

# Платформа для интеллектуального анализа биомедицинских данных

---

Андрей Бурсов

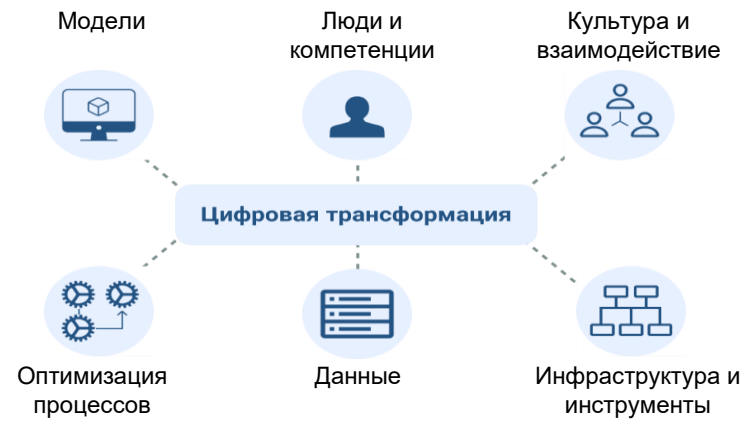
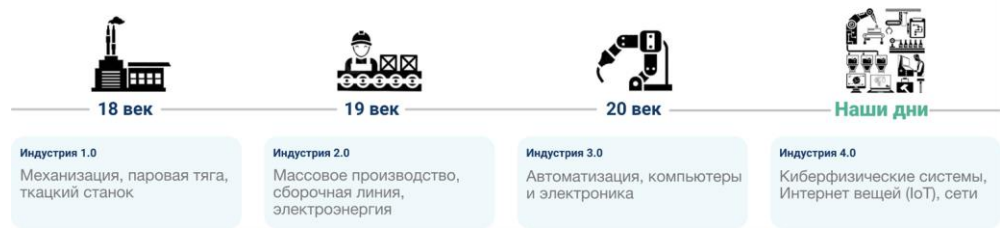
Советник по цифровой медицине



# Цифровая трансформация



**К 2030 году**  
экономика, социальная сфера,  
государственное управление должны  
достичь «цифровой зрелости».



## Результаты цифровой трансформации



**Снижение издержек управления**



**Новые модели деятельности**



**Повышение удовлетворенности при взаимодействии с государством**

# Системное ПО как основа цифровой трансформации

## Прикладные области применения



### Проблемы системного ПО

- Эскалация размеров (Astra Linux – более 150 миллионов строк кода)
- Сложность среды разработки и сборки
- Отсутствие изолированных систем

Платформы «интернета вещей», искусственного интеллекта, ...	
	Платформы хранения и обработки «больших» данных 
	Облачные платформы 
	Аппаратура 

### Необходимые качества системного ПО

- Эффективность
- Продуктивность
- Безопасность

**Одно из направлений – проактивный подход с фокусом на здоровье и долголетию, а не на болезнях**

## **КОНЦЕПЦИЯ 4П-МЕДИЦИНЫ\***

- **predictive** (предиктивная). Выяснение генетической предрасположенности к заболеваниям.
- **preventive** (превентивная). Предотвращение возникновения симптомов заболеваний.
- **personalized** (персонализированная). Учёт индивидуальных особенностей пациента.
- **participative** (партнёрская). Активное участие пациента в профилактике и лечении.

## **СИСТЕМНОЕ ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ – ОСНОВА ЦИФРОВОЙ ТРАНСФОРМАЦИИ**

- Облачные технологии
- Искусственный интеллект (анализ больших данных, машинное обучение)
- Сети передачи данных

\*Dr. Leroy Hood, Systems Biology and P4 Medicine: Past, Present, and Future, 2013.

# Инфраструктура: нужна платформа для анализа

## Источники данных

Носимые девайсы	Показатели здоровья
Фитнес-трекеры	Температура тела
Умные часы	ЭКГ
Кардиомониторы	Давление
Тонометры	Уровень глюкозы
Глюкометры	УЗИ

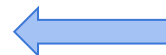
**Интернет вещей, высоко-скоростная сеть**



## Фундамент системы

Дата-центр с платформой для интеллектуального анализа
Хранение и обработка разнородных медицинских данных (в том числе большого объема)
Анализ социальных сетей
ИИ (машинное обучение)
Безопасность
Хранение персональных данных
Разграничение доступа

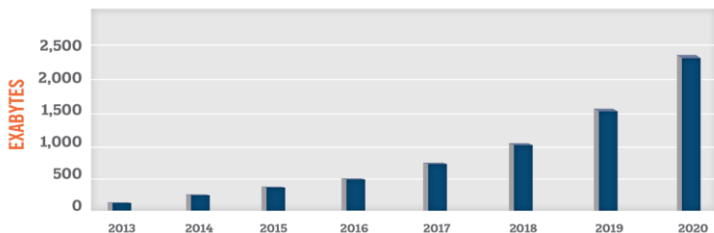
**Проверка гипотез, создание и применение моделей ИИ**



## Пользователи

Пациенты, врачи, ученые-разработчики
Разметка медицинских данных
Персональная дозировка лекарств
Решение об очном визите пациента к врачу
Рекомендация снизить или увеличить физическую нагрузку
Рекомендация срочно вызвать скорую
Срочная корректировка лечения
Регулярное удаленное получение данных для аналитики
Оповещение родственников об опасном состоянии больного

GROWTH IN HEALTHCARE DATA



<https://healthcareappraisers.com/bytes-to-bucks-the-valuation-of-data/>

...

## **АКТУАЛЬНЫЕ ВЫЗОВЫ:**

- безопасное хранение данных
- долгосрочное развитие и жизнедеятельность моделей
- совместная работа распределенных команд
- доверие к ИИ

**Жизненный цикл модели**



**Получение данных из различных источников**



**Хранение больших биомедицинских данных**



**Преобразование данных**



**Аннотирование биомедицинских данных**



**Обучение моделей**



**Тестирование и развёртывание моделей**



**Публикация моделей и сервисов**

**Функции**

Интеграция с внешними системами (МИС, ЛИС и т.п.)  
Интеграция с носимыми устройствами  
Сбор данных из открытых источников

Отказоустойчивость  
Масштабирование  
Версионирование  
Сертификация  
Реестр моделей  
Реестр наборов данных

Конвертация форматов данных  
Домен-специфичная предобработка (приведение частоты дискретизации ЭКГ, предобработка данных МРТ и т.д.)  
Выбор подмножества данных (пол, возраст, фильтрация шумных данных)

Домен-специфичные инструменты аннотирования биомедицинских данных  
Формирование правил аннотирования  
Автоматизация аннотирования, распределение задач и тестирование экспертов  
Создание верифицированных наборов данных  
Интеграция вычисления согласия экспертов в процесс аннотирования

Управление данными и конфигурацией эксперимента  
Эксплоративный анализ и подготовка данных (построение признакового пространства аугментации и т.д.)  
Версионирование моделей  
Low-code средства разработки моделей и AutoML  
Визуализация метрик обучения

Анализ результатов обучения моделей  
Уточнение требований к задачам на аннотирование  
Валидация моделей на доступных наборах данных, полученных с других медицинских устройств, от различных организаций и регионов

Развёртывание и балансировка нагрузки  
Витрина скрининговых сервисов и СППВР  
Получение обратных связей от пользователей моделей и сервисов  
Виртуальные рабочие столы для десктопных сервисов

**Базовая сервис-ориентированная платформа**

API для загрузки данных

API для работы с системами хранения

Управление пайплайнами предобработки данных

API для интеграции инструментов аннотирования и вычисления согласия экспертов

Инфраструктура и сервисы машинного обучения на графических ускорителях (видеокартах), сервисы доверенного ИИ

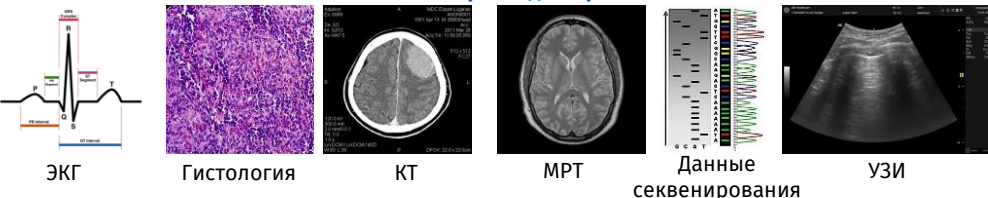
Сервисы MLOps

Сервисы мониторинга

# Создание платформы в рамках НЦМУ совместно с Сеченовским университетом

## Базовая платформа ИСП РАН

### Прикладные решения



SCINOON

### Прикладные инструменты



TALISMAN  
TEXTERRA  
DEDOC

### Технологии анализа данных

Сбор открытых данных  
Анализ медиа  
Анализ текстов  
Анализ изображений  
Анализ временных рядов  
Машинное обучение



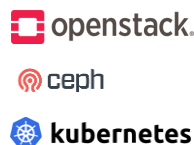
Интеграция



ASPERITAS  
MICHMAN  
FANLIGHT

Облачная инфраструктура, системы оркестрации

Система хранения больших данных



### Аппаратное обеспечение



Медицинские информационные системы



Открытые данные



Государственные порталы







## 1. Организация хранения медицинских данных

- Учет требований 152-ФЗ (ПП 1119, 21 приказ ФСТЭК, 378 приказ ФСБ)
- Обеспечение возможности использования данных для решения задач машинного обучения в рамках платформы
- Обеспечение возможности отзыва права использования ранее переданных данных

## 2. Организация процесса получения наборов данных из сырых медицинских данных

- Методическая и программная поддержка от постановки задачи до получения готового набора данных, пригодного для обучения нейросетевых моделей
- Поддержка интеграции нескольких наборов данных между собой с сохранением приватности

## 3. Организация процесса построения объяснимых и доверенных моделей машинного обучения на основе полученных ранее наборов данных

- Доступ к наборам данных из привычных сред разработки (R, Python)
- Прозрачный доступ к аппаратным ресурсам платформы
- Доступ к специализированным версиям популярных библиотек, адаптированных в рамках работ по доверенному ИИ

## 4. Развитие более обобщенного процесса проведения исследований над медицинскими данными:

- Поддержку итерационного процесса формулирования и проверки гипотез
- Предоставление возможности анализа релевантной академической литературы
- Инструментарий подготовки научных отчетов/статей по результатам проведенных исследований

## ▪Анализ сигналов (ЭКГ)

Совместно с ООО “ТИС”, университетом Технион и др.

2022: подписано соглашение о сотрудничестве с Минздравом Татарстана

2023: планируется подписание соглашения со Станцией скорой и неотложной медицинской помощи им. А.С. Пучкова Департамента здравоохранения г. Москвы

## ▪Анализ генетических данных

Совместно с РУДН, НМИЦ АГП им.В.И.Кулакова и НИИ Морфологии человека и др.

## ▪Анализ изображений

Совместно с МФТИ, медицинской школой Гарварда, НИИ АГиР им. Д.О.Отта и др.

## Назначение сервиса: подготовка верифицированных наборов данных 12-канальных ЭКГ для обучения нейросетевых моделей

Разработан макет системы разметки 12-канальных ЭКГ и нейросетевая модель классификации 12-канальных ЭКГ, опубликован ряд статей

Система подготовлена для интеграции в облачную платформу ИСП РАН Asperitas для обеспечения возможности прозрачного масштабирования мощностей по хранению и анализу ЭКГ



## Назначение сервиса: анализ ЭКГ и синдромальная классификация

№	Медицинская организация	Преприемное население	Врачи	Ставка	Проведено	Обработано	% обработки	Обработано ИИ	Обработано ФД	Подтверждено ФД	Опровергнуто ФД	Без ИИ	Пациенты без Д.учёта
1	ГАУЗ ГП №8 г. Казань	68 086	4	2,5	7.564	6.164	81	423	204	204	0	331	458
2	ДГП №2 г. ИЧ	20 095	3	1,5	1.963	1.962	100	475	13	13	0	21	9
3	ДГП №4 г. ИЧ	15 000	3	0,5	2.644	2.643	100	544	64	64	0	52	93
4	ДГП №3 г. ИЧ	19 800	3	2	3.246	3.243	100	842	272	272	0	60	284
5	ДГП №2, Казань	17 098	2	1,75	1.702	1.700	100	425	78	77	1	9	121
6	ДЛ ПДБ с ПД г. Альметьевск	47 000	2	0	428	380	89	2	0	0	0	0	0
7	ДГП №4 г. ИЧ	19 834	3	2	3.573	3.567	100	493	54	54	0	0	0
8	КБ №2 г. Казань	66 784	4	3	6.078	1.484	24	558	3	3	0	0	0
9	Вернеевская ЦРБ	16 806	2	2	1.172	34	3	0	0	0	0	0	0
10	ГАУЗ Кумурская ЦРБ	0	2	1	5.234	5.233	100	41	1	1	0	0	0
11	ГБ №11 г. Казань	20 798	4	3	5.321	5.300	100	322	212	212	0	0	0
Итого		326 369	392	154,75	318 275	241 684	76	18 385	5 352	5 352	0	0	0

- Нейросетевая модель классификации 12-канальных ЭКГ обучена на 1,5 млн ЭКГ из разных регионов (Республика Татарстан, Москва, Великий Новгород)
- Подписано соглашение о развитии искусственного интеллекта в системе здравоохранения с Министерством здравоохранения Республики Татарстан
- Модель интегрирована в режиме опытной эксплуатации систему «Единый Кардиолог» и апробирована на данных ЭКГ Республики Татарстан

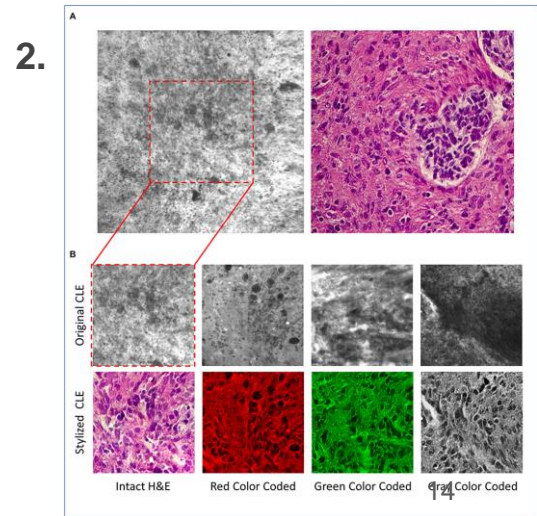
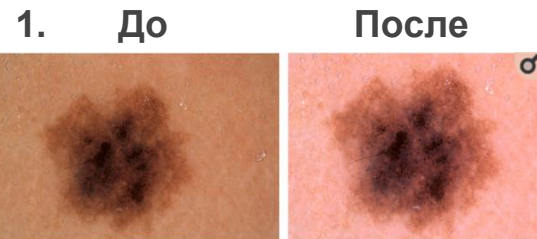
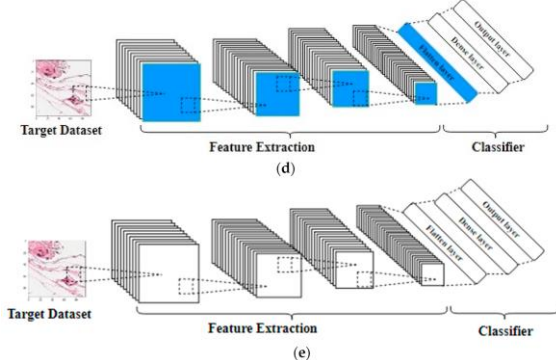
Синдром	Специфичность	Чувствительность	ROC-AUC
Фибрилляция предсердий	0.980	0.988	0.984
Блокада левой ножки пучка Гиса	0.966	0.901	0.932
Блокада передней ветви левой ножки пучка Гиса	0.911	0.960	0.935
Полная блокада правой ножки пучка Гиса	0.982	0.980	0.981
Неполная блокада правой ножки пучка Гиса	0.874	0.943	0.909
Синусовая тахикардия	0.925	0.975	0.950
Синусовая брадикардия	0.921	0.989	0.956
АВ-блокада 1 степени	0.927	0.969	0.947
Желудочковая экстрасистолия	0.984	0.986	0.984
Локализация ИМ в нижней области	0.890	0.949	0.912
Локализация ИМ в передней стенке левого желудочка	0.918	0.964	0.941
Наджелудочковые экстрасистолы	0.947	0.929	0.938

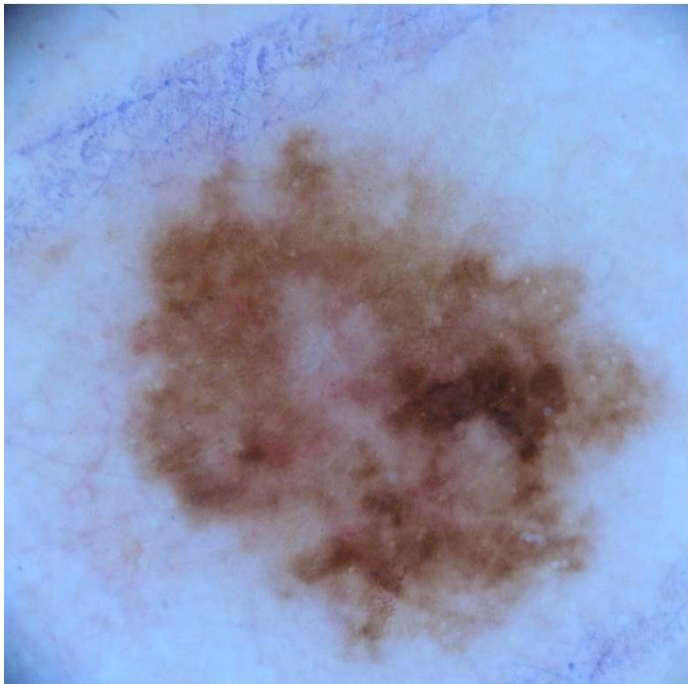
# Задача раннего обнаружения меланомы

**Задача распознавания меланомы на основе анализа меланоцитарных изображений включает следующие подзадачи:**

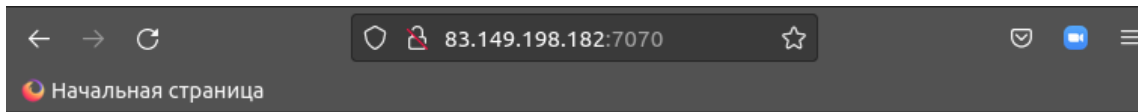
- 1. Предобработка изображения** — улучшение изображения
- 2. Обнаружение объекта** — локализация объекта, осуществляемая посредством сегментации изображений и определения положения интересующего объекта
- 3. Извлечение признаков и обучение** — использование статистических методов или методов глубокого обучения для выявления наиболее значимых шаблонов изображения и уникальных функций, которые помогут классификатору различить разные классы
- 4. Классификация изображения по predetermined классам** с использованием подходящего метода, который сравнивает шаблоны изображений с целевыми шаблонами

3,4.





Диагноз: **меланома**



Идентификатор\_пациента:

Возраст:

Пол:

Вероятность меланомы:

Вероятность невуса:

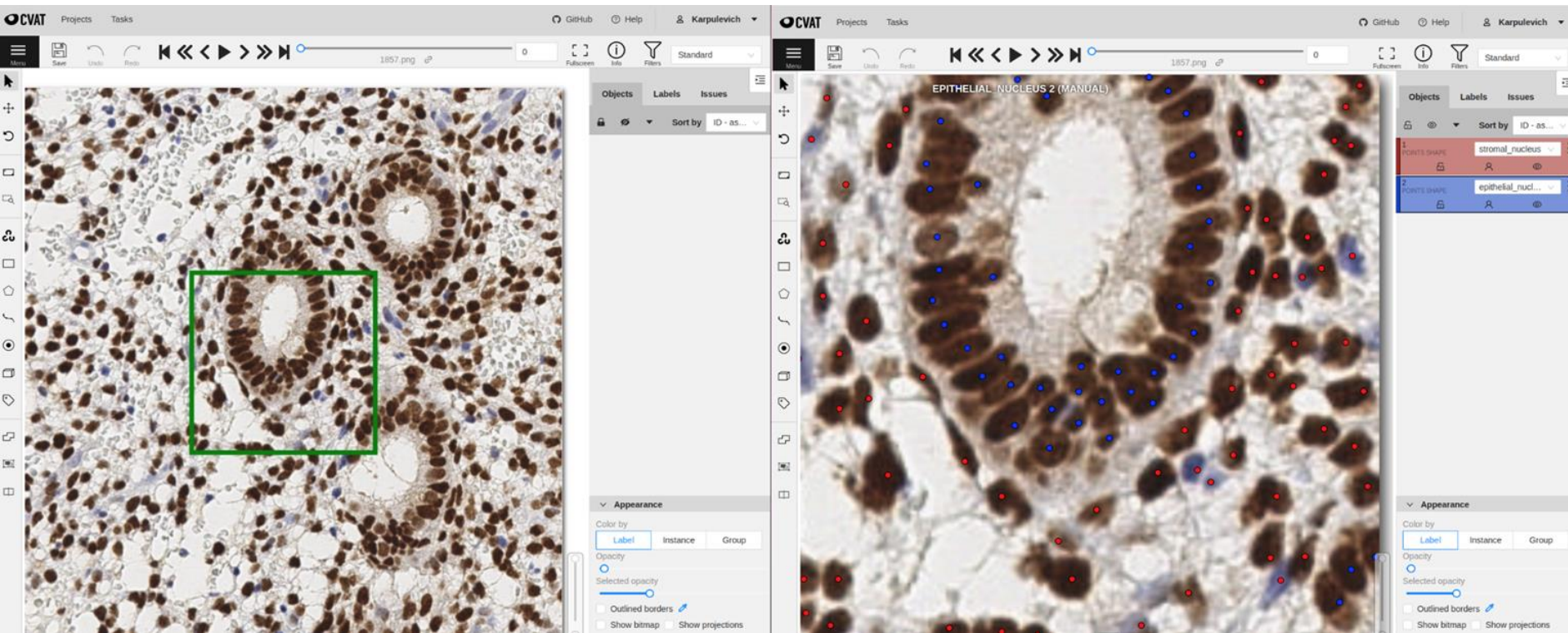
[Вернуться в начало](#)



**Рисунок.** Результаты анализа меланоцитарных изображений разработанным прототипом системы раннего обнаружения меланомы

# Разметка с помощью CVAT (Computer Vision Annotation Tool)

<https://cvat.org/>





## Набор данных

Для обучения использовались сканы тканей от 12 доноров. Каждый слайд делился на более мелкие тайлы. Пустые тайлы убирались из набора. Для разметки были задействованы 7 человек. Во время разметки на каждом изображении размечались ядра клеток стромы и эпителия.

## Количество размеченных ядер:

	Строма	Эпителий
Эстроген	117 435	29 757
Прогестерон	53 561	7 965

## Текущие результаты

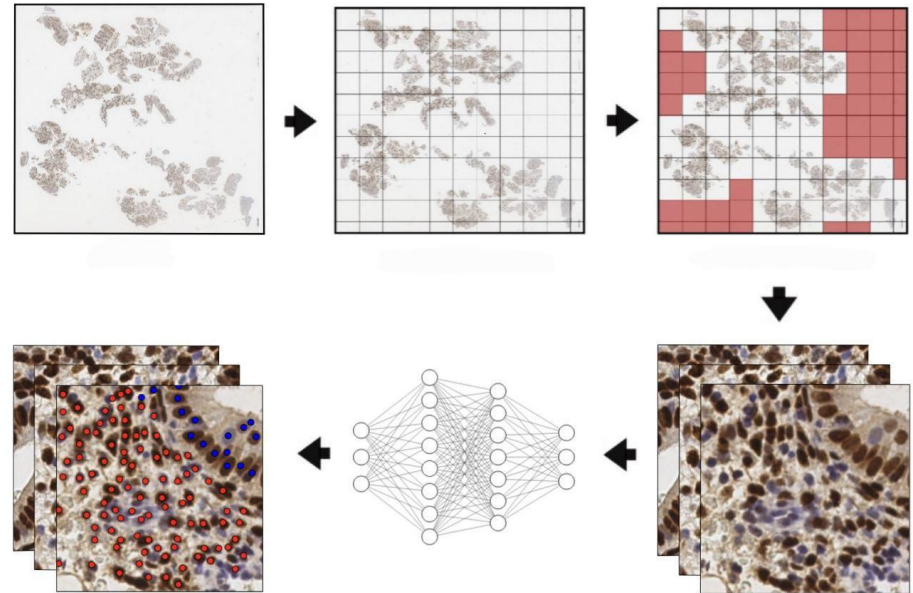
В качестве архитектуры модели была выбрана **CenterNet**

В качестве бэкбона хорошо показала себя сеть **ResNet**

Для оценки качества работы обученной модели использовалась метрика mean Average Precision (mAP).

mAP = 0.77

Справа приведен пример работы обученной модели детекции. На тайлах красные точки - ядра клеток стромы, а синие точки - ядра клеток эпителия.

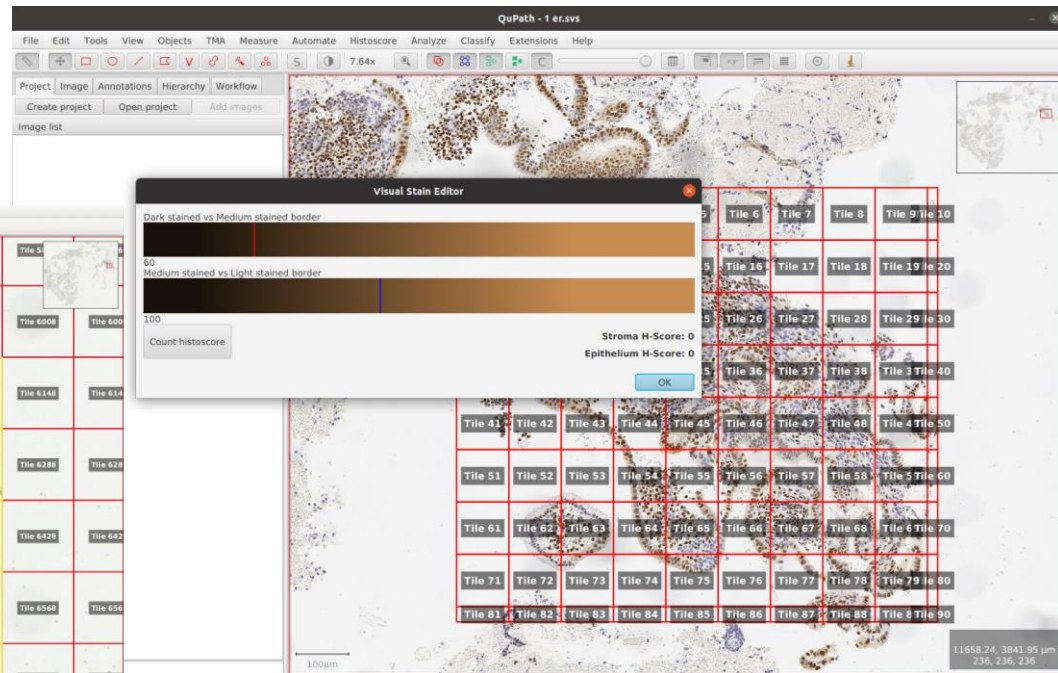
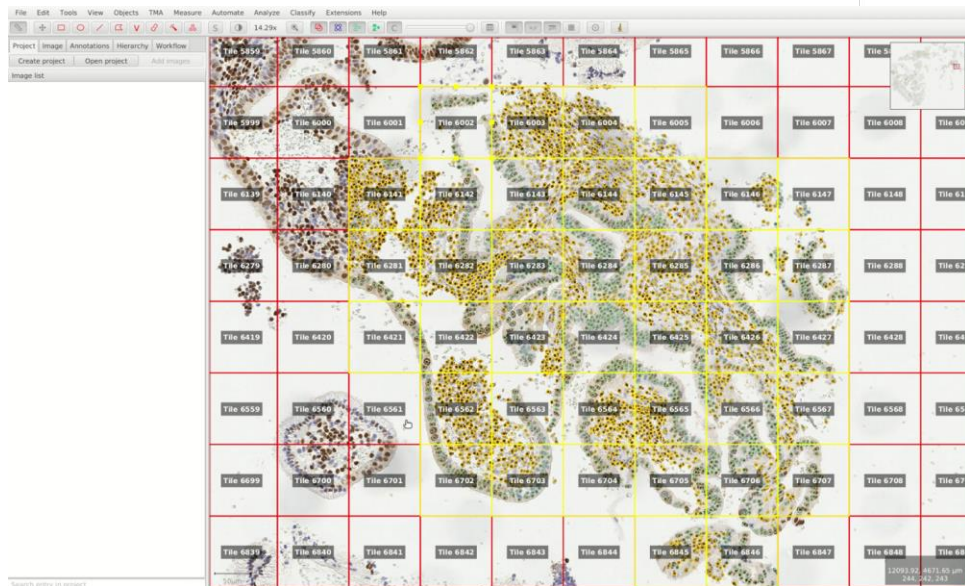


**Опубликована статья про датасет:** Naumov, A.; Ushakov, E.; Ivanov, A.; Midiber, K.; Khovanskaya, T.; Konyukova, A.; Vishnyakova, P.; Nora, S.; Mikhaleva, L.; Fatkhudinov, T.; Karpulevich, E.

**EndoNuke: Nuclei Detection Dataset for Estrogen and Progesterone Stained IHC Endometrium Scans.** *Data* 2022, 7, 75. <https://doi.org/10.3390/data7060075>

## Интеграция модели и подсчета H-score в QuPath

QuPath с интегрированной нейросетевой моделью детекции был интегрирован в Fanlight - SaaS платформу организации единой Web-среды исследований, разработок и образования и развернут в облаке ИСП РАН





Спасибо за внимание!

---

