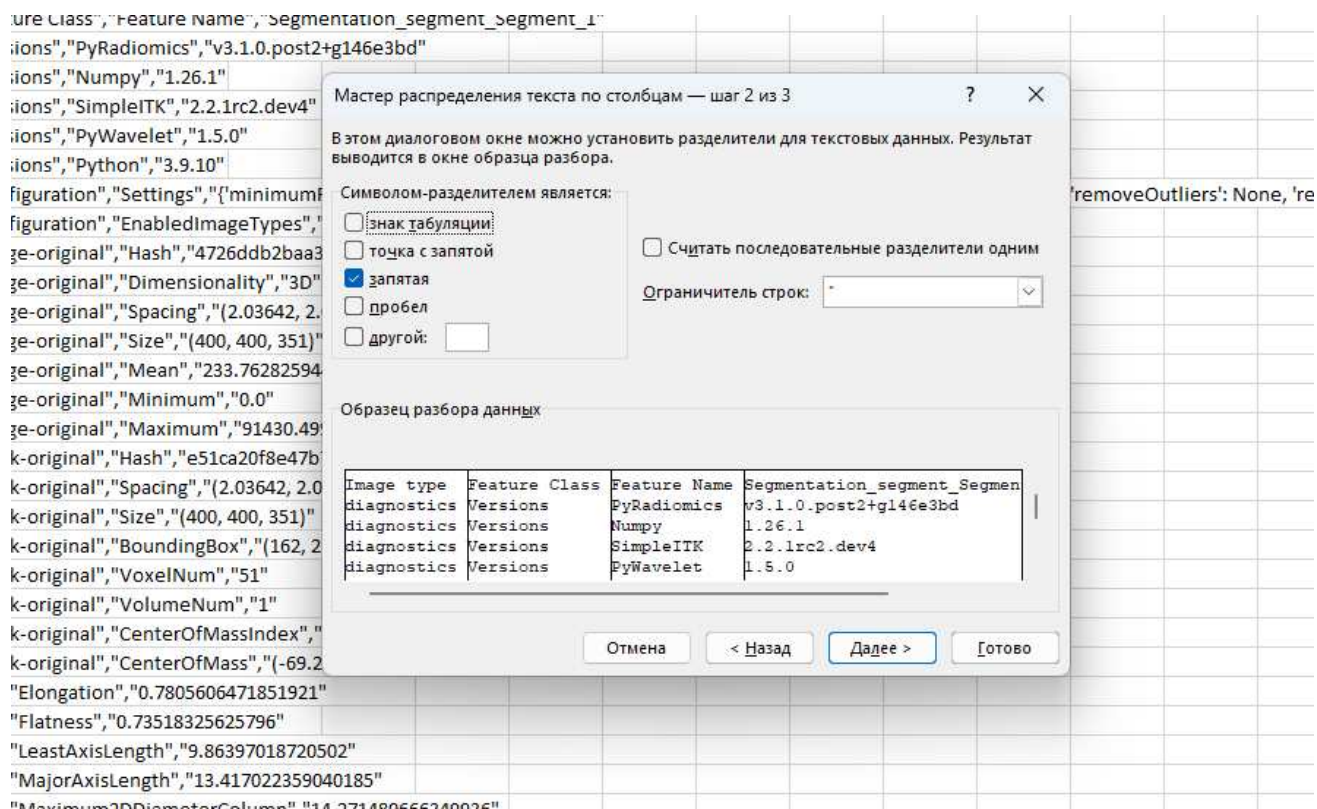


Рисунок 22 – Распределение данных по столбцам: шаг 2

На выходе должна получиться таблица с четырьмя колонками (рисунок 23):

- 1) Image type – оригинальное изображение или с использованием фильтра (в случае если использовали также Wavelet-based features);
- 2) Feature class – группа, куда относится радиомический признак;
- 3) Feature name – название радиомического признака;
- 4) Segmentation_segment_Segmen – численное значение радиомического признака.

Первые строки таблицы, описывающие программные характеристики (в столбике Image type, обозначенные как diagnostics), нужно удалить, так как они не участвуют в дальнейшем анализе радиомических признаков.

	A	B	C	D	E
1	Image type	Feature Class	Feature Name	Segmentation_segment_Segment_1	
24	original	shape	Elongation	0.7805606471851921	
25	original	shape	Flatness	0.73518325625796	
26	original	shape	LeastAxisLength	9.86397018720502	
27	original	shape	MajorAxisLength	13.417022359040185	
28	original	shape	Maximum2DDiameter	14.271480666349936	
29	original	shape	Maximum2DDiameter	14.271480666349936	
30	original	shape	Maximum2DDiameter	14.684833456760757	
31	original	shape	Maximum3DDiameter	16.432506293088707	
32	original	shape	MeshVolume	959.8591934625834	
33	original	shape	MinorAxisLength	10.4727996558706	
34	original	shape	Sphericity	0.8656376671911433	
35	original	shape	SurfaceArea	543.6086154643563	
36	original	shape	SurfaceVolumeRatio	0.5663420417981826	
37	original	shape	VoxelVolume	1057.4866361819995	
38	original	firstorder	10Percentile	3069.7661279999998	
39	original	firstorder	90Percentile	4458.4743849999995	
40	original	firstorder	Energy	772674944.1197819	
41	original	firstorder	Entropy	5.108604018399648	

Рисунок 23 – Итоговая таблица с радиомическими признаками для одного пациента

Каждое вычисление радиомических признаков записывается в отдельную таблицу. Для сопоставления радиомических признаков разных пациентов удобно объединить их в одну таблицу (рисунок 24), используя Copy-Paste.

	A	B	C	D	Строка формул E	F	G	H
1				3374083986	3374083986	1004562743	1004562743	номер экземпляра
2				high uptake	high uptake	no uptake	no uptake	поглощение
3				CT	PET	CT	PET	тип изображения
4	Image type	Feature Class	Feature Name	Segmentation_segment	Segmentation_segment_Se	Segmentation_segment_S	Segmentation_segment_Segment_1	
29	original	shape	LeastAxisLength	14.732311537844128	14.661577257676282	10.514218149012121	9.86397018720502	
30	original	shape	MajorAxisLength	18.955609666971807	19.55438394692931	13.808889999483549	13.417022359040185	
31	original	shape	Maximum2DDiameterColumn	22.26542797336072	22.442708401839557	16.03993581706688	14.271480666349936	
32	original	shape	Maximum2DDiameterRow	21.142306670457135	23.43698425545402	16.03993581706688	14.271480666349936	
33	original	shape	Maximum2DDiameterSlice	21.460523070176006	20.96622713170875	15.471690505315996	14.684833456760757	
34	original	shape	Maximum3DDiameter	23.080628834559114	25.308622486243692	16.647407403205193	16.432506293088707	
35	original	shape	MeshVolume	3398.14204788208	3389.3137857369225	1217.7712955474854	959.8591934625834	
36	original	shape	MinorAxisLength	17.62668929548487	17.88789936320569	12.39953262097976	10.4727996558706	
37	original	shape	Sphericity	0.8080541889920145	0.8256577009882606	0.7987704139183822	0.8656376671911433	
38	original	shape	SurfaceArea	1352.7098670227315	1321.5753834041836	690.408255889255	543.6086154643563	
39	original	shape	SurfaceVolumeRatio	0.398073373026245	0.38992417549702907	0.5669441038835304	0.5663420417981826	
40	original	shape	VoxelVolume	3433.5692825317383	3566.425518103999	1242.8256912231445	1057.4866361819995	
41	original	firstorder	10Percentile	-84.79999999999998	12083.998017599999	-453.0	3069.7661279999998	
42	original	firstorder	90Percentile	61.0	53339.11605140001	40.0	4458.4743849999995	
43	original	firstorder	Energy	9197246.0	241281668379.23053	28408436.0	772674944.1197819	
44	original	firstorder	Entropy	2.6862654737565346	7.340480525038304	3.982912648806399	5.108604018399648	
45	original	firstorder	InterquartileRange	40.0	21750.1061495	218.0	715.2829590000001	

Рисунок 24 – Таблица радиомических признаков с несколькими столбцами, соответствующим двум пациентам (КТ-серия и ПЭТ-серия)

Для создания радиомических моделей, позволяющих решать задачи диагностики и планирования терапии, необходимо большое число пациентов в итоговой таблице: минимальное число составляет 25 пациентов, максимальное число не ограничено, оптимальное составляет несколько сотен пациентов [25].

Важно следить, чтобы были загружены именно нужные вам объем (серия изображений) и сегментация (область интереса), а также чтобы названия разных таблиц с результатами не совпадали друг с другом, иначе программа перезаписывает таблицу, не сохраняя предыдущий результат. Таким образом, целесообразно анализировать по одному исследованию, экспортируя таблицу с радиомическими признаками, а далее перезагружать 3D Slicer для анализа нового исследования.

Следующим этапом в радиомическом анализе является создание математической модели на основе данных с известными вводными, которая будет использоваться врачом для классификации новых пациентов. Для построения такой модели исследователю необходимо иметь подготовку в области статистического анализа. В связи с этим целесообразно привлекать медицинских статистиков для построения модели радиомического анализа. Практический пример построения математической модели с использованием текстурных показателей приведен в приложении А.

ЗАКЛЮЧЕНИЕ

Радиомика – перспективный метод медицинской диагностики, который может стать полезным рутинным инструментом врачей. В открытом доступе находятся удобные инструменты для самостоятельного извлечения радиомических признаков врачом-рентгенологом. Настоящие рекомендации представляют собой подробную пошаговую инструкцию по извлечению радиомических признаков из медицинских изображений с помощью 3D Slicer и его расширения SlicerRadiomics.

В работе проведен сравнительный анализ программ для извлечения радиомических признаков. PyRadiomics является в настоящий момент наиболее широко используемым инструментом. Приведено подробное описание радиомических признаков. Даны инструкции по установке необходимого программного обеспечения, загрузке медицинских изображений и собственно процессу извлечения радиомических признаков. Представленная методическая разработка позволит пользователям извлекать радиомические признаки из доступных медицинских изображений.

Наиболее сложный этап в радиомическом анализе – построение радиомической модели. Для этого пользователю необходима подготовка в области методов статистического анализа или помощь медицинского статистика. Радиомический анализ чувствителен к параметрам загружаемых медицинских изображений и настройкам при извлечении радиомических признаков, что ограничивает применение создаваемой статистической модели на новых данных. Также существует огромное число опубликованных работ по созданию моделей для каждой отдельной патологии. Основное направление развития в радиомике – стандартизация радиомических признаков и генерализация радиомических моделей для широкого использования их на практике.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Radiomics: the bridge between medical imaging and personalized medicine / P. Lambin, R. T. H. Leijenaar, T. M. Deist, et al. // *Nature Reviews Clinical Oncology*. – 2017. – Vol. 14 (12) – P. 749–762. – DOI: 10.1038/nrclinonc.2017.141.
2. Radiomics: the facts and the challenges of image analysis / S. Rizzo, F. Botta, S. Raimondi, et al. // *European Radiology Experimental*. – 2018. – Vol. 2 (1). – P. 36. – DOI: 10.1186/s41747-018-0068-z.
3. Radiomics: Extracting more information from medical images using advanced feature analysis / P. Lambin, E. Rios-Velazquez, R. Leijenaar, et al. // *European Journal of Cancer*. – 2012. – Vol. 48 (4) – P. 441–446. – DOI: 10.1016/j.ejca.2011.11.036.
4. Radiomics: the process and the challenges / V. Kumar, Y. Gu, S. Basu, et al. // *Magnetic Resonance Imaging*. – 2012. – Vol. 30 (9). – P. 1234–1248. – DOI: 10.1016/j.mri.2012.06.010.
5. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach / H. J. W. L. Aerts, E. R. Velazquez, R. T. H. Leijenaar, et al. // *Nature Communications*. – 2014. – Vol. 5. – P. 4006. – DOI: 10.1038/ncomms5006.
6. Rutman A. M., Kuo M. D. Radiogenomics: creating a link between molecular diagnostics and diagnostic imaging // *European Journal of Radiology*. – 2009. – Vol. 70 (2). – P. 232–241. – DOI: 10.1016/j.ejrad.2009.01.050.
7. Radiogenomics: radiobiology enters the era of big data and team science / B. S. Rosenstein, C. M. West, S. M. Bentzen, et al. // *International Journal of Radiation Oncology, Biology, Physics*. – 2014. – Vol. 89 (4). – P. 709–713. – DOI: 10.1016/j.ijrobp.2014.03.009.
8. Formenti S. C., Demaria S. Combining radiotherapy and cancer immunotherapy: a paradigm shift // *Journal of the National Cancer Institute*. – 2013. – Vol. 105 (4). – P. 256–265. – DOI: 10.1093/jnci/djs629.

9. Pathology-based non-small cell lung cancer radiomics signature describing the local tumor immune environment: discovery and validation / C. Tang, A. Amer, B. Hobbs, et al. // *International Journal of Radiation Oncology, Biology, Physics*. – 2016. – Vol. 96 (2S). – P. S42–S43. – DOI: 10.1016/j.ijrobp.2016.06.114.
10. Delta-radiomics features for the prediction of patient outcomes in non-small cell lung cancer / X. Fave, L. Zhang, J. Yang, et al. // *Scientific Reports*. – 2017. – Vol. 7 (1). – P. 588. – DOI: 10.1038/s41598-017-00665-z.
11. Radiomics of CT Features May Be Nonreproducible and Redundant: Influence of CT Acquisition Parameters / R. Berenguer, M. R. Pastor-Juan, J. Canales-Vázquez, et al. // *Radiology*. – 2018. – Vol. 288 (2). – P. 407–415. – DOI: 10.1148/radiol.2018172361.
12. Приоритетные параметры радиомического анализа для компьютерной томографии при злокачественных новообразованиях головы и шеи: систематический обзор / Ю. А. Васильев, О. Г. Нанова, И. А. Блохин [и др.]. // *Digital Diagnostics*. – 2024. – Т. 5 (2). – С. 255–268. – DOI: 10.17816/DD623240.
13. Image Biomarker Standardisation Initiative (IBSI). – URL: <https://theibsi.github.io> (дата обращения: 20.05.2024).
14. Обзор современных средств разметки цифровых диагностических изображений / Ю. А. Васильев, Е. Ф. Савкина, А. В. Владзимирский [и др.]. // *Казанский медицинский журнал*. – 2023. – Т. 104 (5). – С. 750–760. – DOI: 10.17816/KMJ349060.
15. Сегментация артериальных сосудов на основании данных КТ-ангиографии с использованием программы 3D Slicer / авт.-сост. М. Р. Коденко, Ю. А. Васильев, А. В. Владзимирский [и др.] // Серия «Лучшие практики лучевой и инструментальной диагностики». – Вып. 138. – М. : ГБУЗ «НПКЦ ДиТ ДЗМ», 2024. – 64 с.
16. Computational radiomics system to decode the radiographic phenotype / J. J. M. van Griethuysen, A. Fedorov, Ch. Parmar, et al. // *Cancer*

Research. – 2017. – Vol. 77 (21). – P. e104–e107. – DOI: 10.1158/0008-5472.CAN-17-0339.

17. LIFEx: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity / C. Nioche, F. Orhac, S. Boughdad, et al. // *Cancer Research*. – 2018. – Vol. 78 (16). – P. 4786–4789. – DOI: <https://doi.org/10.1158/0008-5472.CAN-18-0125>.

18. MaZda – a software package for image texture analysis / P. M. Szczypiński, M. Strzelecki, A. Materka, A. Klepaczko // *Computer Methods and Programs Biomedicine*. – 2009. – Vol. 94 (1). – P. 66–76. – DOI: 10.1016/j.cmpb.2008.08.005.

19. IBEX: An open infrastructure software platform to facilitate collaborative work in radiomics / L. Zhang, D. V. Fried, X. J. Fave, et al. // *Medical Physics*. – 2015. – Vol. 42 (3). – P. 1341–1353. – DOI: 10.1118/1.4908210.

20. Reliability and prognostic value of radiomic features are highly dependent on choice of feature extraction platform / I. Fornaçon-Wood, H. Mistry, C. J. Ackermann, et al. // *European Radiology*. – 2020. – Vol. 30 (11). – P. 6241–6250. – DOI: 10.1007/s00330-020-06957-9.

21. The Design of SimpleITK / B. C. Lowekamp, D. T. Chen, L. Ibáñez, D. Blezek // *Frontiers in Neuroinformatics*. – 2013. – Vol. 7. – P. 45. – DOI: 10.3389/fninf.2013.00045.

22. Beare R., Lowekamp B. C., Yaniv Z. Image segmentation, registration and characterization in R with SimpleITK // *Journal of Statistical Software*. – 2018. – Vol. 86 (8). – P. 1-35. – DOI: 10.18637/jss.v086.i08.

23. SimpleITK Image-Analysis Notebooks: a Collaborative Environment for Education and Reproducible Research / Z. Yaniv, B. C. Lowekamp, H. J. Johnson, R. Beare // *Journal of Digital Imaging*. – 2018. – Vol. 31 (3). – P. 290–303. – DOI: 10.1007/s10278-017-0037-8.

24. Радиомика магнитно-резонансной томографии при раке предстательной железы: что известно в настоящее время? / П. Б. Гележе,

И. А. Блохин, С. С. Семенов, D. Caruso // Digital Diagnostics. – 2021. – Т. 2 (4). – С. 441–452. – DOI: 10.17816/DD70170.

25. Norman G., Monteiro S., Salama S. Sample size calculations: should the emperor's clothes be off the peg or made to measure? // BMJ. – 2012. – Vol. 345. – P. e5278. – DOI: 10.1136/bmj.e5278.

26. The R Project for Statistical Computing. – URL: <https://www.r-project.org/> (дата обращения: 17.05.2024).

27. Mosquera C. Radiomics for LGG dataset. – URL: <https://kaggle.com/competitions/glioma-radiomics> (дата обращения: 17.05.2024).

Пример статистического анализа радиомических признаков

Методы анализа радиомических признаков для создания на их основе предсказательных моделей

Основной задачей радиомики является классификация данных (например, злокачественный/доброкачественный характер новообразований, определение стадии заболевания).

Перед статистической обработкой радиомических признаков необходимо произвести их стандартизацию. Стандартизация изменяет масштаб набора данных таким образом, чтобы иметь среднее значение 0 и стандартное отклонение 1. Для этого используется следующая формула:

$$X_{\text{стандарт}} = (x_i - X_{\text{среднее}}) / \text{std},$$

Где $X_{\text{стандарт}}$ – стандартизованное значение признака, x_i – исходное значение признака, $X_{\text{среднее}}$ – средневывборочное значение признака, std – стандартное отклонение в выборке.

Особенность работы при анализе радиомических признаков заключается в том, что исследователь получает данные с очень большим числом переменных (радиомических признаков) при относительно небольшом числе исследований.

При такой структуре данных остро стоит проблема переобучения (*overfitting*) модели (рисунок А.1) – это случай, когда модель, построенная на большом числе признаков с малым числом пациентов, учитывает все случайные закономерности конкретной выборки, отсутствующие в генеральной совокупности (т. е. на большом числе пациентов). В итоге построенная модель хорошо работает на данных, на которых она была построена, но не работает на других (тестовых) данных из той же генеральной совокупности.

Такой результат связан с тем, что только некоторые признаки вносят значимый вклад в классификацию групп. Часть из них не являются значимыми для классификации и вносят шум в создаваемую модель, т. е. должны быть исключены. Кроме того, некоторые признаки тесно скоррелированы (связаны) между собой, т. е. включение всей этой группы также не дает новой информации для создаваемой модели. Предлагается оставлять один признак (т. н. архетипический) из группы тесно скоррелированных признаков.

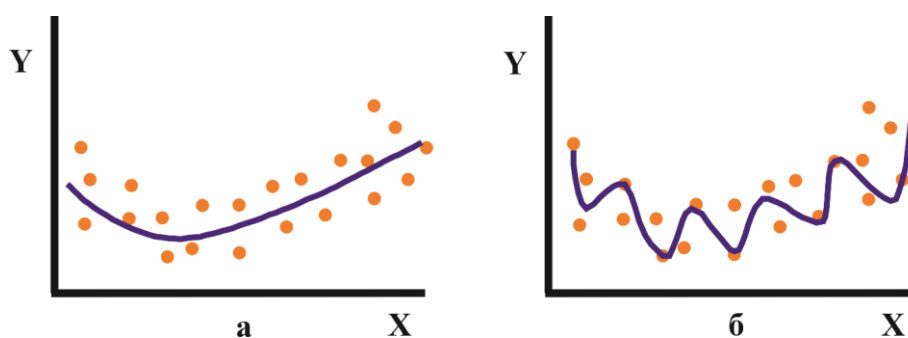


Рисунок А.1 – Пример переобучения модели: а) корректно подобранная модель; б) модель с эффектом переобучения

Исходя из вышесказанного, первым этапом в построении радиомической модели является редукция числа признаков. Существует довольно большое разнообразие методов для решения этой задачи. Рассмотрим некоторые из них.

Метод главных компонент (Principal component analysis) – это один из методов ординации данных. Этот метод вычисляет новые комплексные признаки – главные компоненты, ГК (principal components, PC) – таким образом, что на первую ГК приходится наибольшая доля изменчивости в выборке, на вторую – наибольшая доля изменчивости, следующая за первой, и т. д. На последние ГК приходится наименьшая доля изменчивости. В этом и заключается ординация – перераспределение изменчивости таким образом, что существенная ее часть, как правило, несущая смысловую нагрузку, приходится на первые ГК, а на последние – минимальная доля

изменчивости, как правило, включающая в себя случайную изменчивость, т. е. «шум». Исследователей интересуют первые ГК, объясняющие в совокупности 80–90 % от общей изменчивости; остальные исключаются из анализа. Анализ главных компонент позволяет оценить, какие исходные радиомические признаки коррелируют с отдельными из них. Таким образом осуществляется отбор значимых радиомических признаков. Кроме того, анализ ГК показывает, какие из них (а соответственно, и радиомические признаки) вносят вклад в различия (если они имеются) между интересующими группами пациентов.

Еще одним важным методом для отбора радиомических признаков является внутрикласовый коэффициент корреляции (intraclass correlation coefficient, ICC). ICC позволяет оценивать устойчивость радиомических признаков на основе их воспроизводимости, например, между разными исследователями. В анализе оставляются радиомические признаки с высоким ICC (например, выше 0,8) и исключаются обладающие низкой воспроизводимостью.

После отбора значимых радиомических признаков для построения предсказательных моделей используют линейную модель. Обычная регрессионная модель выглядит как:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon,$$

где y – зависимая переменная; x_{1-p} – независимые переменные, факторы; β_0 – свободный коэффициент; β_{1-p} – коэффициенты (slope); ϵ – регрессионные остатки.

Линейная регрессия не может быть использована в случаях, когда число оцениваемых параметров больше числа пациентов, что характерно именно для радиомики.

Поэтому для радиомического анализа используют усложненный вариант моделирования – LASSO-регрессию (least absolute shrinkage and selection operator). Это метод машинного обучения, который выполняет

одновременно отбор переменных и регуляризацию, чтобы повысить точность прогнозирования и интерпретируемость полученной статистической модели. В LASSO-регрессию вводится гиперпараметр λ , который оценивается исследователем с помощью кросс-валидации и позволяет исключить признаки с нулевой предсказательной силой, тем самым решая проблему переобучения модели. Кросс-валидация представляет собой разделение обучающей выборки данных на блоки, на которых происходит обучение модели, а затем финальная оптимизация результата.

На выходе LASSO исследователь получает отобранный набор признаков, имеющих наибольшую предсказательную ценность, и коэффициенты для них.

Исходные данные должны быть разделены на две части – обучающие и тестовые (рекомендованные в разных работах соотношения составляют 3:1 или 4:1). На обучающих данных проводится построение модели, на тестовых – проверяется качество классификации.

Для проведения статистической обработки данных рекомендован открытый пакет программ R: <https://www.r-project.org/> [26].

Пример анализа главных компонент

Для анализа взяты данные, находящиеся в открытом доступе [27]. Данные представляют собой таблицу радиомических признаков, вычисленных в PyRadiomics, включая фильтры постобработки (wavelet). Данные содержат 640 радиомических признаков для 105 пациентов с диагнозом «глиома низкой степени злокачественности». Также для каждого пациента имеется информация о наличии или отсутствии мутации 1p19q, которая влияет на выбор лечения. Задачей является установить, могут ли радиомические признаки предсказать факт наличия/отсутствия мутации.

Данные стандартизировали. В качестве разведочного провели анализ главных компонент. В таблице А.1 и на рисунке А.2 дана информация о распределении изменчивости по осям ГК.

Таблица А.1 – Распределение долей изменчивости по осям ГК

Номер ГК	% объясненной изменчивости	% кумулятивная изменчивость
1	38,16419	38,1642
2	14,75457	52,9188
3	8,00228	60,9210
4	5,70190	66,6229
5	3,61867	70,2416
6	2,78613	73,0277
7	2,40508	75,4328
8	2,16201	77,5948
9	1,76083	79,3557
10	1,58869	80,9444
11	1,37968	82,3240
12	1,21027	83,5343
13	0,99759	84,5319
14	0,94665	85,4785
15	0,90477	86,3833
16	0,77814	87,1615
17	0,77567	87,9371
18	0,71278	88,6499
19	0,67009	89,3200
20	0,62188	89,9419
21	0,60120	90,5431
22	0,54304	91,0861
23	0,51662	91,6027
24	0,46188	92,0646
25	0,41428	92,4789
26	0,40601	92,8849
27	0,38225	93,2672

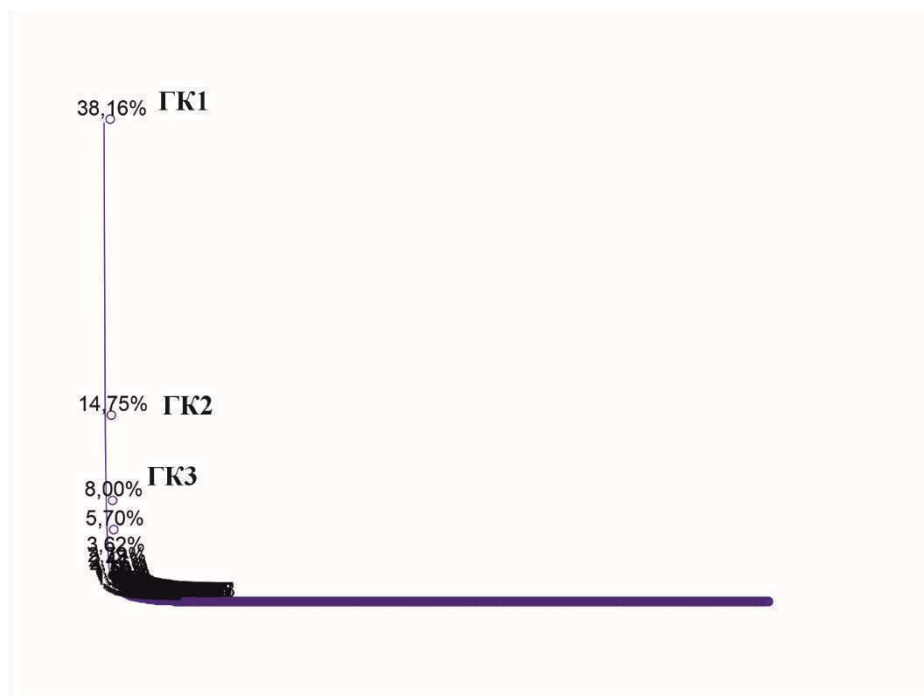


Рисунок А.2 – Графическое представление распределения долей изменчивости по осям главных компонент

На первые 10 главных компонент приходится 80 % всей изменчивости в выборке. Эти компоненты мы включаем в анализ.

Анализ корреляции радиомических признаков с первыми десятью ГК представлен в таблице А.2. Напомним, что главные компоненты – это комплексные признаки, состоящие из исходных радиомических признаков. На данном этапе исследователь решает две задачи:

1. Оценивает, какие ГК максимально эффективно разделяют группы в соответствии с поставленной целью текстурного анализа (в нашем примере – по наличию или отсутствию мутации).
2. Выявляет, какие именно радиомические признаки вносят максимальный вклад в эти ГК.

Таким образом, на основе анализа главных компонент мы находим те радиомические признаки, которые позволяют максимально дифференцировать интересующие группы. В таблице А.2 красным цветом выделены коэффициенты корреляции, которые больше либо равны 0,7. В примере эти корреляции мы считаем значимыми и отбираем соответствующие признаки в качестве значимых в формировании

соответствующих ГК. Чем выше порог для коэффициента корреляции (в идеале 1 по модулю), тем надежней получаемый результат. С другой стороны, если поставить такой жесткий порог, то мы рискуем не обнаружить значимых радиомических признаков в исследуемой выборке. Соответственно, при формировании дизайна исследования важно подобрать компромиссное значение данного параметра, позволяющее извлечь достаточное количество признаков с высокой статистической значимостью.

После отбора значимых радиомических признаков мы можем переходить к построению модели методом LASSO-регрессии.

Таблица А.2 – Коэффициенты корреляции радиомических признаков с первыми десятью ГК (красным выделены коэффициенты корреляции больше или равные 0,7)

Радиомический признак	ГК1	ГК2	ГК3	ГК4	ГК5
original_firstorder_10Percentile	-0,164946	-0,341776	-0,451799	-0,161396	0,185247
original_firstorder_90Percentile	-0,473538	-0,416887	-0,267069	-0,166794	0,412620
original_firstorder_Energy	0,517752	-0,760532	0,210046	0,015737	-0,139282
original_firstorder_Entropy	-0,561776	-0,166605	0,348026	0,111077	0,520563
original_firstorder_InterquartileRange	-0,410384	-0,142719	0,121370	0,066315	0,494286
original_firstorder_Kurtosis	0,055321	-0,347252	-0,306279	-0,106218	-0,464879
original_firstorder_Maximum	-0,669030	-0,427540	-0,076031	-0,079402	0,240189
original_firstorder_Mean	-0,335471	-0,421679	-0,423360	-0,182027	0,321990
original_firstorder_MeanAbsoluteDeviation	-0,580220	-0,335068	-0,016914	-0,073866	0,494199
original_firstorder_Median	-0,282880	-0,421063	-0,466414	-0,173197	0,304958
original_firstorder_Minimum	-0,104396	0,405387	-0,213465	-0,286387	0,050537
original_firstorder_Range	-0,584691	-0,620537	0,038985	0,073899	0,203082
original_firstorder_RobustMeanAbsoluteDeviation	-0,467166	-0,215071	0,054826	0,019728	0,512740
original_firstorder_RootMeanSquared	-0,349336	-0,426483	-0,415948	-0,182240	0,329186
original_firstorder_Skewness	-0,326214	0,231459	0,475677	0,084116	0,039700
original_firstorder_TotalEnergy	0,517752	-0,760532	0,210046	0,015737	-0,139282
original_firstorder_Uniformity	0,438543	-0,058773	-0,478791	-0,164912	-0,534292
original_firstorder_Variance	-0,605194	-0,413545	-0,083583	-0,144667	0,392438
original_glcm_Autocorrelation	-0,258912	-0,621791	-0,283503	0,004865	0,226087
original_glcm_ClusterProminence	-0,535330	-0,443878	-0,144560	-0,205523	0,277200

Радиомический признак	ГК1	ГК2	ГК3	ГК4	ГК5
original_glcm_ClusterShade	-0,036322	0,198045	0,493412	0,057281	-0,126494
original_glcm_ClusterTendency	-0,513067	-0,357592	0,018998	-0,041016	0,421207
original_glcm_Contrast	-0,797741	-0,342986	0,002671	-0,308338	0,081927
original_glcm_Correlation	0,416739	-0,112619	0,068720	0,243722	0,488590
original_glcm_DifferenceAverage	-0,844910	-0,154132	0,235171	-0,203266	0,102551
original_glcm_DifferenceEntropy	-0,831407	-0,101614	0,366569	-0,055722	0,180525
original_glcm_DifferenceVariance	-0,738114	-0,437485	-0,178414	-0,330395	0,096101
original_glcm_Id	0,715888	-0,202325	-0,519308	0,012834	-0,177672
original_glcm_Idm	0,670096	-0,249401	-0,545628	-0,017185	-0,197505
original_glcm_Idmn	0,438804	-0,479954	-0,056301	0,419983	0,083907
original_glcm_Idn	0,401581	-0,612298	-0,260556	0,302604	0,075568
original_firstorder_10Percentile	0,257757	-0,276631	-0,140713	0,048359	-0,234714
original_firstorder_90Percentile	0,162579	-0,352460	0,077611	-0,020474	-0,079326
original_firstorder_Energy	-0,000926	-0,176547	0,077331	0,033388	-0,040215
original_firstorder_Entropy	-0,141033	-0,184013	0,331058	-0,039056	0,155604
original_firstorder_InterquartileRange	-0,183009	-0,270394	0,380800	-0,120962	0,162380
original_firstorder_Kurtosis	0,271949	0,069153	-0,249408	-0,041519	-0,011697
original_firstorder_Maximum	0,115276	-0,066413	0,163120	0,086120	0,070438
original_firstorder_Mean	0,207413	-0,297227	-0,044608	-0,010603	-0,192890
original_firstorder_MeanAbsoluteDeviation	-0,026843	-0,275170	0,276264	-0,078120	0,107359
original_firstorder_Median	0,194564	-0,255284	-0,066000	-0,035066	-0,230657
original_firstorder_Minimum	0,074347	0,136151	0,007754	0,393410	0,089082
original_firstorder_Range	0,071270	-0,134663	0,151813	-0,123450	0,020720
original_firstorder_RobustMeanAbsoluteDeviation	-0,123773	-0,273549	0,354749	-0,105050	0,138600
original_firstorder_RootMeanSquared	0,204432	-0,300780	-0,035298	-0,012680	-0,184703
original_firstorder_Skewness	-0,182010	0,173460	0,178844	0,121624	0,289903
original_firstorder_TotalEnergy	-0,000926	-0,176547	0,077331	0,033388	-0,040215
original_firstorder_Uniformity	0,266681	0,059807	-0,307978	0,045895	-0,083753
original_firstorder_Variance	0,058389	-0,283110	0,211554	-0,043847	0,106410
original_glcm_Autocorrelation	0,121237	-0,366146	-0,088870	-0,282421	-0,246702
original_glcm_ClusterProminence	0,120889	-0,249610	0,123583	-0,005373	0,071252
original_glcm_ClusterShade	-0,039428	-0,034086	0,220301	-0,013028	0,295267
original_glcm_ClusterTendency	0,015160	-0,297433	0,299404	-0,071558	0,164632
original_glcm_Contrast	0,024438	-0,183866	0,076445	-0,039295	0,021166

Радиомический признак	ГК1	ГК2	ГК3	ГК4	ГК5
original_glcm_Correlation	-0,060822	-0,125084	0,269198	-0,059449	0,116843
original_glcm_DifferenceAverage	-0,075529	-0,190510	0,162249	-0,056662	0,077638
original_glcm_DifferenceEntropy	-0,132614	-0,145703	0,174375	-0,076663	0,049339
original_glcm_DifferenceVariance	0,090042	-0,136033	-0,027947	-0,020423	-0,061885
original_glcm_Id	0,226884	0,080603	-0,152340	0,001292	-0,055622
original_glcm_Idm	0,243844	0,050360	-0,138055	-0,015498	-0,038814
original_glcm_Idmn	0,150712	0,100129	-0,025409	-0,011199	-0,048970
original_glcm_Idn	0,211755	0,067791	-0,045801	-0,055052	-0,065816

Основные функции для проведения анализа главных компонент в программе R

Установка пакетов:

- > install.package(MASS)
- > install.package(ade4)

Анализ ГК:

- >df.pca <- princomp(df), где df (data frame) – данные

Построение графика вклада каждой ГК в общий разброс данных:

- >plot(df.pca, main="")

Построение графика распределения экземпляров в пространстве первых двух ГК:

- >biplot(df.pca)

Расчет нагрузок признаков на ГК:

- >loadings(df.pca)

Пример использования LASSO-регрессии

Отбор предикторов для бинарной логистической модели осуществлялся с помощью LASSO-регрессии, гиперпараметр λ определялся с использованием стандартного подхода 10-блочной кросс-валидации при

минимизации среднеквадратичного отклонения модели. Построение модели выполнено в R 4.3.2, пакет *caret* (Classification and Regression Training).

Выбор LASSO-регрессии обусловлен более гибкой регуляризацией входных параметров (первого и второго порядков). Альтернативная RIDGE-регуляризация вводит штраф за большие по модулю значения весов за счет добавления к функции потерь суммы квадратов всех весов. LASSO-регуляризация добавляет к минимизируемой функции потерь только сумму модулей всех весов, что является более уместным в условиях относительной равнозначности входных признаков. Также стоит отметить, что побочным свойством LASSO-регуляризации является обнуление весов для тех признаков, которые не влияют принципиально на результат предсказания. LASSO-регрессия выполняется путем минимизации функции потерь:

$$J_{LASSO} = RMSE + \lambda \cdot \|\beta\|_1,$$

где $RMSE$ – функция среднеквадратического отклонения модели, β – вектор-столбец всех коэффициентов модели, $\|\beta\|_1 = \sum_i |\beta_i|$ обозначает L1 (Манхэттенскую) норму.

Оптимальное значение гиперпараметра λ при минимальном значении среднеквадратичной ошибки (RMSE) было равным 0,07723 (рисунок А.3).

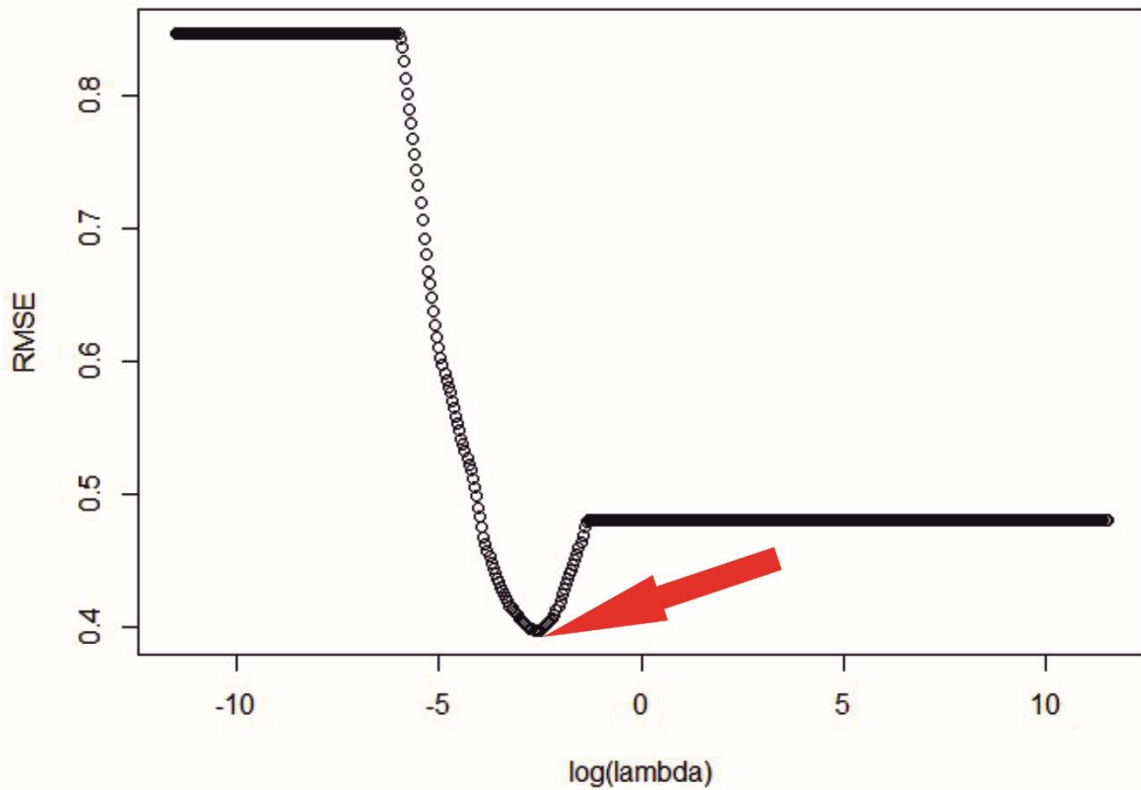


Рисунок А.3 – Подбор оптимального гиперпараметра λ путем минимизации RMSE (минимальное значение указано стрелкой)

Для оптимального гиперпараметра λ определены наиболее важные для модели радиомические признаки (рисунок А.4) по частоте их встречаемости в качестве значимых для классификации исходов в отношении наличия или отсутствия мутации 1p19. Отобраны признаки с ненулевыми коэффициентами в регрессионной модели (таблица А.3), наиболее важные для классификации.

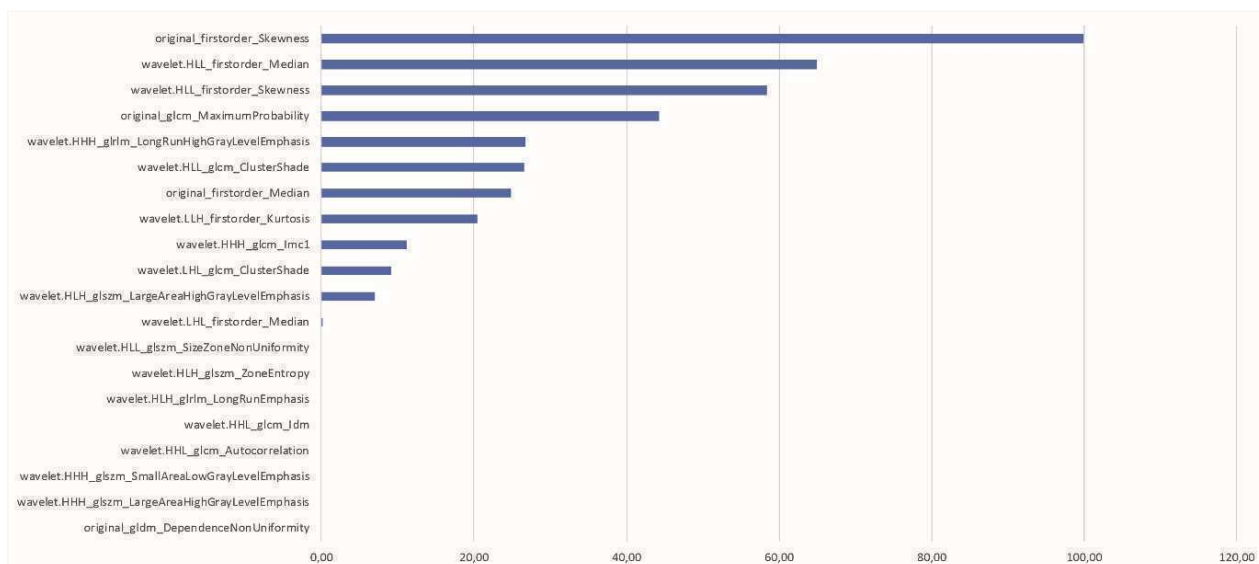


Рисунок А.4 – Номограмма важности радиомических признаков в модели, показаны первые 20 признаков

Таблица А.3 – Значения параметров модели

Признак	Коэффициент корреляции
Intercept	0,64706
original_firstorder_Skewness	0,09737
wavelet.HLL_firstorder_Median	0,06332
wavelet.HLL_firstorder_Skewness	-0,05688
original_glcm_MaximumProbability	-0,04309
wavelet.HHH_glrIm_LongRunHighGrayLevelEmphasis	-0,02600
wavelet.HLL_glcm_ClusterShade	-0,02584
original_firstorder_Median	-0,02415
avelet.LLH_firstorder_Kurtosis	-0,01993
wavelet.HHH_glcm_Imc1	0,01083
wavelet.LHL_glcm_ClusterShade	-0,00898
wavelet.HLH_glszm_LargeAreaHighGrayLevelEmphasis	-0,00686

Из полученных данных следует, что с наличием исхода более всего коррелирует радиомический показатель «Intercept».

Основные функции для построения модели в R

Установка пакета caret:

```
> install.packages(caret)
```

Фиксация случайного числа, служащего начальной точкой для генерации набора случайных чисел:

```
> set.seed(1985)
```

Разделение данных и создание индексной матрицы:

```
> index<-createDataPartition(df$patientID, p=.8, list=FALSE, times=1),
```

где `df` – данные, `patientID` – столбец с идентификационными номерами пациентов, `p=.8` – доля выделяемых экземпляров, `times=1` – число разделений.

Разделение данных на обучающую и тестовую выборки:

```
> train_df<-df[index,]
```

```
> test_df<-df[-index,]
```

K-блочная кросс-валидация

Спецификация 10-блочной кросс-валидации:

```
> ctrlspecs<-trainControl(method="cv", number=10,
```

```
savePredictions="all"),
```

где `method="cv"` – использованный метод (`cv` – cross-validation), `number=10` – число блоков, на которые разделяются данные.

Спецификация и обучение Lasso-модели

Создание вектора потенциальных значений λ :

```
> lambda_vector<-10^seq(5, -5, length=500)
```

Спецификация Lasso-модели:

```
> model1<-train(y ~ .,  
               data=train_df1,  
               preProcess=c("center","scale"),  
               method="glmnet",  
               tuneGrid=expand.grid(alpha=1, lambda=lambda_vector),  
               trControl=ctrlspecs,  
               na.action=na.omit),
```

где y – столбец с переменной наличия или отсутствия мутации, $preProcess=c("center","scale")$ – предобработка данных, если это не было сделано ранее, $method="glmnet"$ – метод обучения модели.

Выбор оптимального параметра (λ):

```
> model1$bestTune$lambda
```

Оценка коэффициентов модели:

```
> coef(model1$finalModel, model1$bestTune$lambda)
```

Выбор числа отображаемых знаков после запятой:

```
> round(coef(model1$finalModel, model1$bestTune$lambda),5)
```

Построение распределения $\log(\lambda)$ vs. RMSE:

```
> plot(log(model1$results$lambda),  
       model1$results$RMSE,  
       xlab="log(lambda)",  
       ylab="RMSE")
```

Важность переменных в модели:

```
> varImp(model1)
```

Визуализация важности переменных в модели

Установка пакета ggplot2:

```
> install.packages("ggplot2")
```

Построение номограммы важности радиомических признаков:

```
> ggplot(varImp(model1))
```

Оценка работы модели на тестовой выборке:

```
> prediction1 <- predict(model1, newdata = test_df)
```

Контроль качества построенной модели

Качество классификации построенной модели, созданной на обучающей выборке, должно быть проверено на тестовой выборке. Исходные данные должны быть разделены на обучающую выборку – на которой модель строится, и на тестовую выборку – на которой модель тестируется. Обучающая и тестовая выборки не должны пересекаться. Качество построенной модели также должно быть проверено на внешних данных – данных, не имеющих отношения к обучающей и тестовой выборке, например, полученных из другого медицинского учреждения, т. е. должна быть проведена внешняя валидация. Таким образом проверяется, насколько полученная модель может быть обобщена, и исключается риск систематических ошибок, например риск переобучения модели.

Наиболее часто используемыми показателями качества построенной модели является ее чувствительность (sensitivity), специфичность (specificity) и площадь под кривой ошибок ROC AUC (ROC – receiver operating characteristic, AUC – area under curve). Расчет чувствительности, специфичности и некоторых других оценок качества модели можно выполнять с помощью функции пакета caret:

```
> confusionMatrix(tab),
```

где tab – таблица сопряженности следующего вида:

		Референс-тест	
		Положительный результат	Отрицательный результат
Индекс-тест	Положительный результат	a	b
	Отрицательный результат	c	d

a – число истинно положительных результатов, b – число ложноположительных результатов, c – число ложноотрицательных результатов, d – число истинно отрицательных результатов,

получаемое на выходе значение чувствительности рассчитывается по формуле $a/(a+c)$,

получаемое на выходе значение специфичности рассчитывается по формуле $d/(b+d)$.

Для использования полученной прогностической модели на практике может быть сформирована отчетность по ее созданию и тестированию, соответствующая стандартам. Для формирования такой отчетности написано руководство TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis), содержащее 22 контрольных пункта рекомендаций для повышения качества отчетности по исследованиям, направленным на создание прогностической модели. Перед началом разработки прогностической модели целесообразно обратиться к этому руководству. Контрольный перечень TRIPOD на английском языке доступен по адресу: <https://www.tripod-statement.org>.

Описание радиомических признаков

Таблица Б.1 – Признаки первого порядка (First order features)

Название признака	Оригинальное описание признака	Описание признака на русском языке
Среднее (Mean)	The average gray level intensity within the ROI.	Средняя интенсивность уровня серого в пределах области интереса (ROI).
Медиана (Median)	The median gray level intensity within the ROI.	Медианное значение интенсивности уровня серого в пределах области интереса (ROI).
Минимум (Minimum)	The minimal gray level intensity within the ROI.	Минимальное значение интенсивности уровня серого в пределах области интереса (ROI).
Максимум (Maximum)	The maximum gray level intensity within the ROI.	Максимальное значение соответственно.
10% процентиль (10th percentile)	The 10th percentile of data sample.	Это первые 10 % значений на гистограмме распределения интенсивностей уровней серого.
90% процентиль (90th percentile)	The 90th percentile of data sample.	Это определенная часть выборки данных, охватывающая указанный процент значений выборки, в данном случае 90 %.
Интерквартильный интервал (Interquartile Range)	Here P25 and P75 are the 25th and 75th percentile of the image array, respectively.	Здесь интервал между 75 % процентилем и 25 % процентилем.
Интервал (Range)	The range of gray values in the ROI.	Разница между максимальным и минимальным значениями интенсивности уровня серого в пределах области интереса (ROI).

Название признака	Оригинальное описание признака	Описание признака на русском языке
Асимметрия кривой (Skewness)	Skewness measures the asymmetry of the distribution of values about the Mean value. Depending on where the tail is elongated and the mass of the distribution is concentrated, this value can be positive or negative.	Показатель асимметрии распределения значений относительно среднего значения. В зависимости от того, где вытянут «хвост» кривой, эта величина может быть положительной или отрицательной.
Коэффициент эксцесса (Kurtosis)	Kurtosis is a measure of the ‘peakedness’ of the distribution of values in the image ROI. A higher kurtosis implies that the mass of the distribution is concentrated towards the tail(s) rather than towards the mean. A lower kurtosis implies the reverse: that the mass of the distribution is concentrated towards a spike near the Mean value.	Это мера «остроконечности» распределения значений. Более высокий эксцесс означает, что масса распределения сосредоточена у «хвостов», а не у среднего значения. Меньший эксцесс подразумевает обратное: масса распределения концентрируется вблизи среднего значения.
Дисперсия (Variance)	Variance is the mean of the squared distances of each intensity value from the Mean value. This is a measure of the spread of the distribution about the mean.	Это мера разброса распределения относительно среднего значения. Рассчитывается как среднее квадратов расстояний каждого значения интенсивности пикселей от среднего значения. По определению дисперсия равна квадрату среднеквадратичного отклонения.
Среднеквадратичное отклонение (Standard Deviation)	Standard Deviation measures the amount of variation or dispersion from the Mean Value.	По определению дисперсия равна квадрату среднеквадратичного отклонения. Значения дисперсии и среднеквадратичного отклонения коррелированы, поэтому PyRadiomic в качестве результата выдает только значение дисперсии и не выдает значение среднеквадратичного отклонения (хотя эта величина в англоязычном мануале к программе описывается как отдельный признак).
Однородность (Uniformity)	Uniformity is a measure of the sum of the squares of each intensity value. This is a measure of the homogeneity of the image array, where a greater uniformity implies a greater homogeneity or a smaller range of discrete intensity values.	Это мера гомогенности массива изображений. Большие значения однородности подразумевают большую гомогенность или меньший диапазон дискретных значений интенсивности. Вычисляется как сумма квадратов каждого значения интенсивности.

Название признака	Оригинальное описание признака	Описание признака на русском языке
Среднее абсолютное отклонение (Mean Absolute Deviation (MAD))	Mean Absolute Deviation is the mean distance of all intensity values from the Mean Value of the image array.	Среднее расстояние всех значений интенсивности от среднего значения массива изображений.
Робастное среднее абсолютное отклонение (Robust Mean Absolute Deviation (rMAD))	Robust Mean Absolute Deviation is the mean distance of all intensity values from the Mean Value calculated on the subset of image array with gray levels in between, or equal to the 10th and 90th percentile.	Это среднее расстояние всех значений интенсивности от среднего значения, рассчитанное для подмножества массива изображений с уровнями серого между ними или равными 10-му и 90-му перцентилям.
Корень среднеквадратичного значения (Root Mean Squared (RMS))	RMS is the square-root of the mean of all the squared intensity values. It is another measure of the magnitude of the image values. This feature is volume-confounded, a larger value of c increases the effect of volume-confounding.	Квадратный корень из среднего значения всех квадратов значений интенсивности.
Энергия (Energy)	Energy is a measure of the magnitude of voxel values in an image. A larger values implies a greater sum of the squares of these values.	Суммарная мера величины значений вокселей в изображении. Большие значения подразумевают большую сумму квадратов этих значений.
Общая энергия (Total Energy)	Total Energy is the value of Energy feature scaled by the volume of the voxel in cubic mm.	Это энергия, масштабированная по объему вокселя в кубических миллиметрах.
Энтропия (Entropy)	Entropy specifies the uncertainty/randomness in the image values. It measures the average amount of information required to encode the image values.	Мера неопределенности/случайности значений изображения. Этот признак измеряет средний объем информации, необходимый для кодирования значений изображения.

Таблица Б.2 – Признаки второго порядка (Second order features)

Название признака	Оригинальное описание признака	Описание признака на русском языке
Характеристики матрицы совпадения уровней серого (Gray Level Co-occurrence Matrix, GLCM)	GLCM of size $N_g \times N_g$ describes the second-order joint probability function of an image region constrained by the mask and is defined as $P(i,j \delta,\theta)$. The $(i,j)^{th}$ element of this matrix represents the number of times the combination of levels i and j occur in two pixels in the image, that are separated by a distance of δ pixels along angle θ .	GLCM характеризует текстуру на основе количества пар пикселей с определенными значениями интенсивности, расположенных в определенных пространственных отношениях.
Контраст (Contrast)	Contrast is a measure of the local intensity variation, favoring values away from the diagonal. A larger value correlates with a greater disparity in intensity values among neighboring voxels.	Мера локального изменения интенсивности. Большое значение контраста коррелирует с большим несоответствием значений интенсивности между соседними вокселями.
Корреляция (Correlation)	Correlation is a value between 0 (uncorrelated) and 1 (perfectly correlated) showing the linear dependency of gray level values to their respective voxels in the GLCM.	Это значение, показывающее линейную зависимость значений уровня серого от соответствующих вокселей в GLCM. Изменяется от 0 (некоррелированное) до 1 (полностью коррелированное).
Разница средних (Difference Average)	Difference Average measures the relationship between occurrences of pairs with similar intensity values and occurrences of pairs with differing intensity values.	Величина, измеряющая взаимосвязь между появлением пар с одинаковыми значениями интенсивности и появлением пар с разными значениями интенсивности.
Сумма квадратов (Sum of Squares)	Sum of Squares or Variance is a measure in the distribution of neighboring intensity level pairs about the mean intensity level in the GLCM.	Мера распределения соседних пар уровней интенсивности относительно среднего уровня интенсивности в GLCM.
Характеристики матрицы размера зон уровня серого (Gray Level Size Zone Matrix, GLSZM).	GLSZM quantifies gray level zones in an image. A gray level zone is defined as the number of connected voxels that share the same gray level intensity.	GLSZM определяет количественно зоны уровня серого на изображении. Зона уровня серого определяется как количество связанных вокселей, которые имеют одинаковую интенсивность уровня серого.

Название признака	Оригинальное описание признака	Описание признака на русском языке
Выраженность небольших зон (Small Area Emphasis, SAE)	SAE is a measure of the distribution of small size zones, with a greater value indicative of more smaller size zones and more fine textures.	Мера распределения зон небольшого размера. Большее значение указывает на большее количество зон меньшего размера и более тонкую текстуру области интереса.
Выраженность больших зон (Large Area Emphasis, LAE)	LAE is a measure of the distribution of large area size zones, with a greater value indicative of more larger size zones and more coarse textures.	Мера распределения зон большого размера. Большее значение указывает на более крупные зоны и более грубую текстуру.
Процент зон (Zone Percentage, ZP)	ZP measures the coarseness of the texture by taking the ratio of number of zones and number of voxels in the ROI.	Измеряет грубость текстуры, беря соотношение количества зон и количества вокселей в области интереса.
Энтропия зон (Zone Entropy, ZE)	ZE measures the uncertainty/randomness in the distribution of zone sizes and gray levels. A higher value indicates more heterogeneity in the texture patterns.	Измеряет неопределенность/случайность в распределении размеров зон и уровней серого. Более высокое значение указывает на большую неоднородность текстурных рисунков.
Матрица длин серий уровней серого (Gray Level Run Length Matrix, GLRLM)	GLRLM quantifies gray level runs, which are defined as the length in number of pixels, of consecutive pixels that have the same gray level value.	Количественно определяет длины наборов последовательных пикселей, имеющих одинаковое значение уровня серого.
Неравномерность уровня серого (Gray Level Non-Uniformity, GLN)	GLN measures the similarity of gray-level intensity values in the image, where a lower GLN value correlates with a greater similarity in intensity values.	Мера сходства значений интенсивности уровней серого на изображении. Более низкое значение GLN коррелирует с большим сходством значений интенсивности.
Неравномерность длины ряда (Run Length Non-Uniformity, RLN)	RLN measures the similarity of run lengths throughout the image, with a lower value indicating more homogeneity among run lengths in the image.	Мера сходства длин серий по всему изображению. Более низкое значение RLN указывает на большую однородность длин серий на изображении.
Дисперсия уровня серого (Gray Level Variance, GLV)	GLV measures the variance in gray level intensity for the runs.	Мера разницы в интенсивности уровня серого для рядов.

Название признака	Оригинальное описание признака	Описание признака на русском языке
Дисперсия рядов (Run Variance, RV)	RV is a measure of the variance in runs for the run lengths.	Мера дисперсии рядов по их длинам.
Матрица различий соседних оттенков серого (Neighbouring Gray Tone Difference Matrix, NGTDM)	NGTDM quantifies the difference between a gray value and the average gray value of its neighbours within distance.	Количественно определяет разницу между значением серого для вокселя и средним значением серого его соседних вокселей на определенном расстоянии.
Грубость (Coarseness)	Coarseness is a measure of average difference between the center voxel and its neighbourhood and is an indication of the spatial rate of change. A higher value indicates a lower spatial change rate and a locally more uniform texture.	Это мера средней разницы между центральным вокселем и его окрестностями, а также показатель пространственной скорости изменений. Более высокое значение указывает на более низкую скорость пространственных изменений и локально более однородную текстуру.
Контрастность (Contrast)	Contrast is a measure of the spatial intensity change, but is also dependent on the overall gray level dynamic range. Contrast is high when both the dynamic range and the spatial change rate are high, i.e. an image with a large range of gray levels, with large changes between voxels and their neighbourhood.	Это мера пространственного изменения интенсивности, но она также зависит от общей динамики диапазона уровней серого. Контраст высок, когда высоки как динамический диапазон, так и скорость пространственных изменений, то есть на изображении с большим диапазоном уровней серого, с большими изменениями между вокселями и их окрестностями.
Занятость (Busyness)	A measure of the change from a pixel to its neighbour. A high value for busyness indicates a 'busy' image, with rapid changes of intensity between pixels and its neighbourhood.	Мера изменения пикселя относительно соседнего. Высокое значение указывает на «занятое» изображение с быстрыми изменениями интенсивности между пикселями и их окрестностями.
Сложность (Complexity)	An image is considered complex when there are many primitive components in the image, i.e. the image is non-	Изображение считается сложным, если в нем много примитивных компонентов, т. е. изображение

Название признака	Оригинальное описание признака	Описание признака на русском языке
	uniform and there are many rapid changes in gray level intensity.	неоднородно и имеется множество быстрых изменений интенсивности уровня серого.
Сила (Strength)	Strength is a measure of the primitives in an image. Its value is high when the primitives are easily defined and visible, i.e. an image with slow change in intensity but more large coarse differences in gray level intensities.	Это мера примитивов изображения. Значение высоко, когда примитивы легко определяются и видны, то есть изображение имеет медленное изменение интенсивности, но при этом содержит более грубые различия в интенсивности уровней серого.
Матрица зависимости уровней серого (Gray Level Dependence Matrix, GLDM)	GLDM quantifies gray level dependencies in an image. A gray level dependency is defined as the number of connected voxels within distance that are dependent on the center voxel.	Количественно определяет зависимости уровней серого в изображении. Зависимость уровня серого определяется как количество вокселей, которые зависят от центрального вокселя.
Выраженность малых зависимостей (Small Dependence Emphasis, SDE)	A measure of the distribution of small dependencies, with a greater value indicative of smaller dependence and less homogeneous textures.	Мера распределения небольших зависимостей. Большее значение указывает на меньшую зависимость и на менее однородную текстуру.
Выраженность малых зависимостей (Large Dependence Emphasis, LDE)	A measure of the distribution of large dependencies, with a greater value indicative of larger dependence and more homogeneous textures.	Мера распределения больших зависимостей. Большее значение указывает на большую зависимость и большую однородность текстуры.
Неравномерность уровня серого (Gray Level Non-Uniformity, GLN)	Measures the similarity of gray-level intensity values in the image, where a lower GLN value correlates with a greater similarity in intensity values.	Измеряет сходство значений интенсивности уровней серого на изображении, где более низкое значение GLN коррелирует с большим сходством значений интенсивности.
Зависимость неравномерности (Dependence Non-Uniformity, DN)	Measures the similarity of dependence throughout the image, with a lower value indicating more homogeneity among dependencies in the image.	Измеряет сходство зависимостей по всему изображению. Более низкое значение указывает на большую однородность зависимостей в изображении.

Таблица Б.3 – Признаки формы (Shape features)

Название признака	Оригинальное описание признака	Описание признака на русском языке
Удлинение (Elongation)	Elongation shows the relationship between the two largest principal components in the ROI shape. The values range between 1 (where the cross section through the first and second largest principal moments is circle-like (non-elongated)) and 0 (where the object is a maximally elongated: i.e. a 1 dimensional line).	Рассчитывается как отношение двух наиболее длинных осей области интереса. Значение варьирует от 0 (когда объект плоский или сегментация представляет собой один срез) до 1 (когда объект сферический).
Плоскостность (Flatness)	Flatness shows the relationship between the largest and smallest principal components in the ROI shape. The values range between 1 (non-flat, sphere-like) and 0 (a flat object, or single-slice segmentation).	Рассчитывается как отношение самой короткой и самой длинной осей области интереса. Значение варьирует от 0 (когда объект представляет собой линию) до 1 (когда объект представляет собой сферу, т. е. оси равны между собой).
Наименьшая длина оси (Least Axis Length)	This feature yield the smallest axis length of the ROI-enclosing ellipsoid. In case of a 2D segmentation, this value will be 0.	Наименьшая длина оси эллипсоида, охватывающего область интереса. В случае 2D-сегментации значение признака равно 0.
Наибольшая длина оси (Major Axis Length)	This feature yield the largest axis length of the ROI-enclosing ellipsoid.	Наибольшая длина оси эллипсоида, охватывающего область интереса.
Максимальный 2D диаметр, колонка (Maximum 2D diameter Column)	Maximum 2D diameter (Column) is defined as the largest pairwise Euclidean distance between tumor surface mesh vertices in the row-slice (usually the coronal) plane.	Наибольшее расстояние между двумя вершинами сетки поверхности опухоли в корональной плоскости.
Максимальный 2D диаметр, ряд (Maximum 2D diameter Row)	Maximum 2D diameter (Row) is defined as the largest pairwise Euclidean distance between tumor surface mesh vertices in the column-slice (usually the sagittal) plane.	Наибольшее расстояние между двумя вершинами сетки поверхности опухоли в сагиттальной плоскости.
Максимальный 2D диаметр, срез (Maximum 2D diameter Slice)	Maximum 2D diameter (Slice) is defined as the largest pairwise Euclidean distance between tumor surface mesh vertices in the row-column (generally the axial) plane.	Наибольшее евклидово расстояние между двумя вершинами сетки поверхности опухоли в аксиальной плоскости.

Название признака	Оригинальное описание признака	Описание признака на русском языке
Максимальный 3D диаметр (Maximum 3D diameter)	Maximum 3D diameter is defined as the largest pairwise Euclidean distance between tumor surface mesh vertices.	Наибольшее евклидово расстояние между двумя вершинами сетки поверхности опухоли.
Объем сетки (Mesh Volume)	This feature is calculated from the triangle mesh of the RO.	Рассчитывается на основе треугольной сети области интереса.
Длина малой оси (Minor Axis Length)	This feature yield the second-largest axis length of the ROI-enclosing ellipsoid.	Вторая по длине ось эллипсоида, охватывающего область интереса.
Сферичность (Sphericity)	Sphericity is a measure of the roundness of the shape of the tumor region relative to a sphere. It is a dimensionless measure, independent of scale and orientation. The value range is between 0 and 1, where a value of 1 indicates a perfect sphere (a sphere has the smallest possible surface area for a given volume, compared to other solids).	Мера округлости формы опухолевой области относительно сферы. Величина не зависит от единиц измерений, масштаба и ориентации в пространстве. Это безразмерная мера, независимая от масштаба и ориентации. Диапазон значений изменяется от 0 до 1, где значение 1 указывает на идеальную сферу.
Площадь поверхности (Surface Area)	This feature is calculated as sum of squares of each triangle in the mesh is calculated.	Рассчитывается как сумма площадей всех поверхностных треугольников сети.
Соотношение площади к объему (SurfaceVolumeRatio или Surface Area to Volume ratio)	Here, a lower value indicates a more compact (sphere-like) shape. This feature is not dimensionless, and is therefore (partly) dependent on the volume of the ROI.	Рассчитывается как отношение площади поверхности (Surface Area) к объему сетки (Mesh Volume). Более низкое значение указывает на более компактную (сферическую) форму.
Воксельный объем (Voxel Volume)	The volume of the ROI is approximated by multiplying the number of voxels in the ROI by the volume of a single voxel. This is a less precise approximation of the volume and is not used in subsequent features.	Объем области интереса аппроксимируется путем умножения количества вокселей в области интереса на объем одного вокселя. Это менее точное приближение объема по сравнению с Mesh Volume и не используется в других функциях.

Серия «Лучшие практики лучевой и инструментальной диагностики»

Выпуск 145

Авторы-составители:

*Васильев Юрий Александрович
Нанова Ольга Геннадьевна
Романенко Мария Олеговна
Гележе Павел Борисович
Гончар Анна Павловна
Мынко Олег Игоревич
Блохин Иван Андреевич
Решетников Роман Владимирович*

**РАДИОМИЧЕСКИЙ АНАЛИЗ
ПОСЛЕ ПОЛУЧЕНИЯ МАСКИ ОБЛАСТИ ИНТЕРЕСА:
РУКОВОДСТВО ДЛЯ ВРАЧА-РЕНТГЕНОЛОГА**

Методические рекомендации

Отдел координации научной деятельности ГБУЗ «НПКЦ ДиТ ДЗМ»
Технический редактор В. П. Гамарина
Компьютерная верстка Е. Д. Бугаенко

ГБУЗ «НПКЦ ДиТ ДЗМ»
127051, г. Москва, ул. Петровка, д. 24, стр. 1