



**НМИЦ  
ОНКОЛОГИИ**  
им.Н.Н. Блохина

## **Датасеты: практика и способы сбора, разметки, хранения и защиты от несанкционированного использования**

Докладчик:

**Степан Андреевич Побоженский**

Руководитель отдела компьютерного зрения и  
персонализированной медицины

Исследовательского центра в сфере искусственного  
интеллекта в здравоохранении

ФГБУ «НМИЦ онкологии им. Н.Н.Блохина» Минздрава России

## Основные направления деятельности

- Разработка персонализированных противоопухолевых вакцин.
- Разработка предиктивных моделей на основе первичных медицинских данных для прогнозирования появления или течения заболеваний.
- Разработка моделей с технологиями и инструментами ИИ для диагностики МРТ и КТ-изображений.
- Разработка первого в России программного комплекса по автоматизированной разметке органов и помощи врачу в формировании плана лечения в лучевой терапии.
- Дополнительное профессиональное образования по вопросам разработки и внедрения технологий с искусственным интеллектом в медицине.

# Диагностика и лечение онкологических заболеваний



НМИЦ  
ОНКОЛОГИИ  
им.Н.Н.Блохина

1

Применение технологий искусственного интеллекта для задач оконтуривания и оценки данных в ПЭТ/КТ

2

Применение технологий искусственного интеллекта для задач оконтуривания и оценки данных в лучевой терапии

3

Исследование и разработка методов интеллектуального анализа изображений для повышения качества рентгенологической диагностики новообразований в почках

4

Разработка фреймворка для аннотирования компрессионных рентгеновских изображений ткани молочных желез (маммограмм)

5

Исследование и разработка методов интеллектуального анализа изображений для диагностики опухолей печени



# 1

## Применение технологий искусственного интеллекта для задач оконтуривания и оценки данных в ПЭТ/КТ

### Проблематика



- Для ПЭТ-КТ диагностики рака простаты и лимфом важно оценить динамику на основании суммарного опухолевого объема, что вручную делать трудоемко и сложно.
- В широкой клинической практике используются методики автооконтуривания на основе математических алгоритмов, но получаются не совсем четкие оконтуренные структуры, которые врачу приходится редактировать вручную

### Планируемое решение

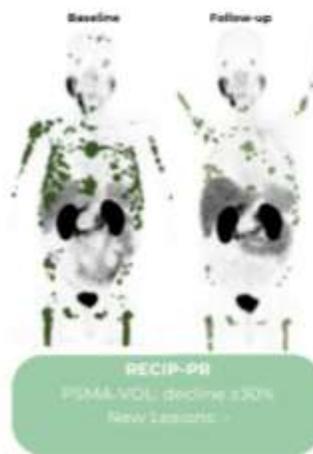


- Разработка программы с технологией ИИ для оконтуривания и расчета суммарного объема опухоли (особенно при лимфомах и ЗНО предстательной железы) .

### Прогнозируемый результат



- стандартизация оконтуривания, а значит и врачебного заключения;
- сокращение трудозатрат врача радиолога, повышение производительности
- объективизация оценки динамики течения онкологического заболевания
- снижение стоимости исследования



## 2

# Применение технологий искусственного интеллекта для задач оконтуривания и оценки данных в лучевой терапии

### Проблематика

Критически большие временные затраты на разметку органов



и зоны интереса радиотерапевтом

- Глиальные опухоли (30-35 минут)
- Рак пищевода (40-45 минут)
- Рак легкого (40-45 минут)
- СТЛТ метастазов в печени (40-45 минут)
- Рак поджелудочной железы (40-45 минут)
- Опухоли органов головы и шеи (60 минут и более...)

### Планируемое решение

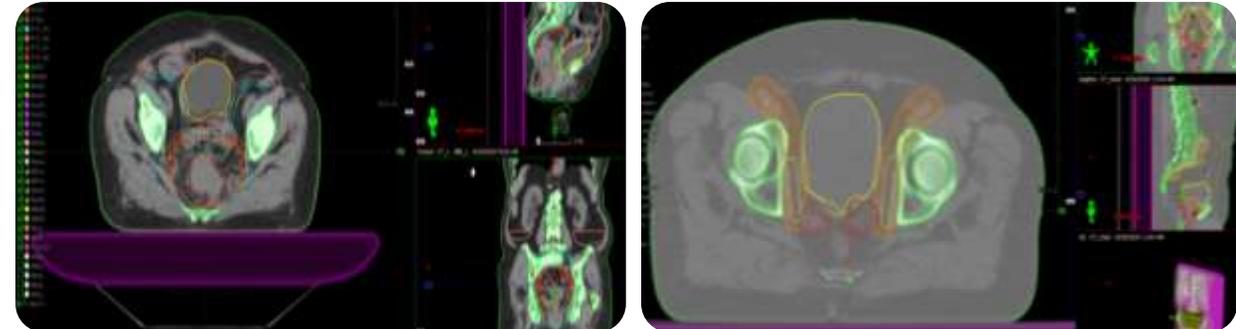
- Разработка новых подходов и решений для задач оптимизации и увеличения точности оконтуривания в лучевой терапии
- Разработка библиотеки алгоритмически-программных модулей для задач оконтуривания в лучевой терапии.



### Прогнозируемый результат



- Сокращение времени на разметку органов > 30%
- Повышение точности разметки
- Минимизация рисков для пациента связанная с ошибками в разметке или расчетах



3

Исследование и разработка методов интеллектуального анализа изображений для повышения качества рентгенологической диагностики новообразований в почках

#### Проблематика

- Из-за схожести признаков злокачественных и доброкачественных опухолей почек, ошибки диагностики достигают 20%. 

Согласно клиническим рекомендациям (2023), пациентам с солидными новообразованиями почек, накапливающими КВ, выявленными по КТ/МРТ, возможно хирургическое лечение без морфологической верификации.

#### Планируемое решение

Разработка программного обеспечения с применением технологий искусственного интеллекта для обработки и анализа КТ исследований почек. 

#### Прогнозируемый результат

Создание автоматизированного программного комплекса позволит: 

- Снизить вероятность совершения диагностических ошибок
- Повысить выявляемость клинически значимых злокачественных новообразований почек
- Повышение точности диагностики поможет избежать необоснованных операций.



## Проблематика

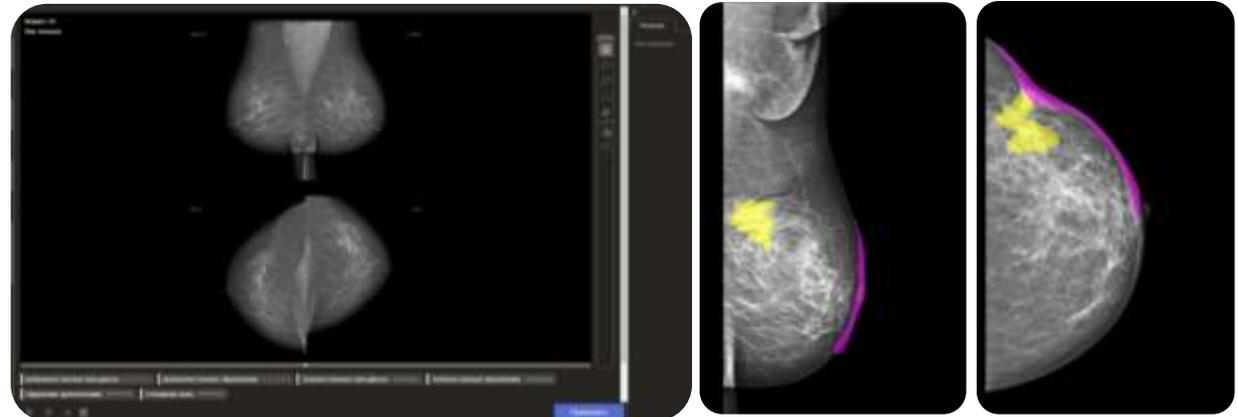
- Сложность процедуры выбора и пропорционального распределения по классам исследований для разметки 
- Сложность контроля корректности разметки изображений
- Ручное агрегирование размеченных исследований из различных учреждений
- Сложность защиты размеченных исследований от несанкционированного доступа и использования

## Планируемое решение

Разработка программного комплекса для распределения, хранения двойного контроля, разметки и защиты медицинских данных 

## Прогнозируемый результат

- Повышение качества и репрезентативности данных 
- Гарантированная корректность разметки
- Эффективное управление данными множества учреждений
- Снижение операционных затрат и ускорение процессов



## Проблематика

Трудности в выявлении и дифференциальной диагностике различных новообразований печени обусловлены:



- Сложностью обнаружения мелких опухолей — очаги <1 см легко пропустить на фоне цирроза или жировой инфильтрации.
- Нетипичным контрастированием.
- Фоновыми изменениями печени — цирроз, фиброз или стеатоз маскируют новообразования.
- Риском инвазивных процедур — биопсия опасна при низких тромбоцитах, но часто необходима для верификации диагноза.

## Планируемое решение

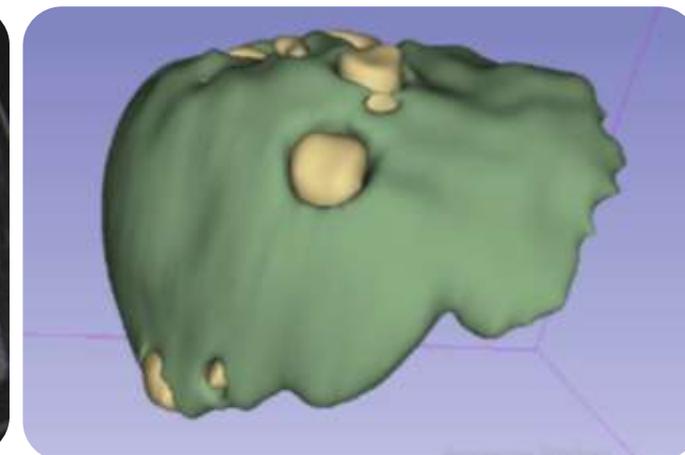
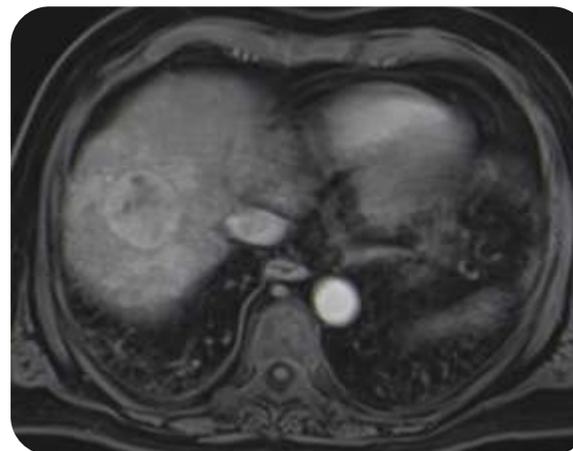


Разработка и внедрение в клиническую практику программного продукта поддержки принятия врачебных решений в диагностике опухолей печени на основе технологий искусственного интеллекта.

## Прогнозируемый результат

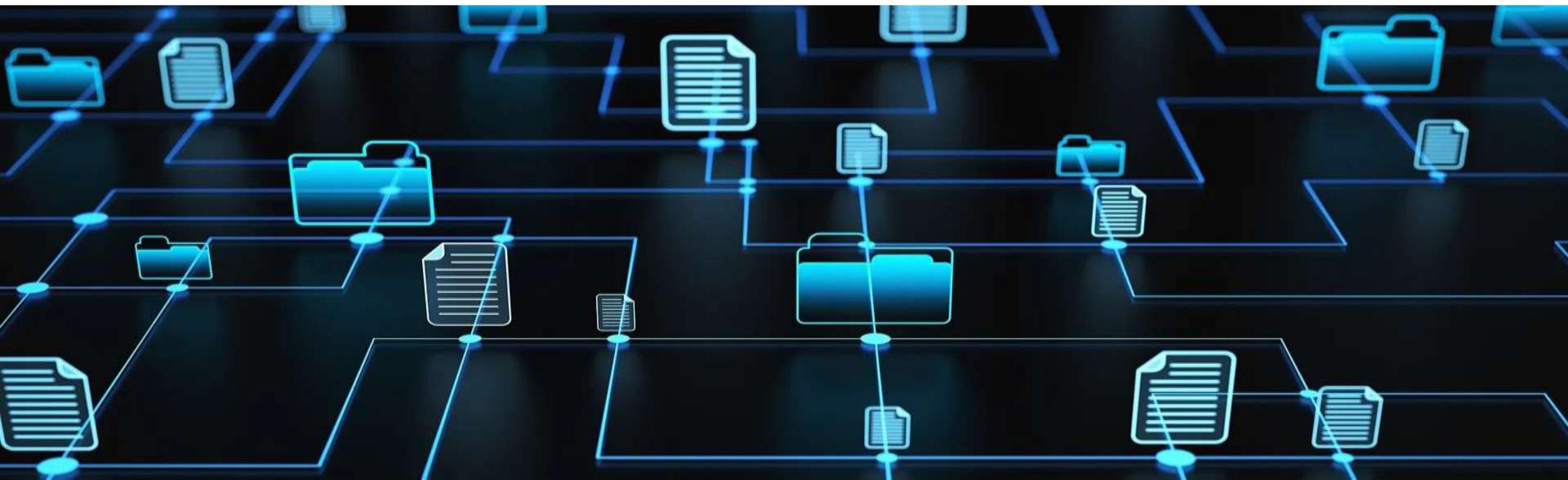


- Увеличение выявляемости малых опухолей (<1 см) даже на фоне цирроза.
- Сокращение времени диагностики
- Уменьшение числа инвазивных биопсий
- Повышение точности дифференциации доброкачественных и злокачественных образований



# Набор данных (Датасет)

Согласно ГОСТ 20886-65 Набор данных (Data set) – это идентифицированная совокупность физических записей, организованная одним из установленных в системе обработки данных способов и представляющая файлы или части файлов в среде хранения



# СБОР МЕДИЦИНСКИХ ИССЛЕДОВАНИЙ ДЛЯ ФОРМИРОВАНИЯ НАБОРА ДАННЫХ В НМИЦ ОНКОЛОГИИ ИМ. Н. Н. БЛОХИНА

- Формирование реестров исследований. Выбираются конкретные случаи в необходимых пропорциях под каждый набор данных;
- Случаи только подтвержденные гистологическими исследованиями;
- Автоматическая анонимизация DICOM.



# Разметка медицинских исследований

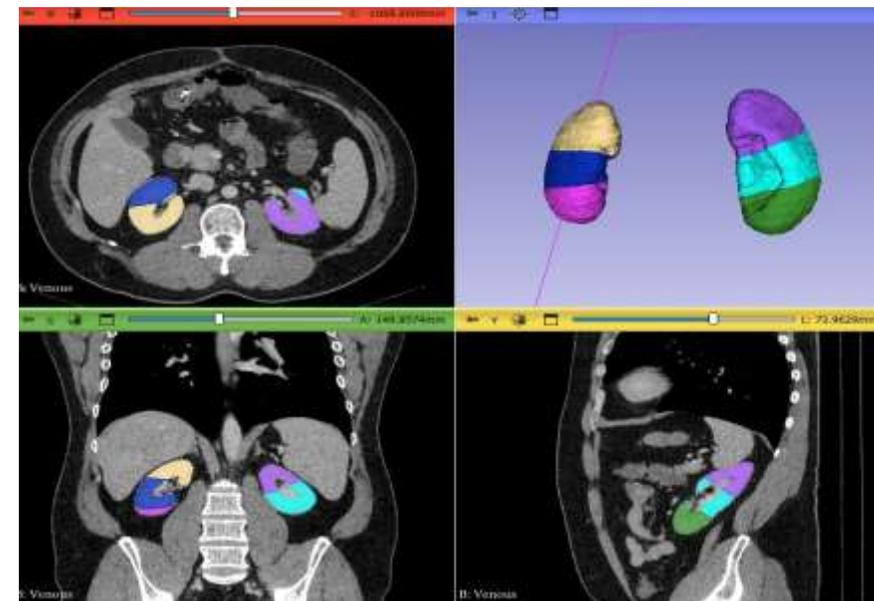
1

Собственный фреймворк для аннотирования компрессионных рентгеновских изображений ткани молочных желез.

В функционал включена возможность администрирования процессов разметки, а также сбор размеченных медицинских исследований всеми пользователями в единую базу данных

2

Open source решение (3D slicer)



# Хранение и регистрация наборов данных

РОССИЙСКАЯ ФЕДЕРАЦИЯ

**RU2023620847**



ФЕДЕРАЛЬНАЯ СЛУЖБА  
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ  
ГОСУДАРСТВЕННАЯ РЕГИСТРАЦИЯ БАЗЫ ДАННЫХ, ОХРАНЯЕМОЙ  
АВТОРСКИМИ ПРАВАМИ

Номер регистрации (свидетельства):  
2023620847  
Дата регистрации: 10.03.2023  
Номер и дата поступления заявки:  
2023620537 01.03.2023  
Дата публикации и номер бюллетеня:  
10.03.2023 Бюл. № 3

Автор(ы):  
Чупков Светлана Васильевна (RU),  
Тупицын Николай Николаевич (RU),  
Шолохова Елена Николаевна (RU)

Правообладатель(и):  
Федеральное государственное бюджетное учреждение «Национальный медицинский исследовательский центр онкологии имени Н.Н. Блохина» Министерства здравоохранения Российской Федерации (ФГБУ «НМИЦ онкологии им. Н.Н. Блохина» Минздрава России) (RU)

Название базы данных:  
База данных клинических, цитоморфологических, молекулярно-биологических параметров при раке молочной железы

- НМИЦ им. Н. Н. Блохина использует для сбора, анализа и хранения наборов данных собственные сервера;
- Для осуществления контроля, и прослеживания модификаций размеченных наборов данных, в специальных реестрах, фиксируется объем и контрольные суммы архивов;
- Каждый размеченный набор данных проходит процедуру патентования;
- Производится автоматический «бэкап» наборов данных.

РОССИЙСКАЯ ФЕДЕРАЦИЯ

**RU2023620846**



ФЕДЕРАЛЬНАЯ СЛУЖБА  
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ  
ГОСУДАРСТВЕННАЯ РЕГИСТРАЦИЯ БАЗЫ ДАННЫХ, ОХРАНЯЕМОЙ  
АВТОРСКИМИ ПРАВАМИ

Номер регистрации (свидетельства):  
2023620846  
Дата регистрации: 10.03.2023  
Номер и дата поступления заявки:  
2023620538 01.03.2023  
Дата публикации и номер бюллетеня:  
10.03.2023 Бюл. № 3

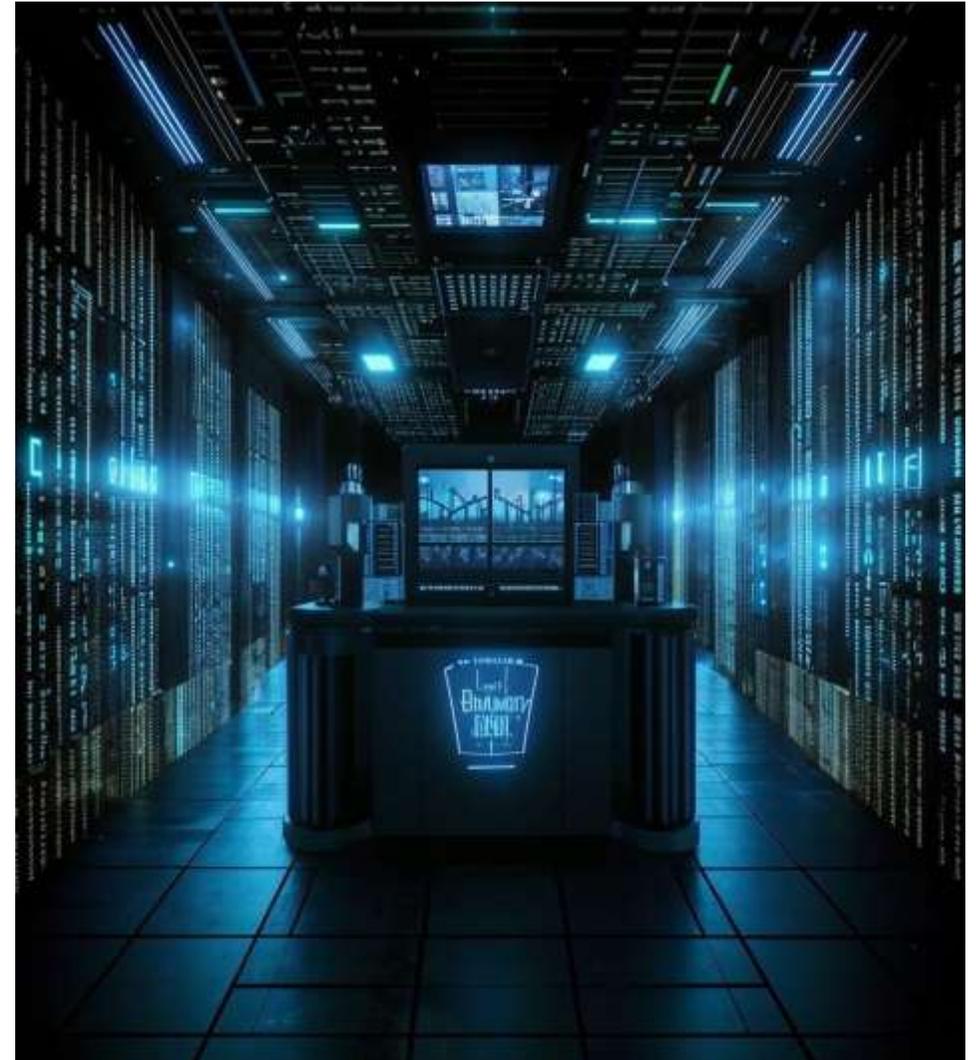
Автор(ы):  
Чупков Светлана Васильевна (RU),  
Тупицын Николай Николаевич (RU),  
Шолохова Елена Николаевна (RU)

Правообладатель(и):  
Федеральное государственное бюджетное учреждение «Национальный медицинский исследовательский центр онкологии имени Н.Н. Блохина» Министерства здравоохранения Российской Федерации (ФГБУ «НМИЦ онкологии им. Н.Н. Блохина» Минздрава России) (RU)

Название базы данных:  
База данных клинических, цитоморфологических, молекулярно-биологических параметров при раке желудка

# Способы защиты и маркировки наборов данных

- Патентование наборов данных;
- Редактирование метаданных DICOM;
- «Открытый» водяной знак;
- «Скрытый» водяной знак;
- Шифрование архивов;
- Соглашение NDA;
- Контроль и администрирование доступа.



## «Открытый» водяной знак

- На медицинское исследование наносится дата и наименование организации, в которой оно было сделано
- Данная операция выполняется в автоматическом режиме медицинским аппаратом;
- Файл DICOM при этом не считается измененным.



## Редактирование метаданных

- При создании файла DICOM, информация, в шестнадцатеричной системе вносится автоматически
- Для защиты данных возможно изменение «профильной строки данных», так и изменение иных строк для улучшения защиты.

- ... (0008,0060) CS Modality = <CT>
- ... (0008,0070) LO Manufacturer = <GE MEDICAL SYSTEMS>
- ... (0008,0080) LO Institution Name = <ПРИМЕР Больницы >
- ... (0008,0090) PN Referring Physician's Name = <>
- ... (0008,009C) PN Consulting Physician's Name = <>
- ... (0008,0201) SH Timezone Offset From UTC = <+0300>
- ... (0008,1010) SH Station Name = <CT01>

# СИСТЕМА МАРКИРОВКИ НЕРАЗМЕЧЕННЫХ МЕДИЦИНСКИХ ИССЛЕДОВАНИЙ



Формирование уникального идентификатора исследования.  
Посредством сбора и компиляции параметров отдельных пикселей на выбранных срезах

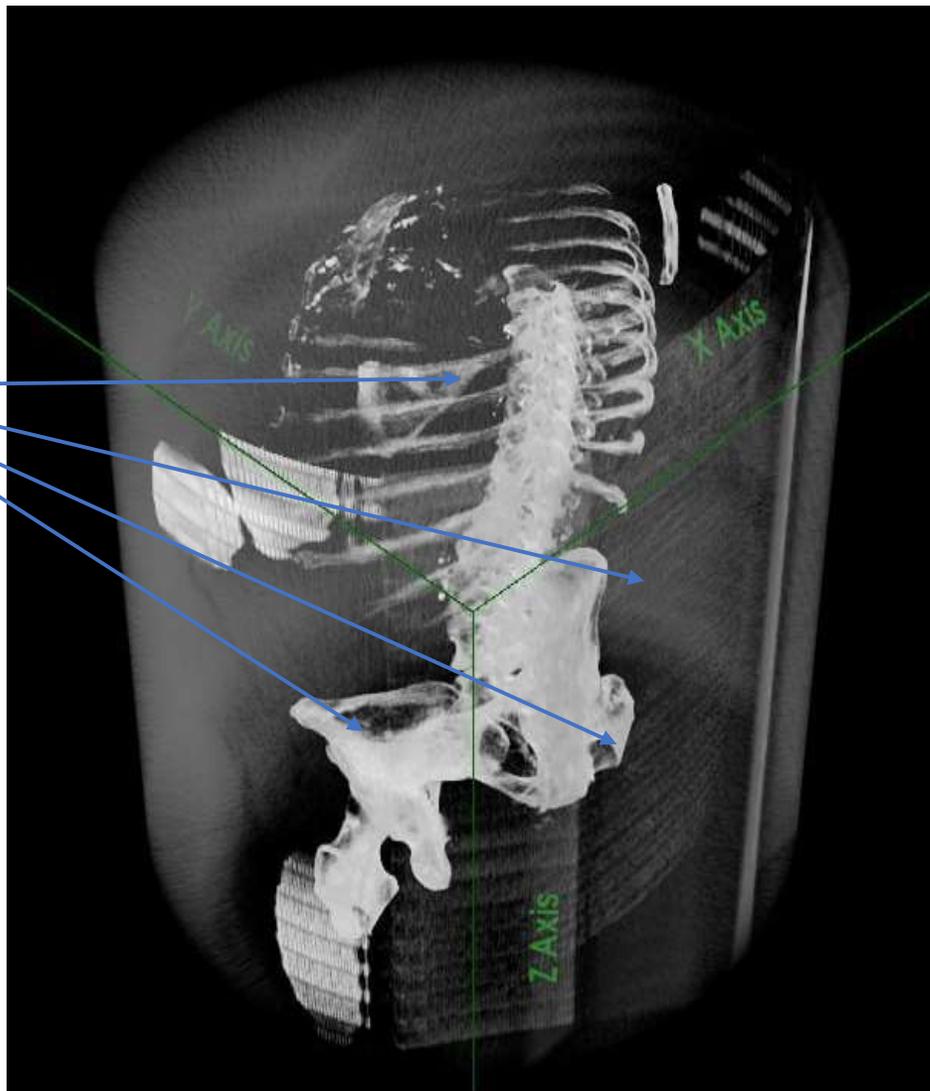


Таблица соответствий			
ФИО пациента	Дата	ID исследования	Идентификатор

# СИСТЕМА МАРКИРОВКИ РАЗМЕЧЕННОГО НАБОРА ДАННЫХ

Скрытое нанесение на слои разметки идентификационных данных лица взявшего набор данных на ответственное хранение, либо использование

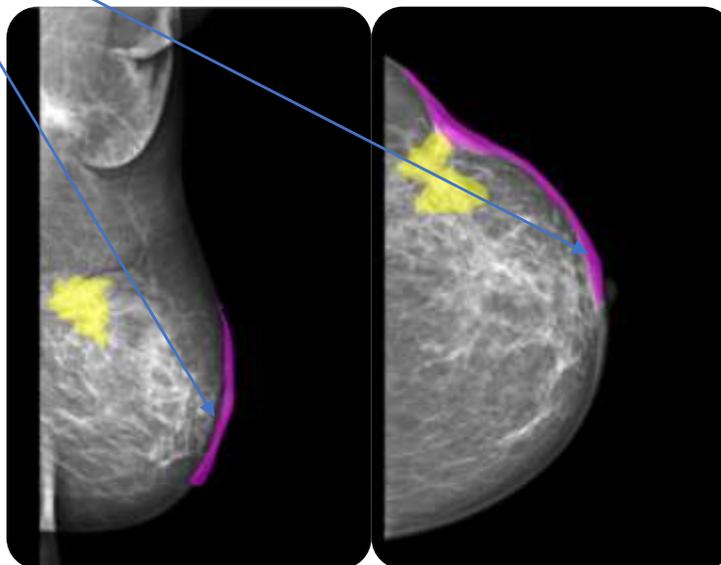
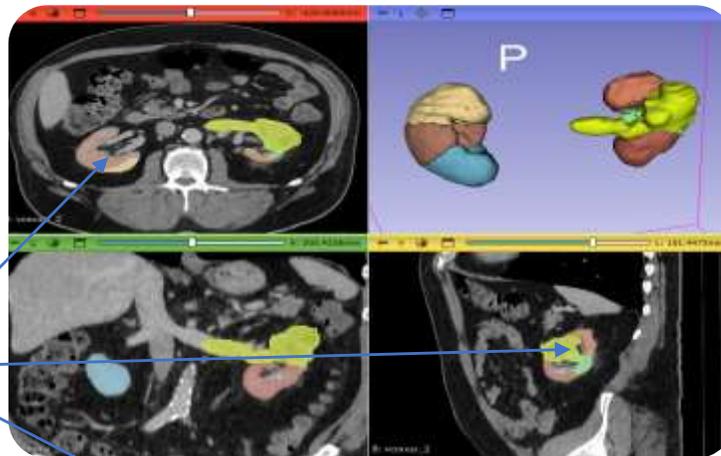


Таблица соответствий			
ID пациента	Дата	ID исследования	Идентификатор

Реестр		
ID пациента	Дата	Координаты метки



**СПАСИБО ЗА ВНИМАНИЕ !**

---