



Машинное обучение в биологии и медицине

Автор: Белокопытова Полина

черный

рыжий

британский

вислоухий

шотландский

смешной

белый

серый

толстый

чеширский

сибирский

персидский

пушистый

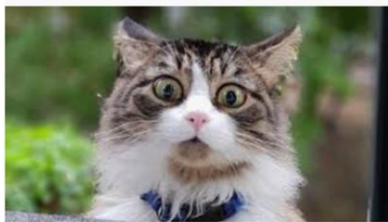
сиамский



Самый старый кот в мире отметил 30-летн...
kp.ru



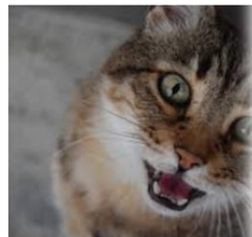
Посмотрите на кота, у которого 1 млн подп...
probusiness.io



Мурлок - кот, который как собака, только сова - ...
youtube.com



В Новокузнецке кот затопил склад с игрушк...
rg.ru



Кот — это вам не кошка! | K1NEV...
k1news.ru



Кот Шрёдингера — знаменитый ...
sciencepop.ru



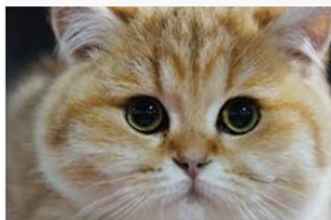
На кого подписаться: Боль...
wonderzine.com



На кого подписаться: Очен...
wonderzine.com



Вежливый кот - кот с глазами...
memepedia.ru

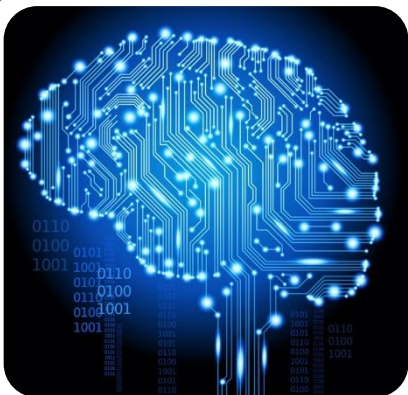


В Карелии кот оставил без электричеств...
ria.ru



Самый большой кот в мире: фото, пород...
24smi.org



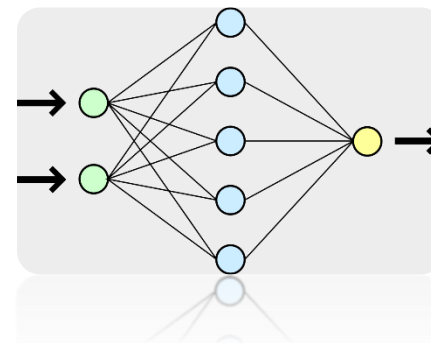


Искусственный
интеллект



Машинное обучение

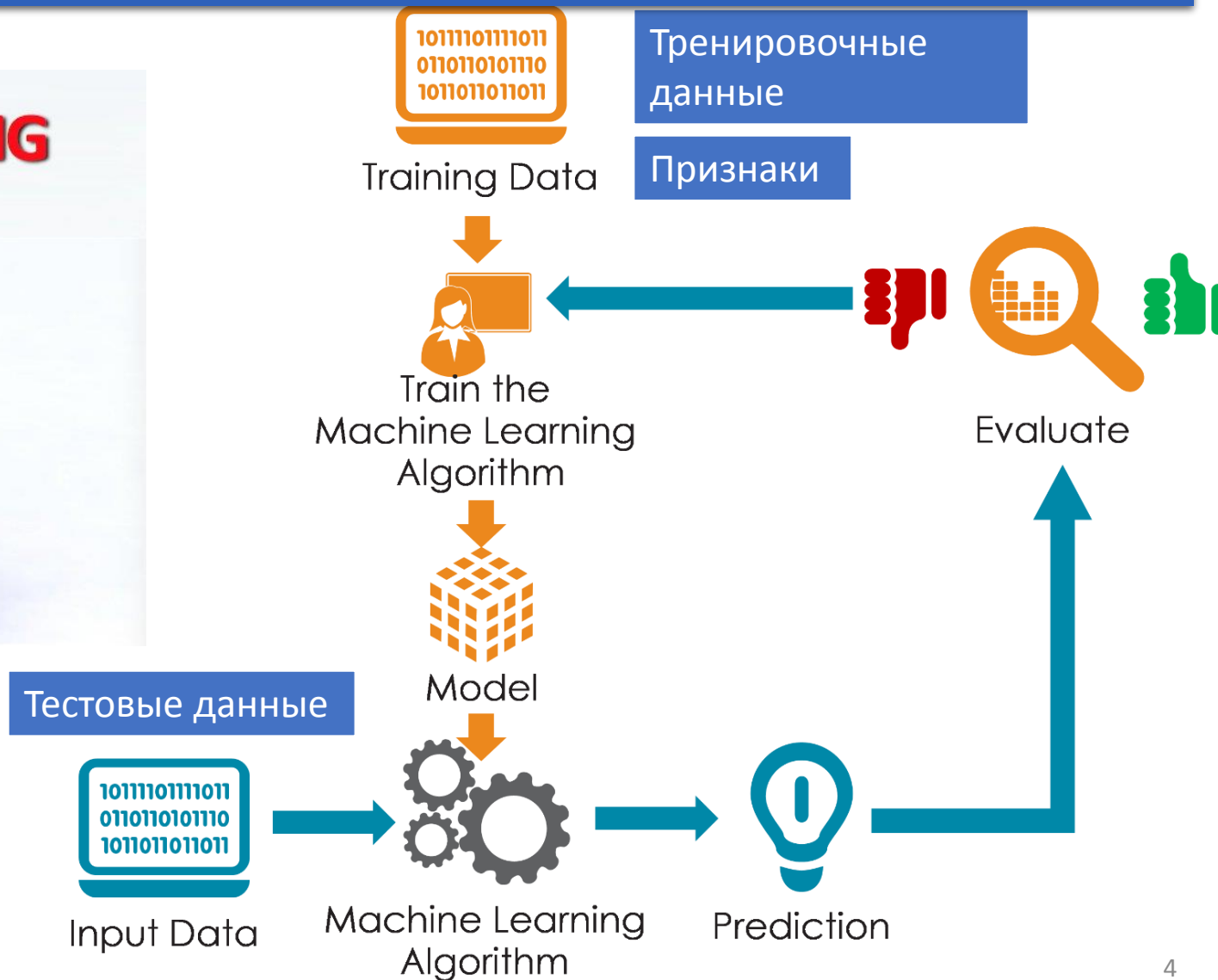
Глубокое обучение



Машинное обучение. Как это работает?

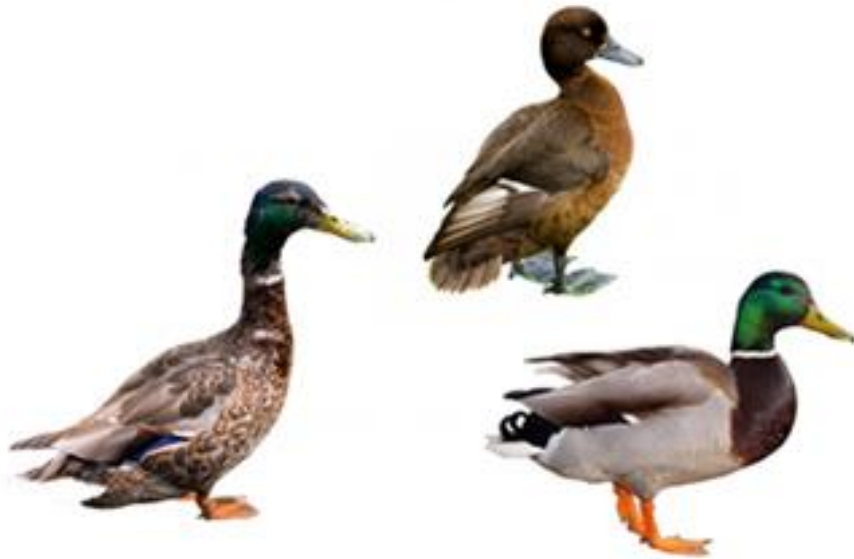


Trash in – trash out



If it Walks/Swims/Quacks like a duck...Then it must be a duck!

✓ Walks ✓ Quacks ✓ Swims



DUCKS



NOT DUCKS

Подходы машинного обучения

Обучение с учителем



- ☒ Классификация
- ☐ Регрессия
- ☐ Ранжирование

Обучение без учителя



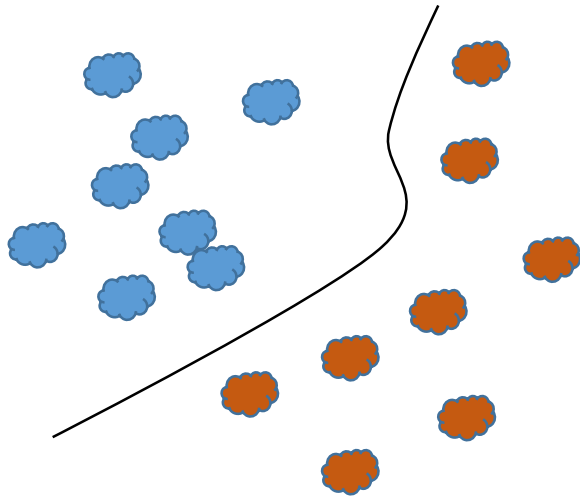
- ☐ Кластеризация
- ☐ Уменьшение размерности

Обучение с подкреплением

- Самоуправляемые автомобили
- Роботы пылесосы

Обучение с частичным привлечением учителя

Классификация



Примеры:

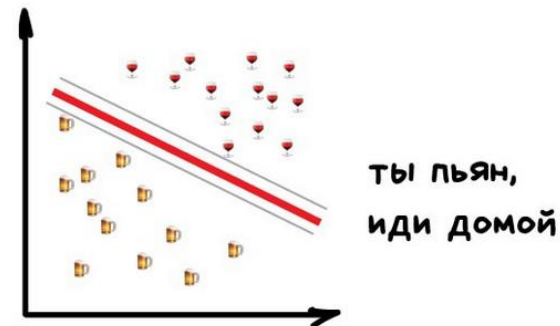
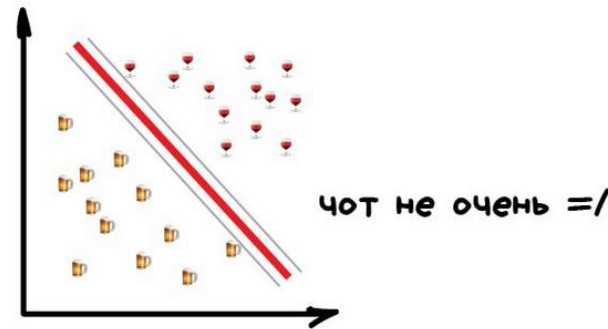
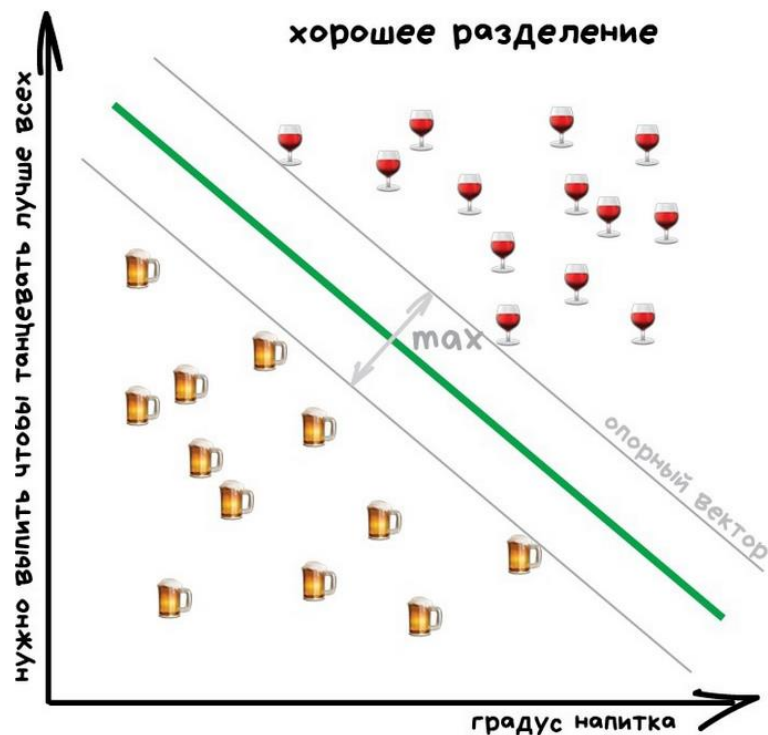
- Спам-фильтр
- Определение объекта на фото
- Определение языка
- Предсказание промотор-энхансерных взаимодействий
- Анализ тональности
- Предсказание эффекта мутаций
- Предсказание состояния хроматина
- Предсказание вторичной структуры белка
- и т.д.

Методы:

- ☐ Логистическая регрессия
- ☐ Машина опорных векторов
(Support vector machine, SVM)
- ☐ Наивный Байес
- ☐ Деревья решений
- ☐ Нейросети
- ☐ Ансамбли (Random Forests,
Gradient Boosted Trees)
- ☐ k-NN

SVM (метод опорных векторов)

Разделяем виды алкоголя

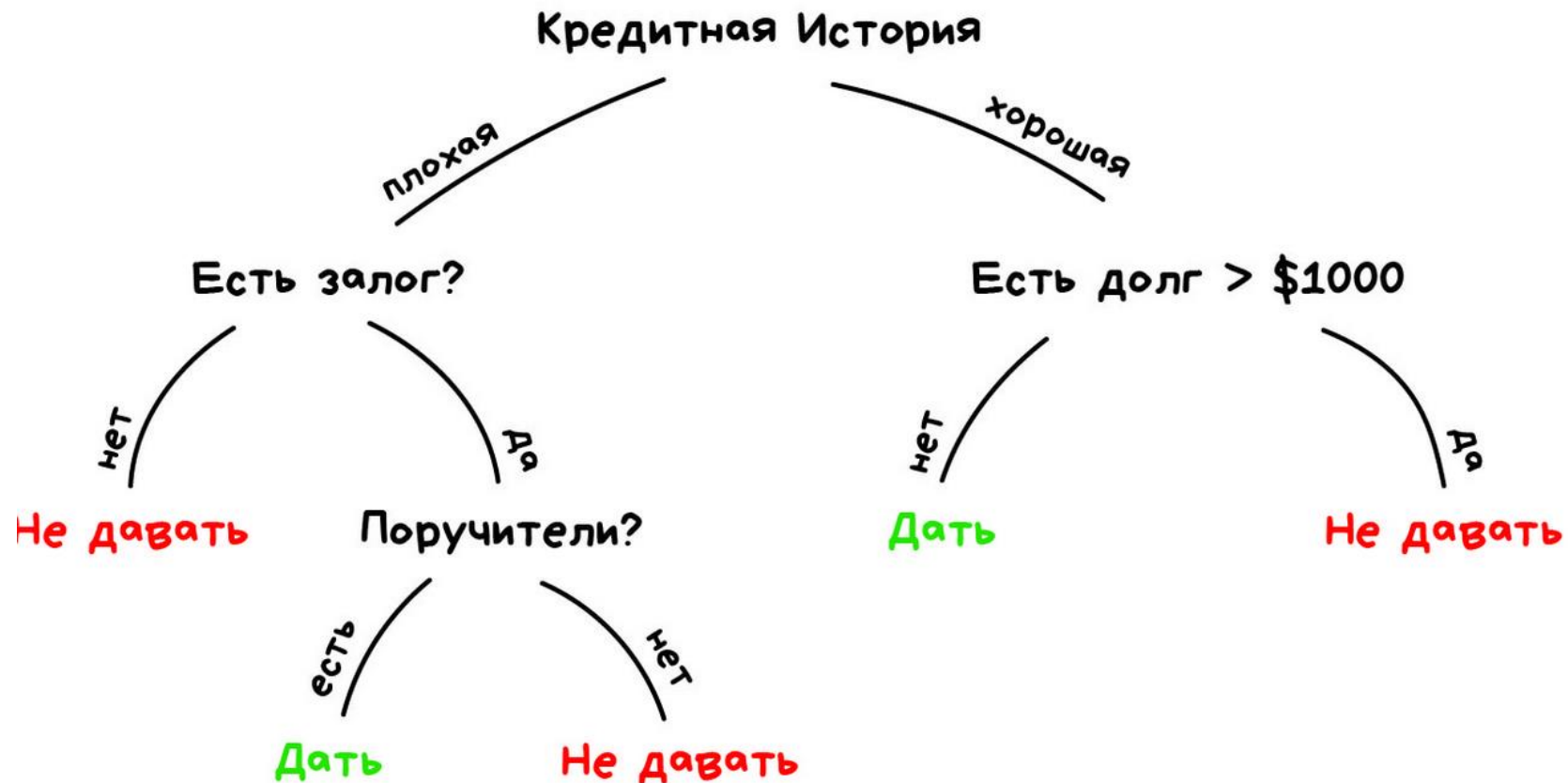


Различных разделяющих плоскостей может быть много, надо выбрать “лучшую”

как так провести две прямые между категориями, чтобы между ними образовался наибольший зазор.

Дерево решений

Давать ли кредит?



Подходы машинного обучения

Обучение с учителем



- ☐ Классификация
- ☒ Регрессия
- ☐ Ранжирование

Обучение без учителя



- ☐ Кластеризация
- ☐ Уменьшение размерности

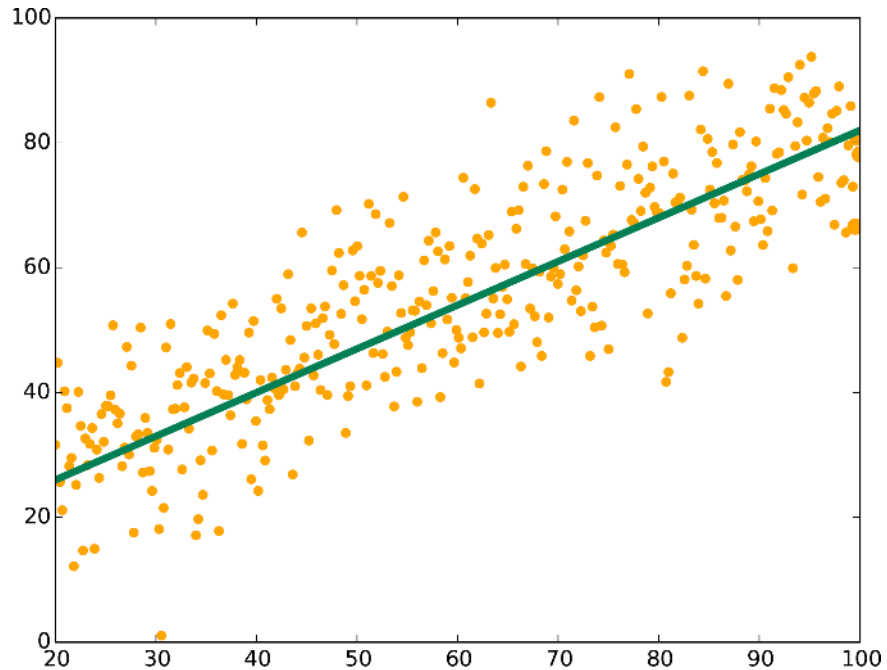
Обучение с подкреплением

- Самоуправляемые автомобили
- Роботы пылесосы

Обучение с частичным привлечением учителя

Регрессия

Геометрически определяет прямую (в случае линейной регрессии), наиболее близко проходящую ко всем точкам.



Примеры:

- Прогноз стоимости ценных бумаг
- Предсказание трехмерной архитектуры хроматина
- Предсказание температуры воздуха
- Предсказание уровня экспрессии гена
- И т.д.

Подходы машинного обучения

Обучение с учителем



- ☐ Классификация
- ☐ Регрессия
- ☐ Ранжирование

Обучение без учителя



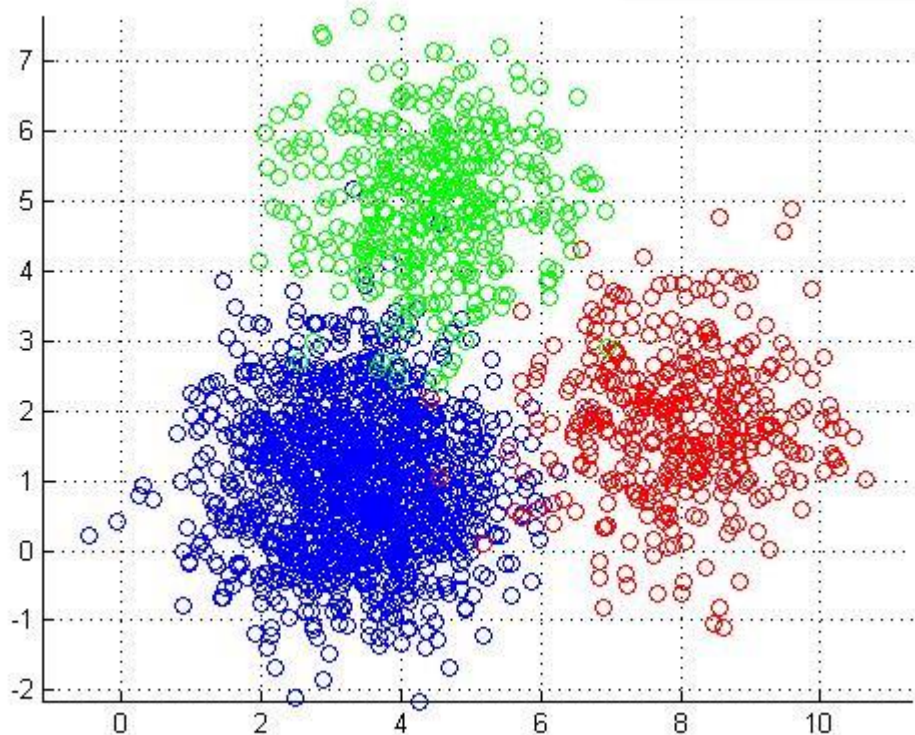
- ☐ Кластеризация
- ☐ Уменьшение размерности

Обучение с подкреплением

- Самоуправляемые автомобили
- Роботы пылесосы

Обучение с частичным привлечением учителя

Кластеризация



Примеры:

- Объединение близких точек на карте
- Анализ и разметки новых данных
- Группировка покупателей по потребительскому поведению (частота покупок, виды товаров в корзине, средний чек,...)
- Группировка клеток по профилю экспрессии
- И т.д.

Подходы машинного обучения

Обучение с учителем



- ☐ Классификация
- ☐ Регрессия
- ☐ Ранжирование

Обучение без учителя



- ☐ Кластеризация
- ☐ Уменьшение размерности

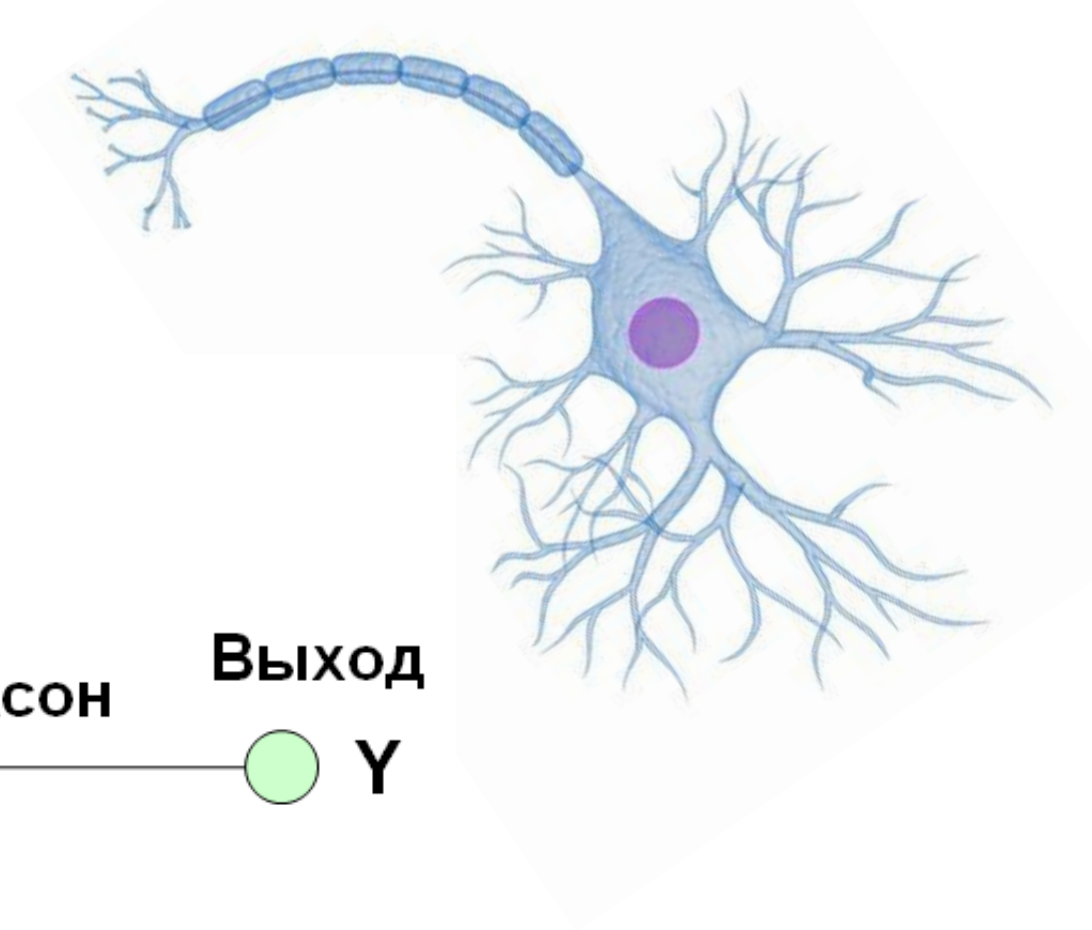
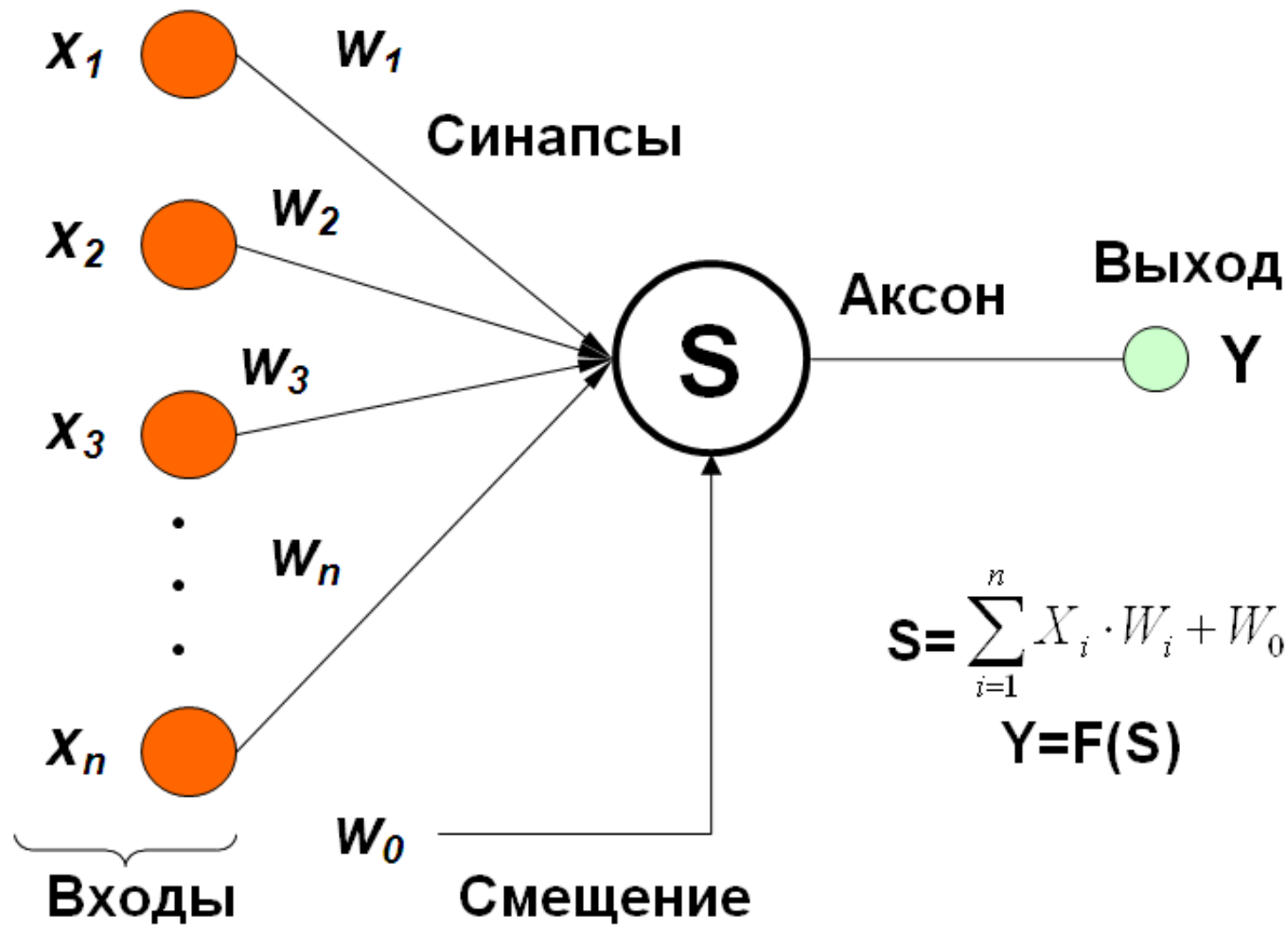
Обучение с частичным привлечением учителя

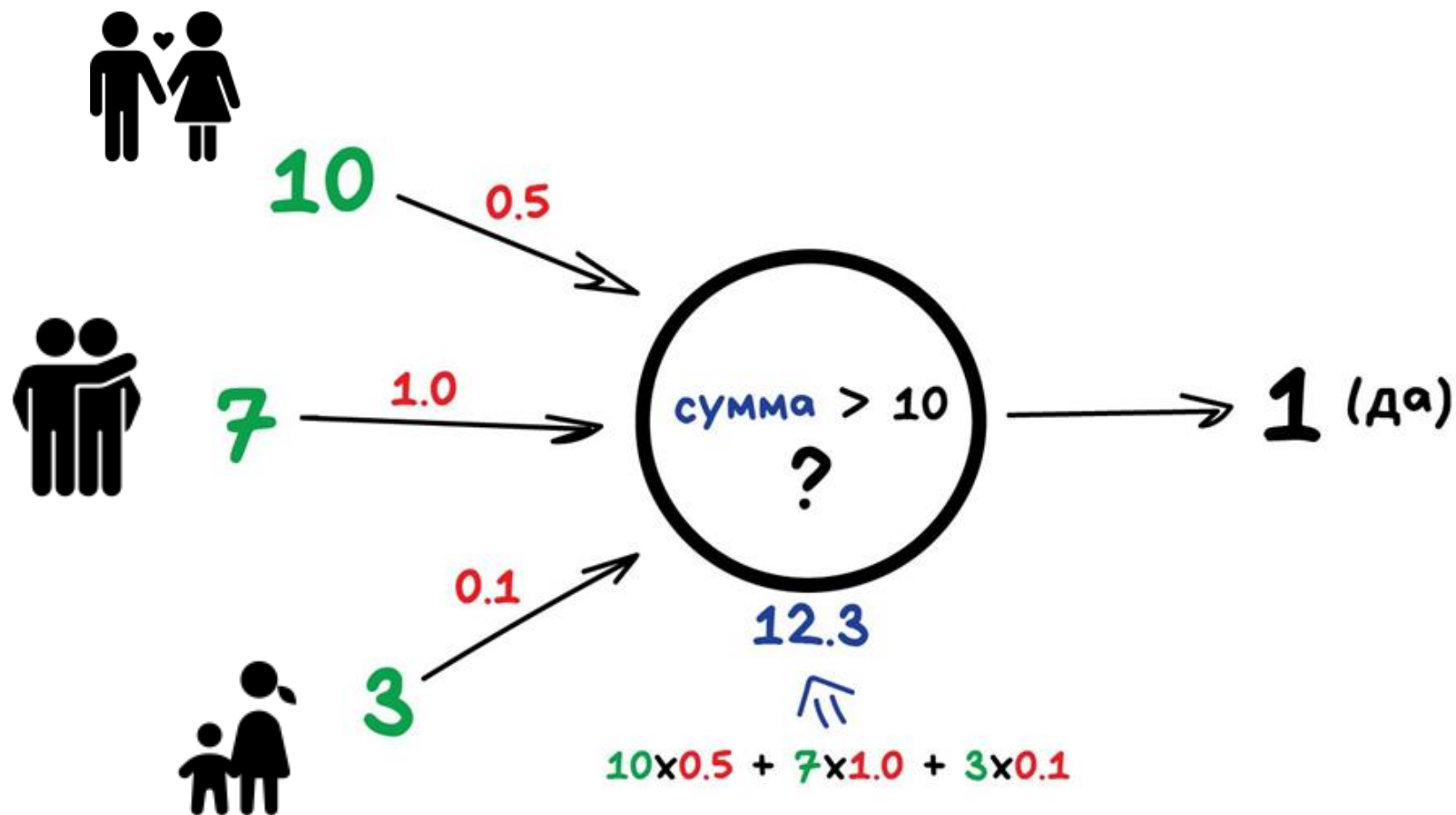
Обучение с подкреплением

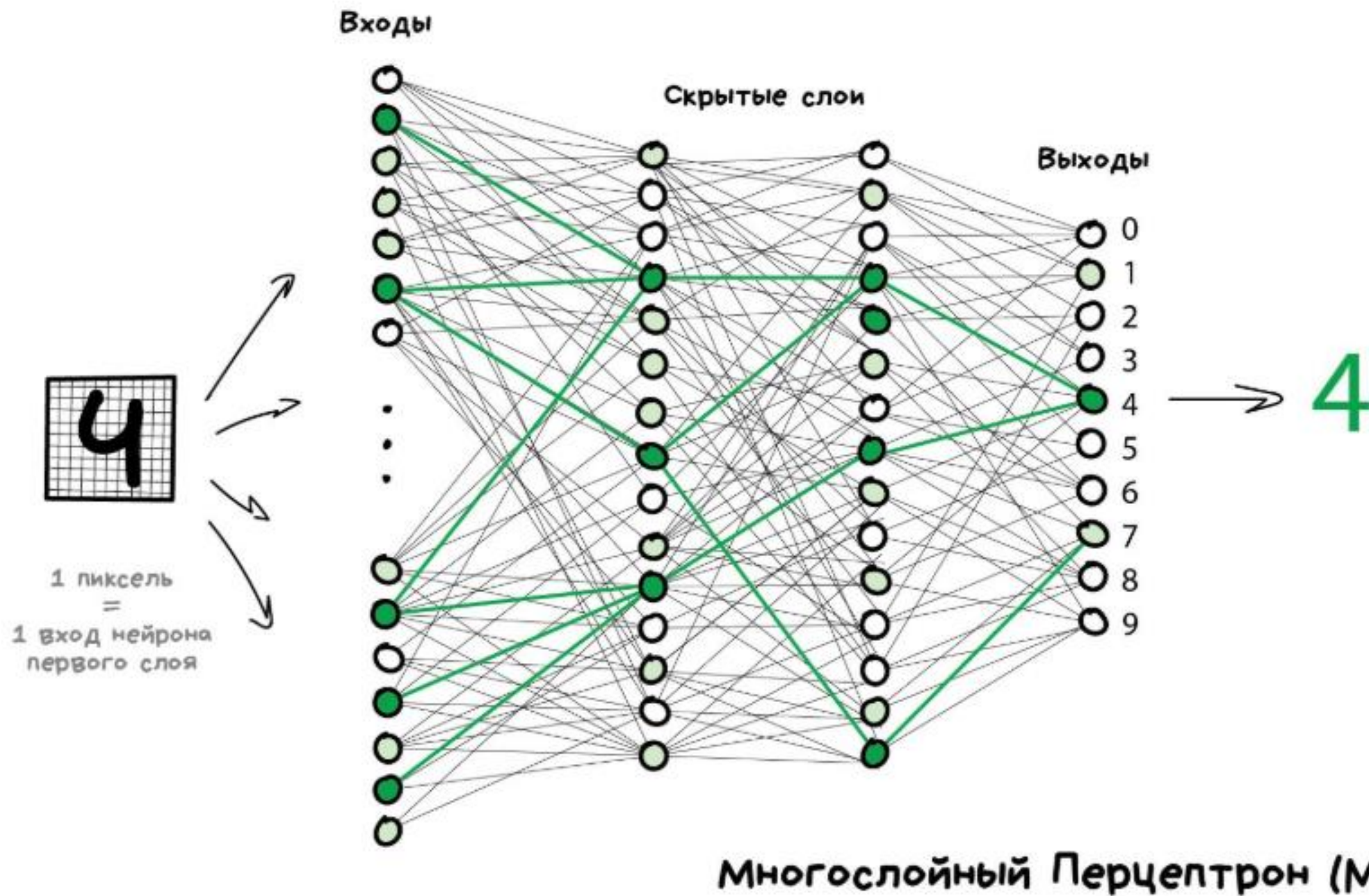
- Самоуправляемые автомобили
- Роботы пылесосы



Нейросети

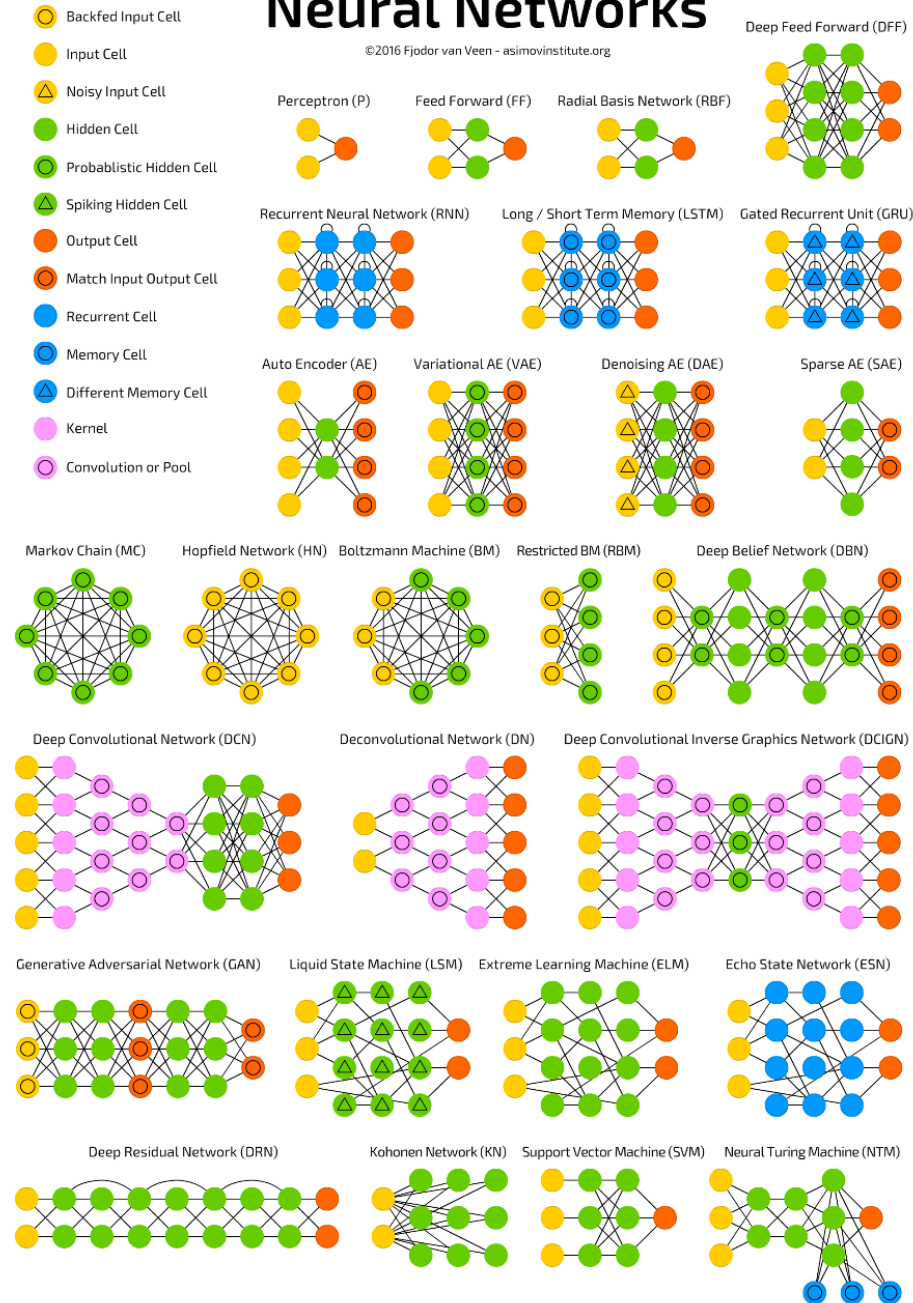




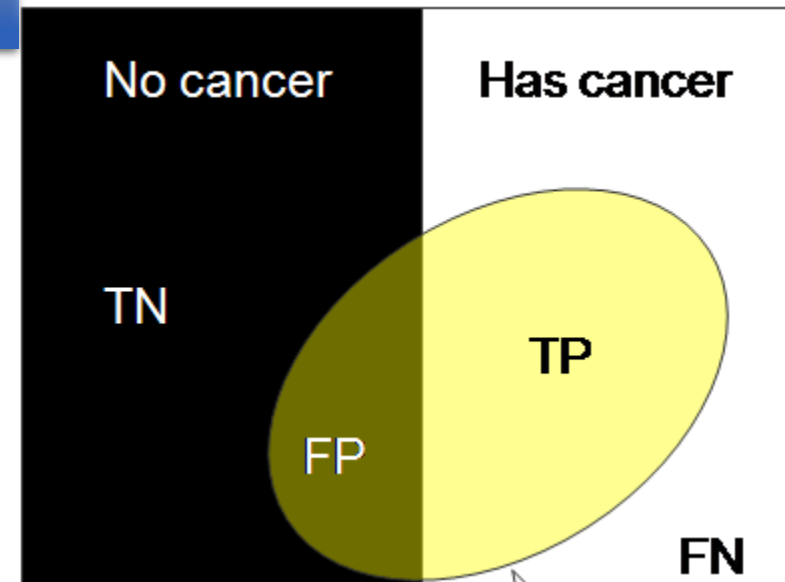
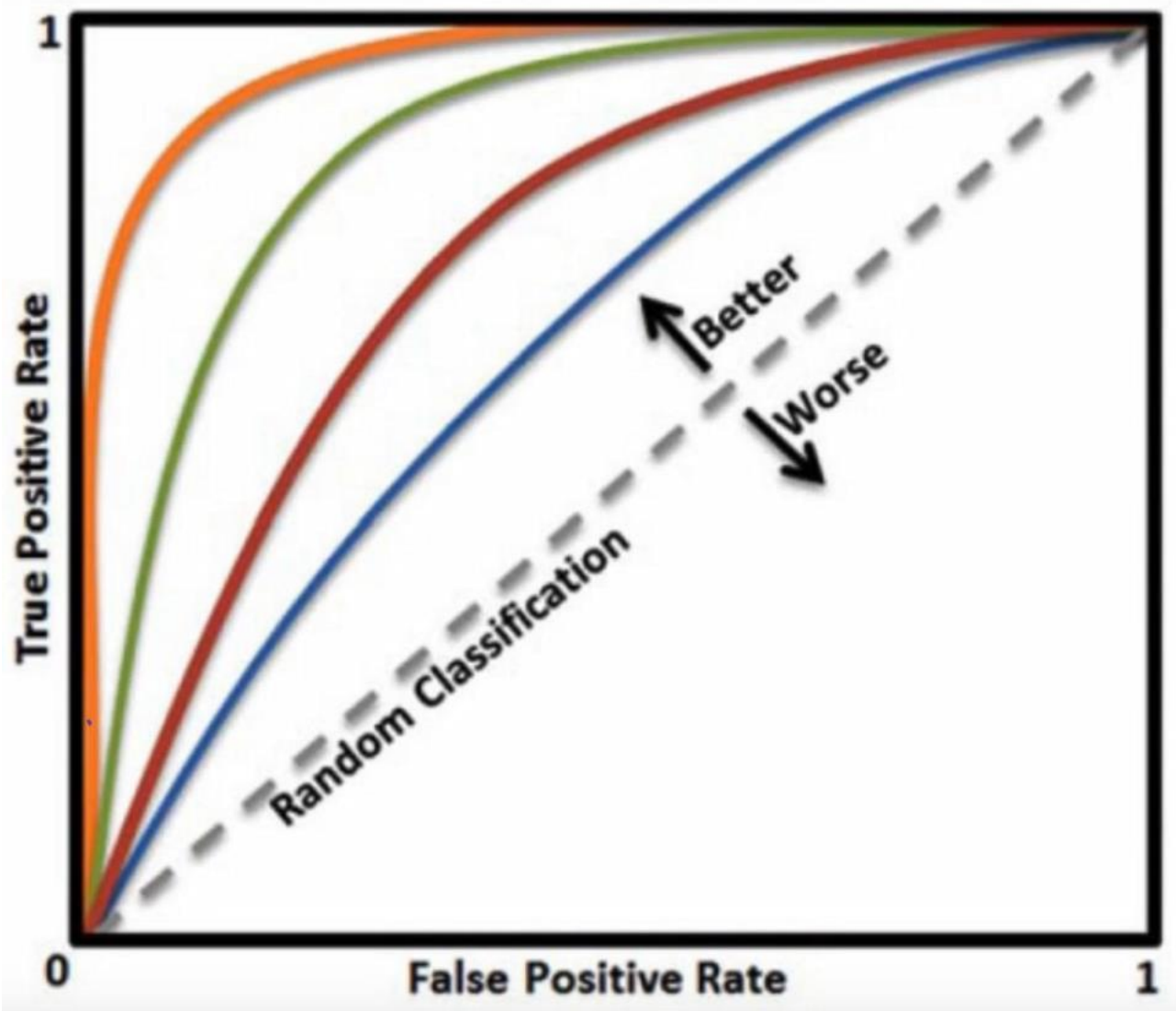


A mostly complete chart of Neural Networks

©2016 Fjodor van Veen - asimovinstitute.org



Оценка точности предсказания



	Has cancer	No cancer
Predict cancer	TP	FP
Predict No cancer	FN	TN

$$FPR = \frac{FP}{TN + FP}$$

$$\text{Чувствительность (TPR)} = \frac{TP}{TP + FN}$$

$$\text{Специфичность} = \frac{TN}{TN + FP}$$

Так что там насчет биологии и медицины?



> 10 млн
изображений

> 21 тыс. категорий



Применение ИИ в медицине



Диагностика
болезней

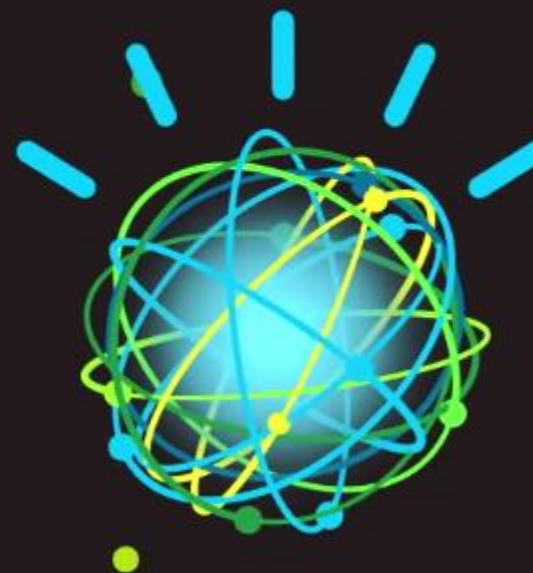
Персонализированная
медицина

Разработка лекарств

Предсказание
вспышек заболеваний

“Интеллектуальные
медицинские записи”

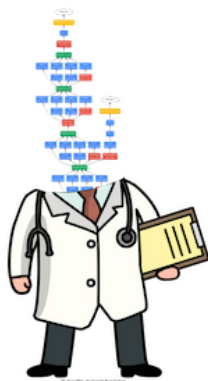
IBM.WATSON



> 600 тыс. медицинских
заключений

> 2млн страниц текста

> 42 медицинских
журналов



Chester the AI Radiology Assistant

NOT FOR MEDICAL USE. This is a web based (but locally run) prototype system for diagnosing chest X-ray images. The patient data remains on your computer and all computation occurs in your browser. The goals of this system are:

1. Let people play with deep learning tools to know how they work and their limitations.
2. Show the potential of open data (needed to build a public system like this).
3. Create a tool to help teach radiology.
4. Demonstrate a model delivery system that can scale to provide free medical tools to the world.

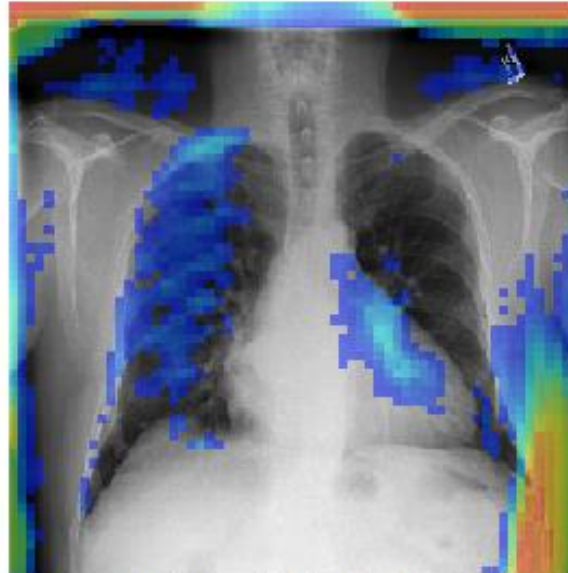
Example Image (00000001_001-Cardiomegaly-Emphysema.png)

Input Image



Out Of Distribution reconstruction error

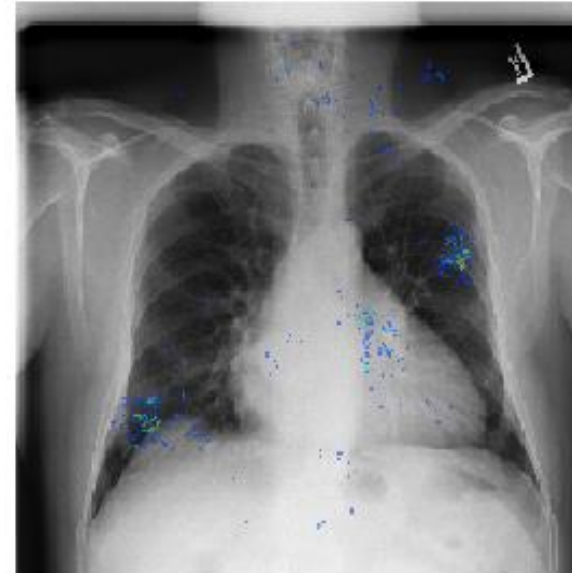
Heatmap where the image varies from the training distribution.



recScore:0.27, ssim:0.39

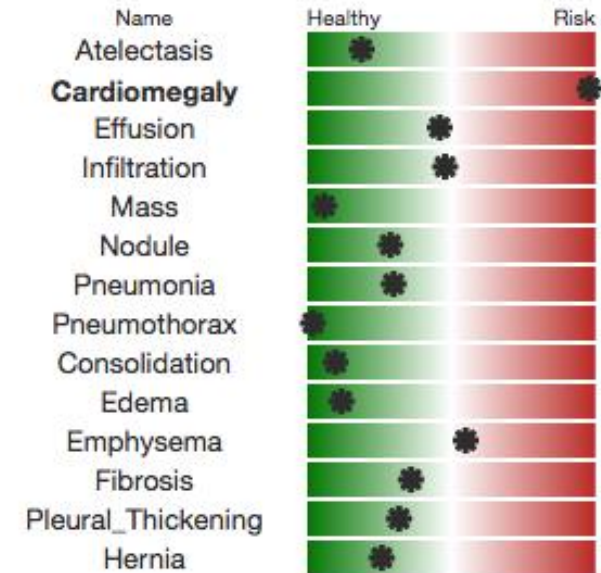
Predictive image regions

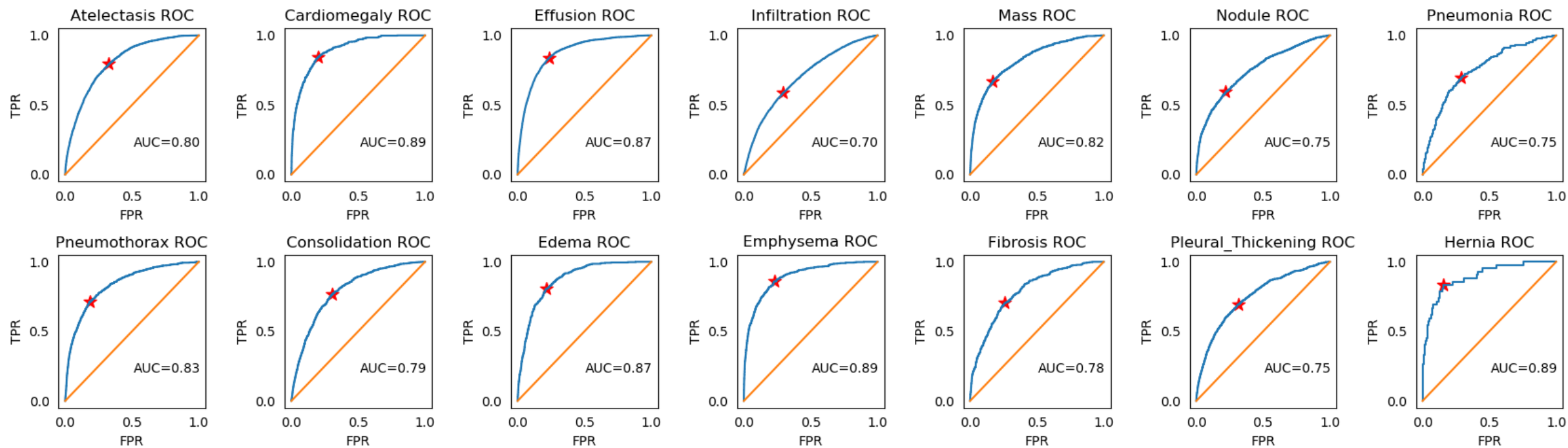
Heatmap of image regions which influence the prediction.



Disease Predictions

Probability of a disease.





Применение ИИ в медицине

A diagram illustrating the application of Artificial Intelligence (AI) in medicine. The background is a close-up photograph of various colorful pills and capsules. A blue banner at the top contains the title. Five blue arrows originate from the banner and point to different text boxes: 'Диагностика болезней' (Disease diagnosis), 'Персонализированная медицина' (Personalized medicine), 'Разработка лекарств' (Drug development), 'Интеллектуальные медицинские записи' (Intelligent medical records), and 'Предсказание вспышек заболеваний' (Prediction of disease outbreaks). The 'Персонализированная медицина' box is highlighted with a red oval.

Диагностика
болезней

Персонализированная
медицина

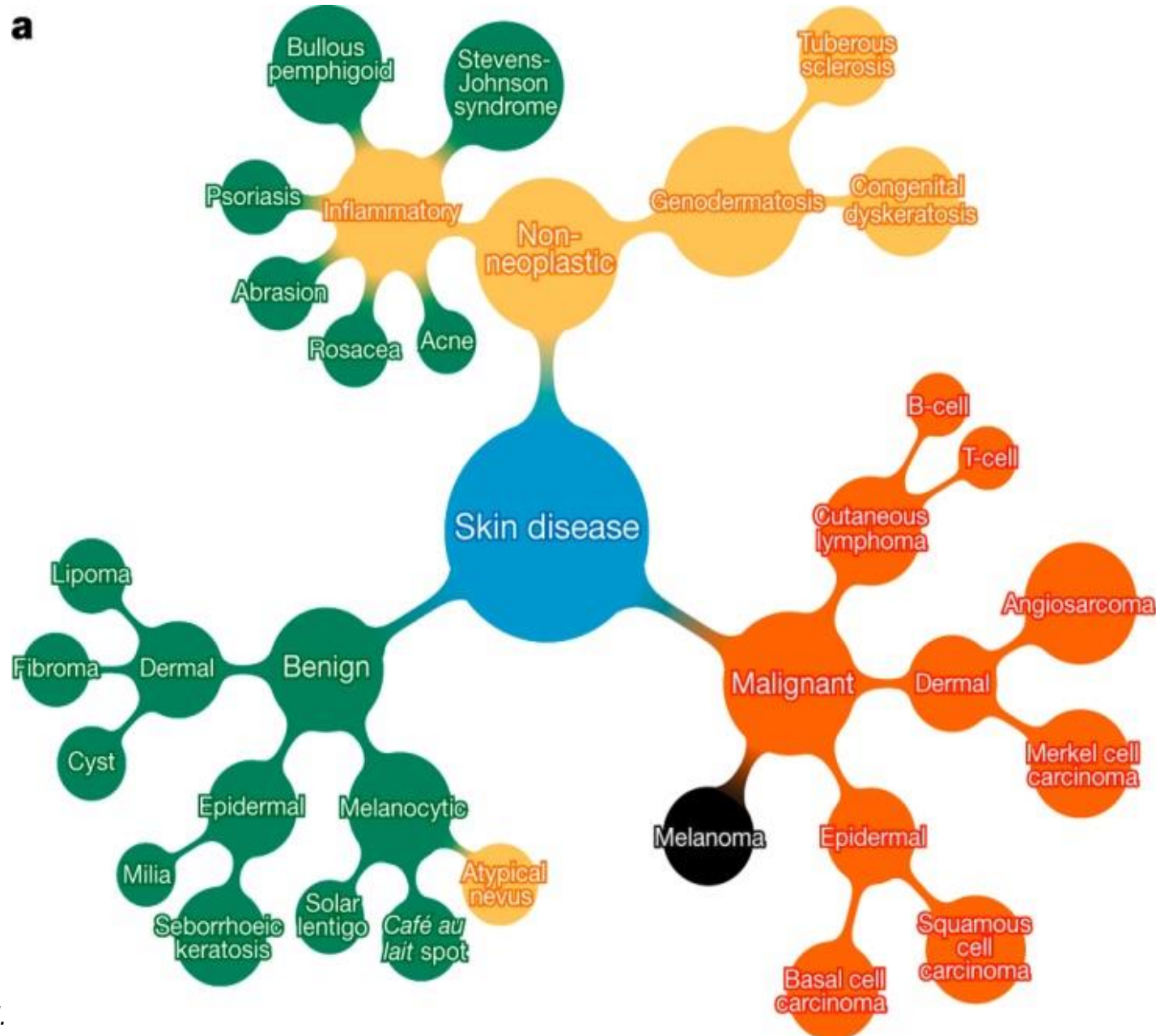
Разработка лекарств

“Интеллектуальные
медицинские записи”

Предсказание
вспышек заболеваний

Предсказание рака кожи

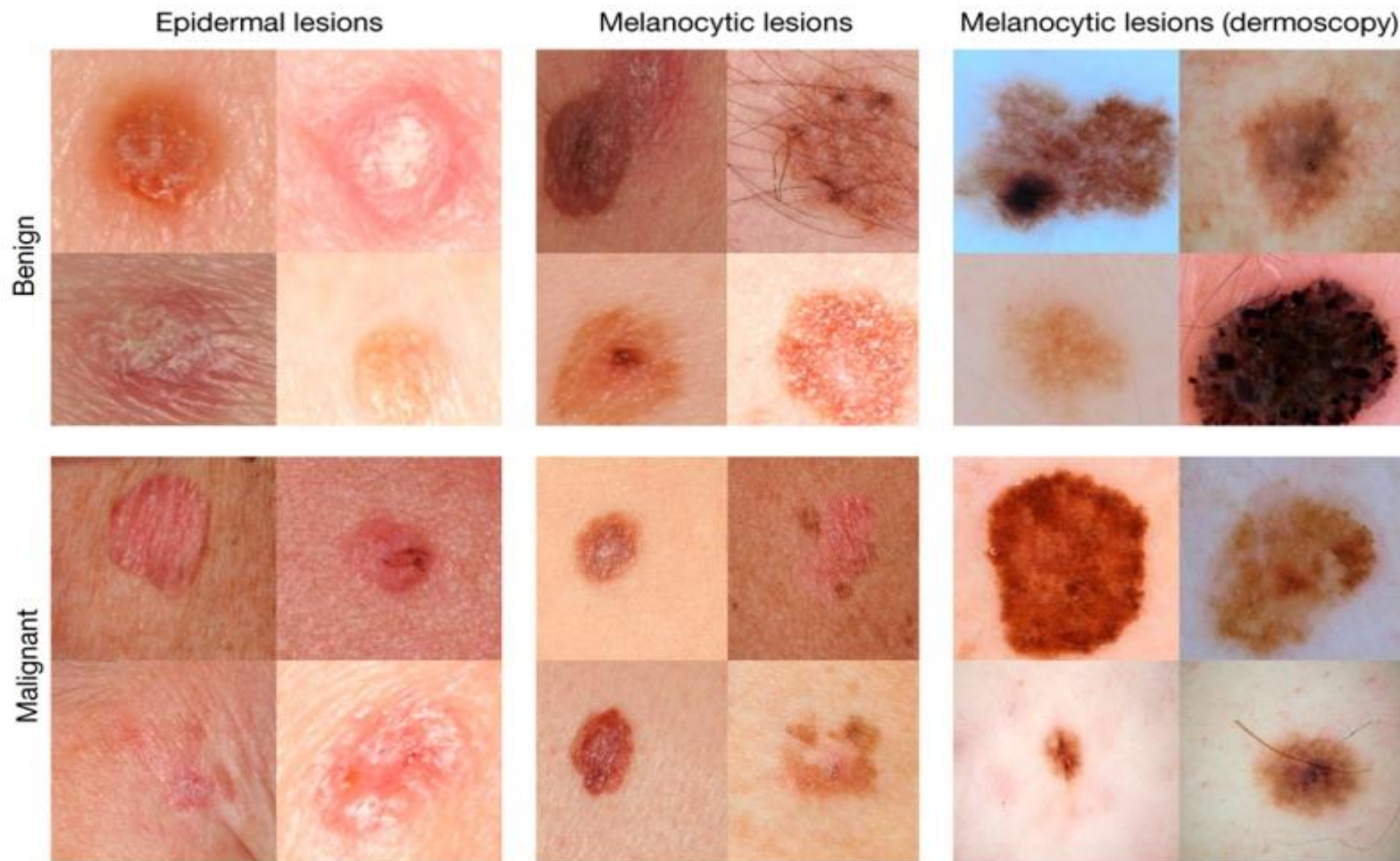
a



99%
VS
14%

Предсказание рака кожи

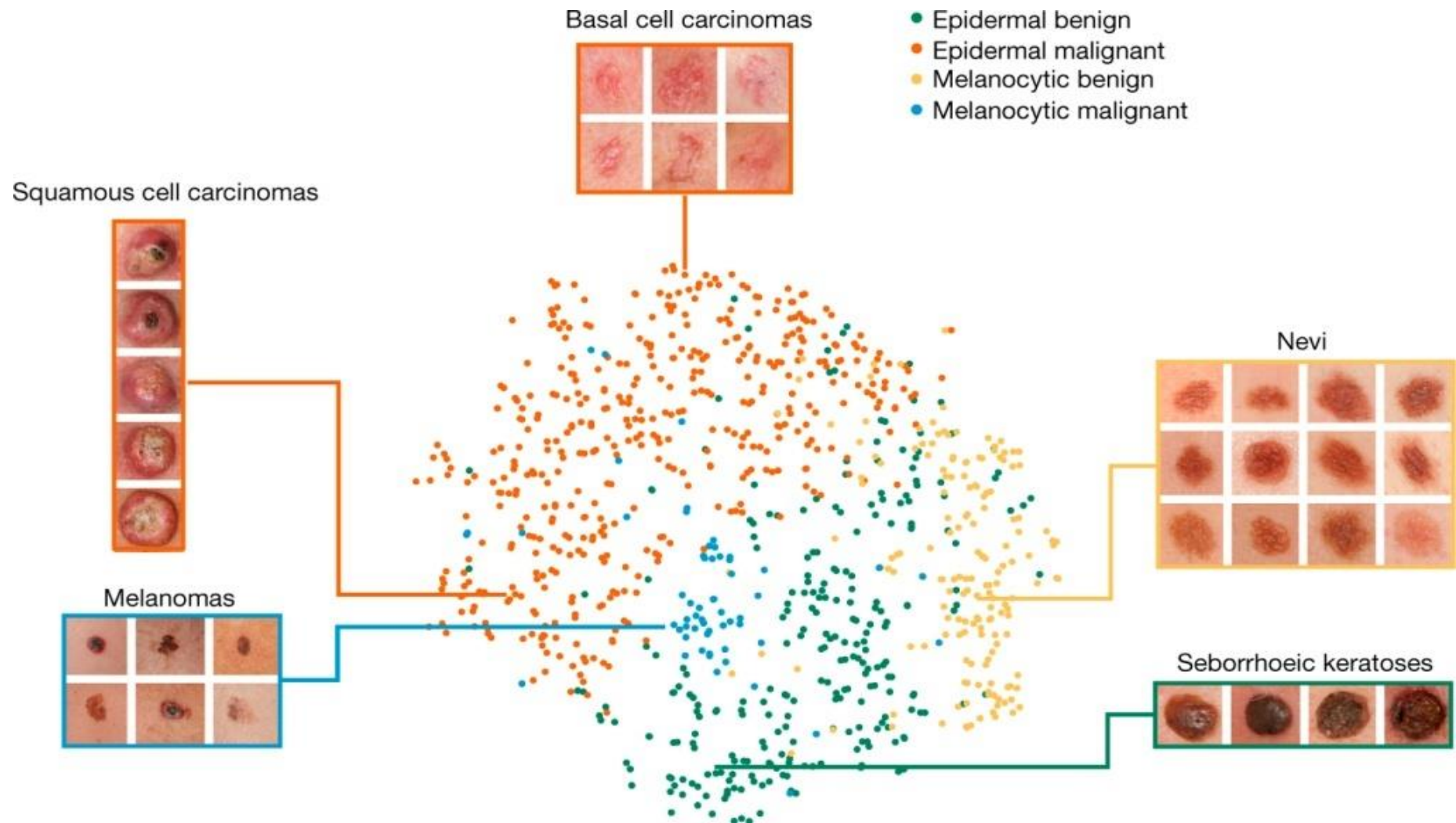
b



Train:
29,450 clinical
images

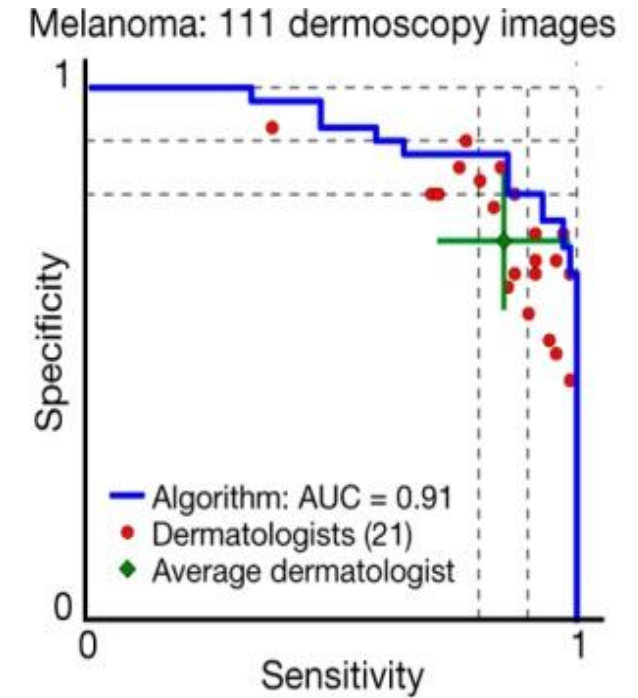
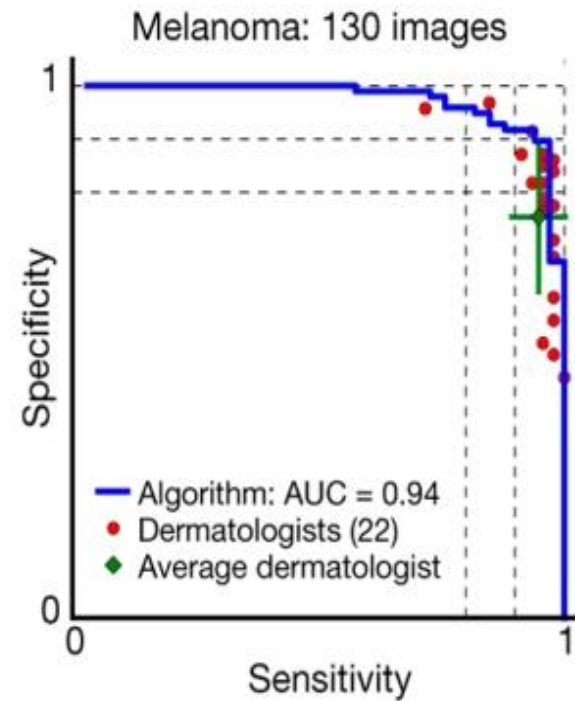
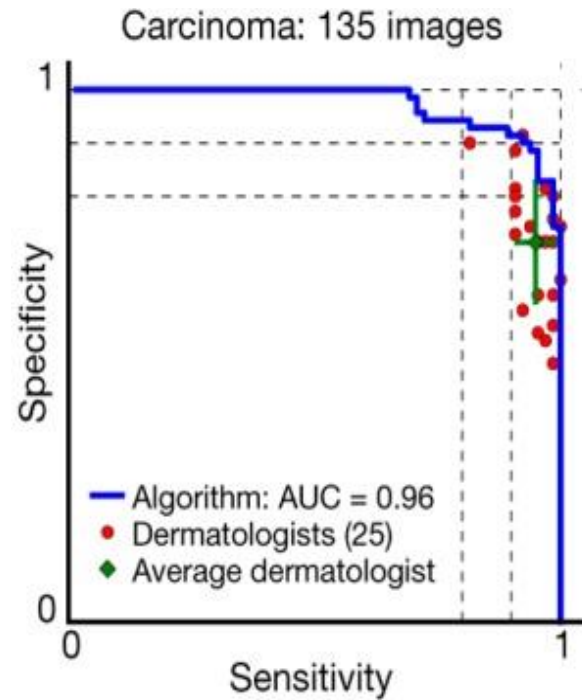
21 board-certified
dermatologist

Предсказание рака кожи



A Esteva et al. *Nature* 1–4 (2017)
doi:10.1038/nature21056

Предсказание рака кожи



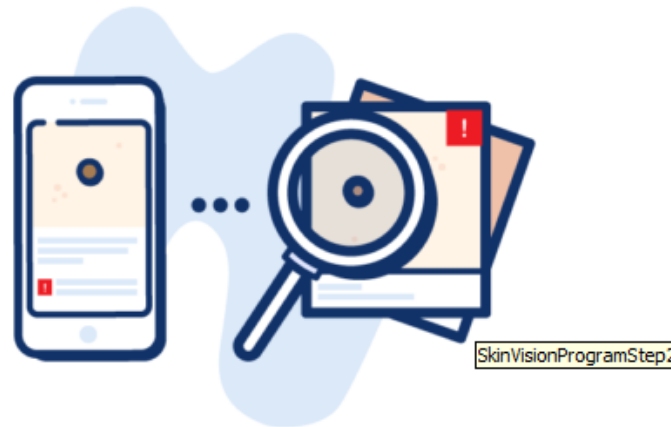
GETTING STARTED

SkinVision Service: Keep your skin healthy with regular self-checks



Take a photo of your skin spot

Open the SkinVision camera and take a photo of the skin spot that concerns you. You may need the help of someone else if your skin spot is in a difficult place.



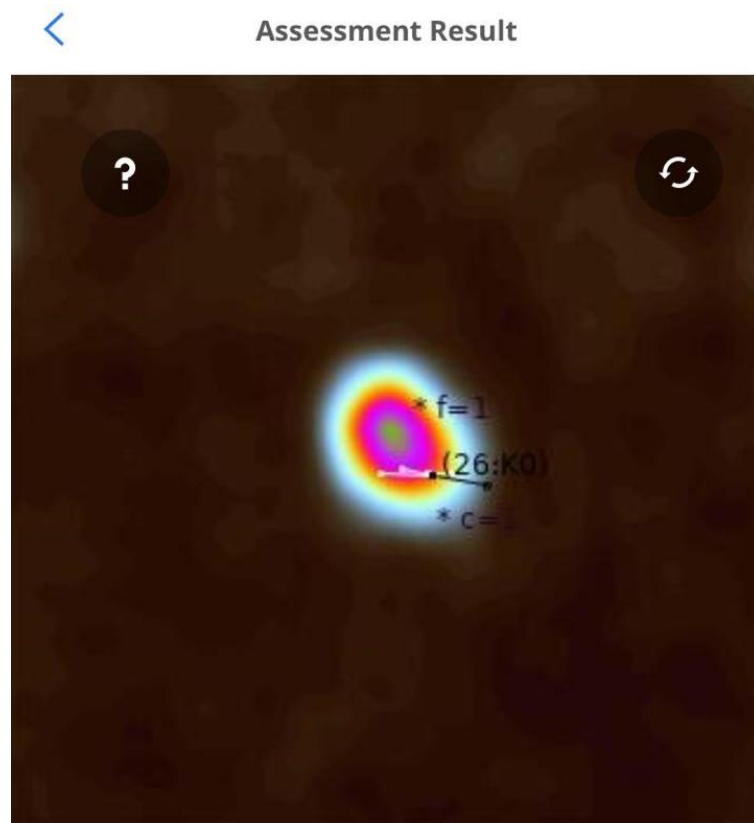
Receive your risk indication

SkinVision gives you a low, medium, or high-risk indication within 30 seconds. In case of a high-risk indication, you get free advice from our in-house dermatologists on the next steps to take.



Schedule your next check

Store your photos securely and set up reminders for your next skin checks. Stay on top of your skin health with regular checks.



Your assessment is

● MEDIUM RISK

What does this mean?

Our algorithm performed a check on your photo and did not find signs of skin cancer

Применение ИИ в медицине



```
graph TD; A[Применение ИИ в медицине] --> B[Диагностика болезней]; A --> C[Персонализированная медицина]; A --> D[Разработка лекарств]; A --> E["Интеллектуальные медицинские записи"]; A --> F[Предсказание вспышек заболеваний];
```

Диагностика
болезней

Персонализированная
медицина

Разработка лекарств

“Интеллектуальные
медицинские записи”

Предсказание
вспышек заболеваний

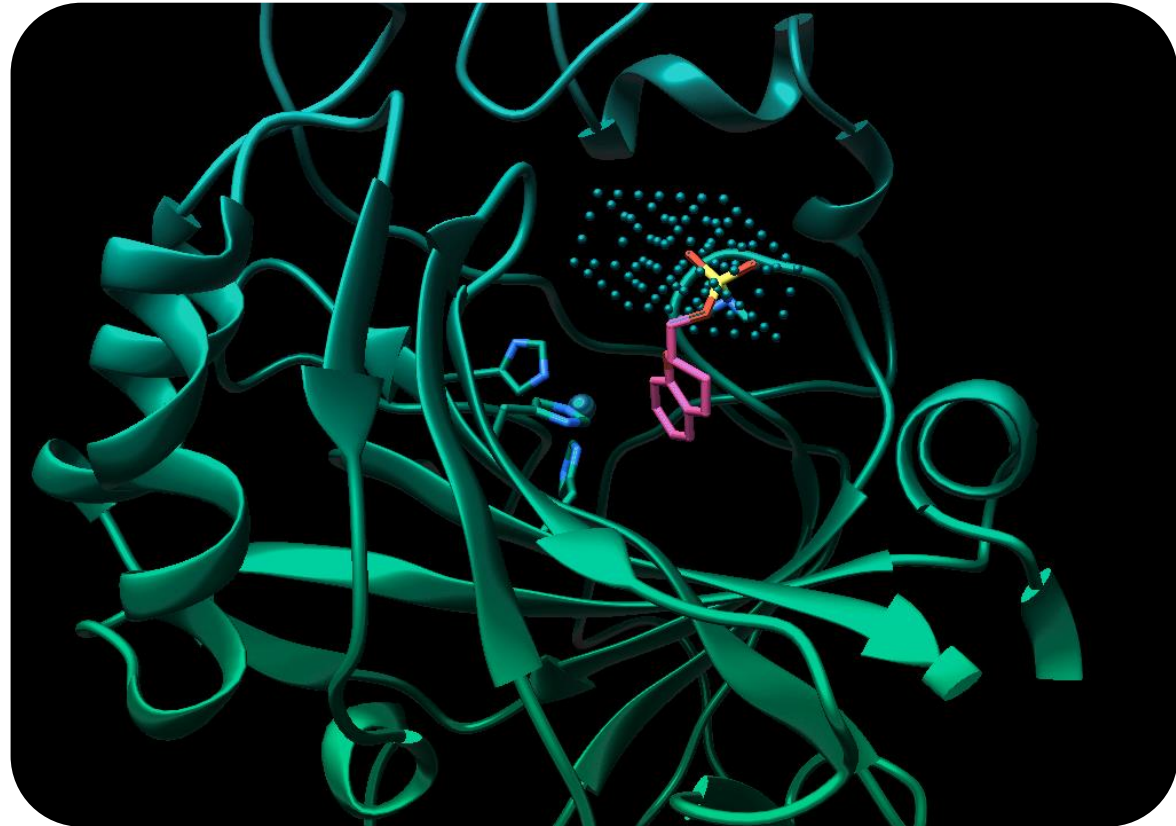
Разработка лекарств

> 10 лет

- 1) Генерируем гипотезы.
 - определяем белки мишени
- 2) Перебираем соединения, связывающиеся с мишенью.
- 3) Выбираем лучшие варианты
- 4) Тестируем в биологической системы
- 5) Клинические испытания

Чем может помочь искусственный интеллект?

- Выявление биологических мишеней
- Создание новых молекул с определенным набором свойств.
- **Перебор молекул, связывающихся с мишенью**



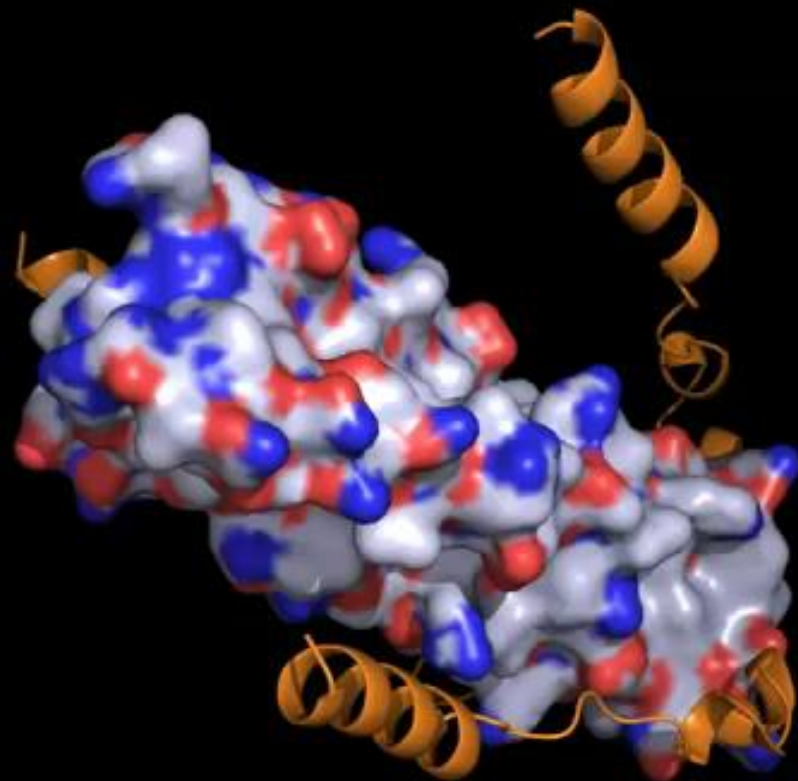
Нейронные сети и лихорадка Эбола в 2014-2015г.

- 28,183 случая
- 11,306 смертей
- (40% смертность)

37

Мишень:
glycoprotein 2

Нейронные сети и лихорадка Эбола в 2014-2015г.



7,000 молекул, прошедших испытания на безопасность и использующихся для лечения других болезней

Применение ИИ в медицине

The diagram features a background image of a tipped-over glass bottle of pills and a pile of various colored capsules (yellow, green, pink, white, and dark red) on a white surface. A blue banner at the top contains the title. Five blue arrows originate from the banner and point to different text boxes: 'Диагностика болезней' (pointing to the left), 'Персонализированная медицина' (pointing down into the pile of pills), 'Разработка лекарств' (pointing down towards the center), 'Предсказание вспышек заболеваний' (pointing down and to the right, and is circled in red), and '“Интеллектуальные медицинские записи”' (pointing to the right).

Диагностика
болезней

Персонализированная
медицина

Разработка лекарств

“Интеллектуальные
медицинские записи”

Предсказание
вспышек заболеваний

Предсказание вспышек малярии



- 2017 г. 219 миллионов случаев
- 2017 г. 435 000 смертей

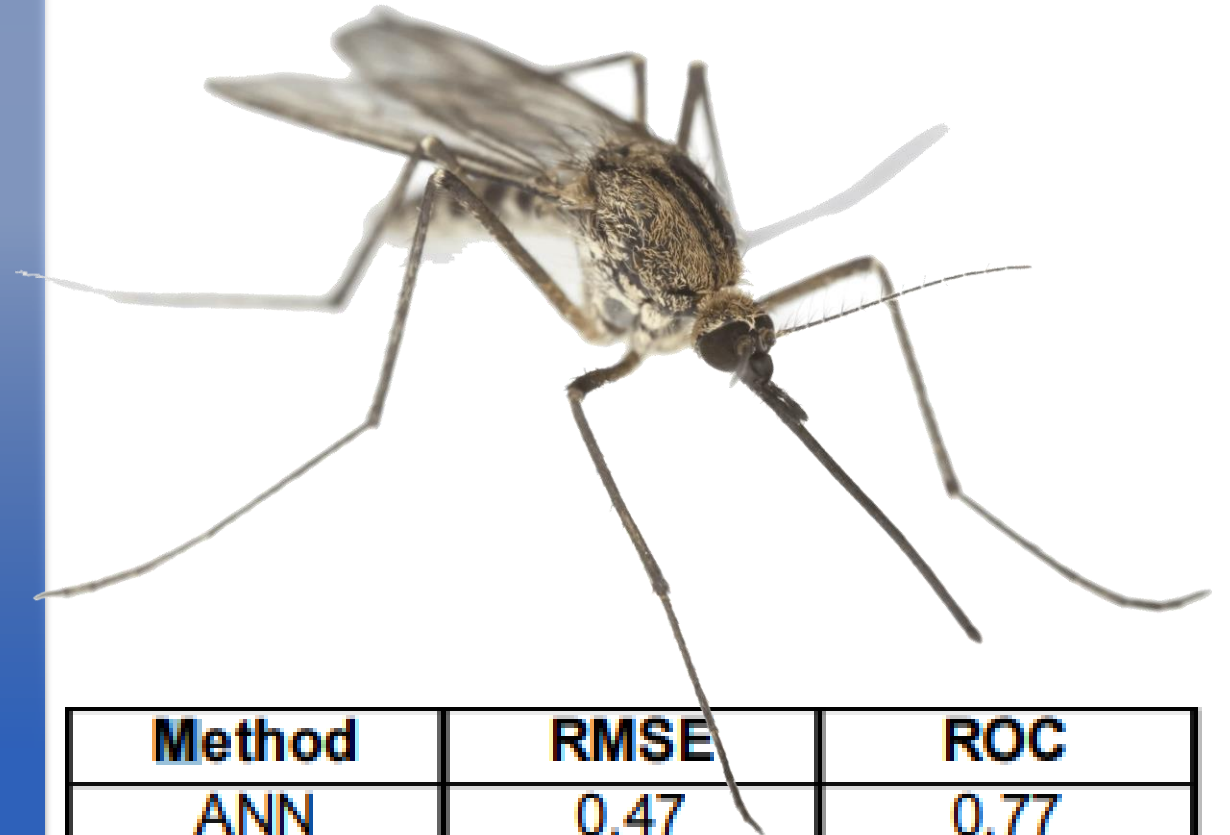
Факторы, приводящие к вспышке

☐ Климатические

- Температура
- Осадки
- Влажность
- Наводнение
- Засуха
- стихийные бедствия

☐ Не климатические

- различия между людьми-хозяевами
- миграция людей
- строительные работы



Method	RMSE	ROC
ANN	0.47	0.77
SVM	0.12	0.89

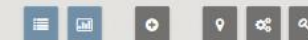
2011 – 2014:
1680 samples

Max.Temperature	Min.Temperature	Avg. Humidity	Rainfall	Positive	pf	Outbreak
29	18	49.74	0.00	2156	112	No
34	23	83.27	15.22	10717	677	Yes
40	23	50.74	0.00	1257	127	No
34	24	59.16	9.06	4198	211	No
34	27	73.23	0.00	11808	712	Yes
31	24	88.77	41.40	10881	648	Yes
33	24	77.94	23.88	8830	459	Yes
31	24	84.57	11.15	9693	482	No
36	24	53.40	2.12	9310	549	No
32	23	57.50	0.00	13154	838	Yes
34	18	59.40	0.00	2197	136	No
42	24	49.43	2.19	3362	213	No
45	32	34.74	0.38	416	26	No
43	28	69.07	4.65	7514	410	No
33	23	80.97	6.92	10990	390	Yes
32	24	87.32	11.92	6536	338	No
40	27	63.97	0.00	11169	776	Yes
39	25	47.52	0.00	8131	312	No
36	26	72.78	3.54	5138	213	No
31	23	73.35	4.97	10659	612	Yes
30	23	86.81	7.21	9041	418	No
30	22	78.80	3.12	11265	404	Yes
33	22	73.71	1.75	9233	212	No

Table 1.1: Sample data collected from different sources.

1028 alerts for Alerts from past week ▾

9 языков



проверенные
официальные
отчеты

отчеты
очевидцев

онлайн-
агрегаторы
новостей

дискуссии
экспертов



HealthMap

О нас Mobile Проекты Disease Daily



Войти



Применение ИИ в медицине



Диагностика
болезней

Персонализированная
медицина

Разработка лекарств

Предсказание
вспышек заболеваний

“Интеллектуальные
медицинские записи”

Пример электронной медицинской карты

Handy patients enterprise edition

File Edit View Help

 David (8 month and 10 day)
John (2 years and 3 month)
Mother: Teacher
Father: Financial advisor
Parents: Married

Last: **Anderson** P
First: **David** Boy
Birth: 5 January 2009
Age: 8 month and 10 days Patient nb: 3

Appointments

Forms: Meeting (Doctor), Full status (Doctor), Assistant, Billing, Reports, Statistics

Sheets: O: Neurologic, O: Vascular, O: Cardiac, O: Respiratory, O: Abdomen, Exams, Radiology, Summary, Patient documents, Letter

SOAP: Sum. T, R-V T, P, PC, Admission Agenda

Meetings

Meeting	Date	Time
2 month checkup	5 Mar 09	2m.0d
1 month checkup	5 Feb 09	1m.0d
Respiration problem	22 Jan 09	17d
10 days checkup	13 Jan 09	8d
Control for return at home	9 Jan 09	4d
Birth	5 Jan 09	0d

Diagnosis

General ☒
My Diagnosis ☐
Social ☐

New documents

- Abdomen palpat - 15 Sep 2009
- Cardiac auscul - 15 Sep 2009

To Do: Send checkup

Assist: 1 Doc: 0

Notes

Father ask many questions, add 10 minutes to consultation

Current doctor: Dr Herman

Menu 1 Menu 2 Menu 3 Search

Digestive

Thursday, 22 Jan 2009

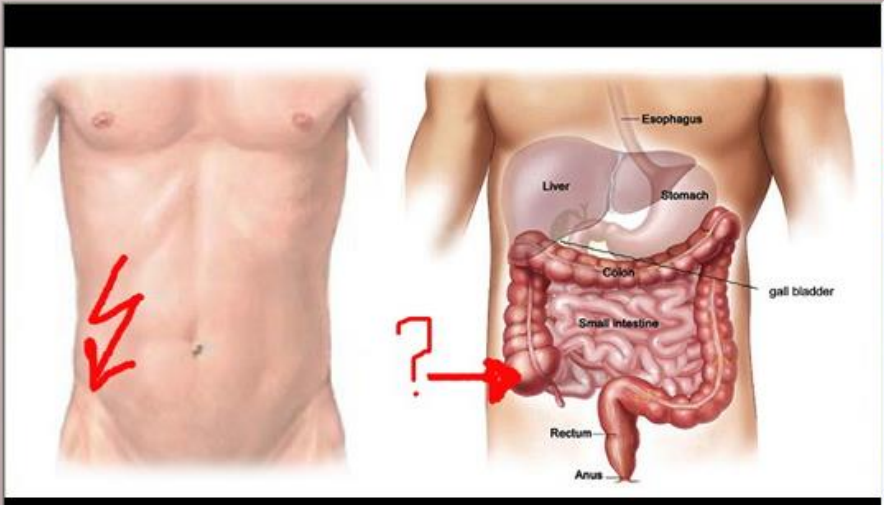
Digestive inspection
Normal

Digestive auscultation
Normal abdomen noises

Digestive palpation
Little pain on the right lower area

Liver
No hepatomegaly.

Rectal

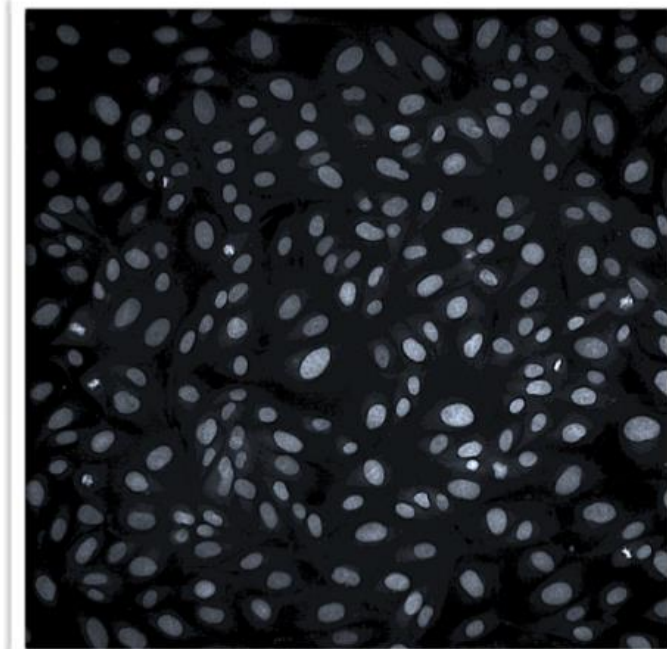


Page 1/1
Draw ☒
Mark ☐
Color
Pen 8

Documents manager

Previous page Next page

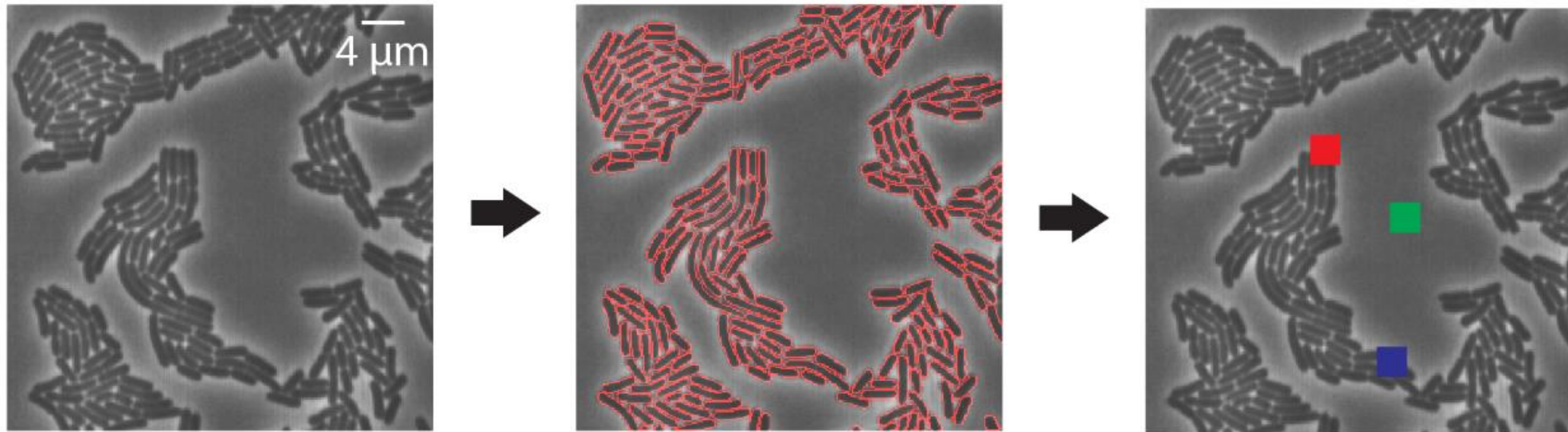
Применение в микроскопии



.....▶
Manually tuned
algorithm

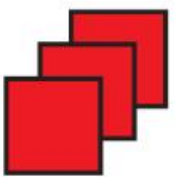


Применение в микроскопии



Manually annotate
training image

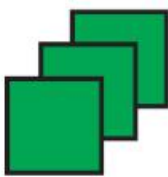
Sample training image
to construct training data



Cell boundary
images



Cell interior
images



Background
images

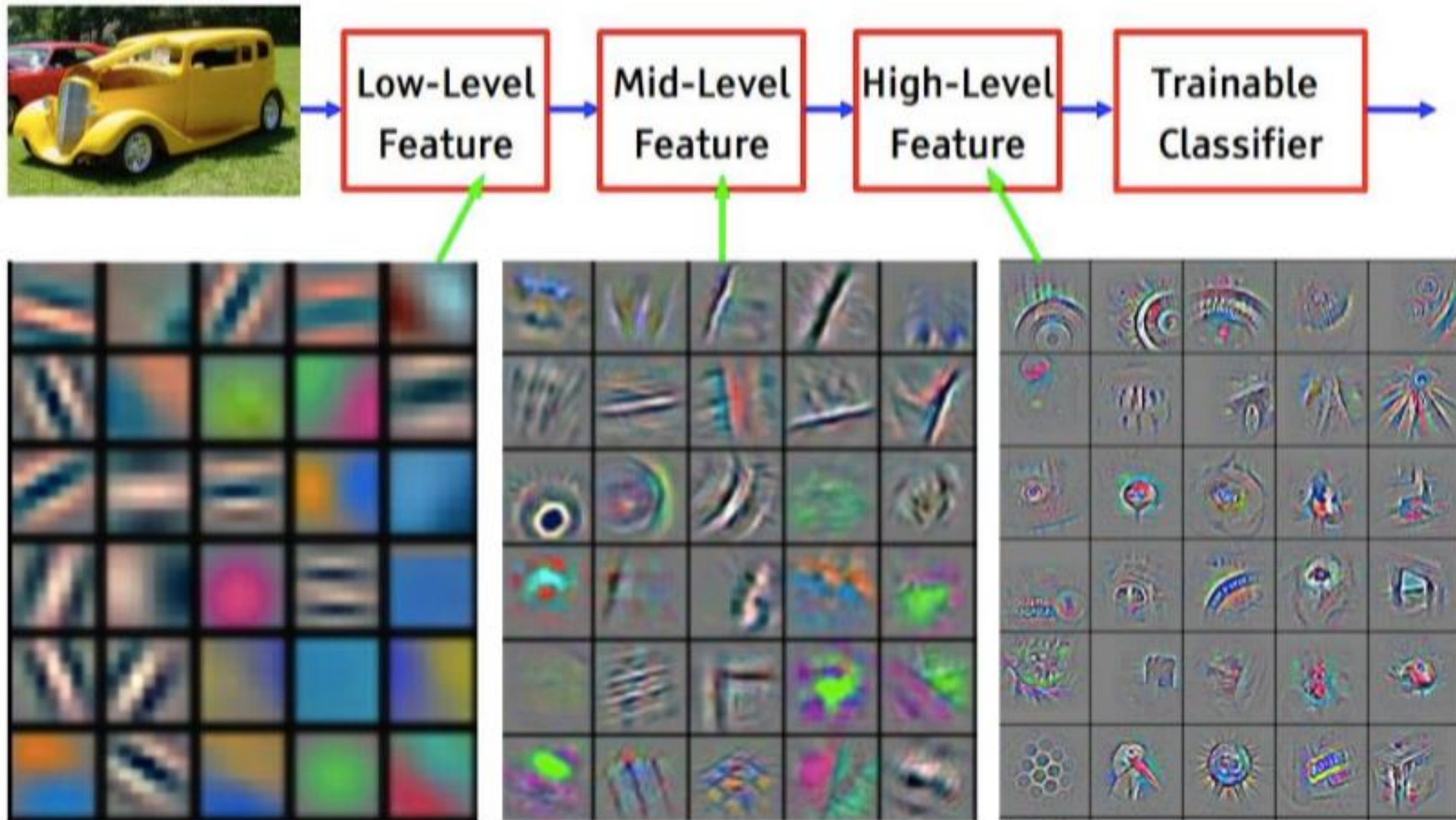


Train image classifier



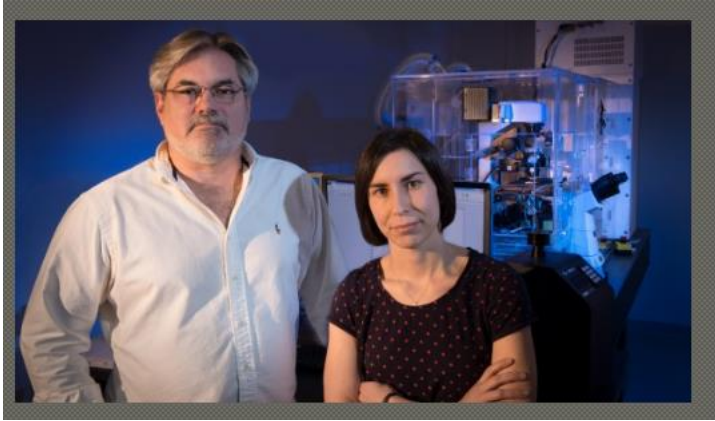
Apply image classifier
to new images

Deep Learning Automates the Quantitative Analysis of Individual Cells in Live-Cell Imaging Experiments
David A. Van ValenTakamasa Kudo... Markus W. Covert

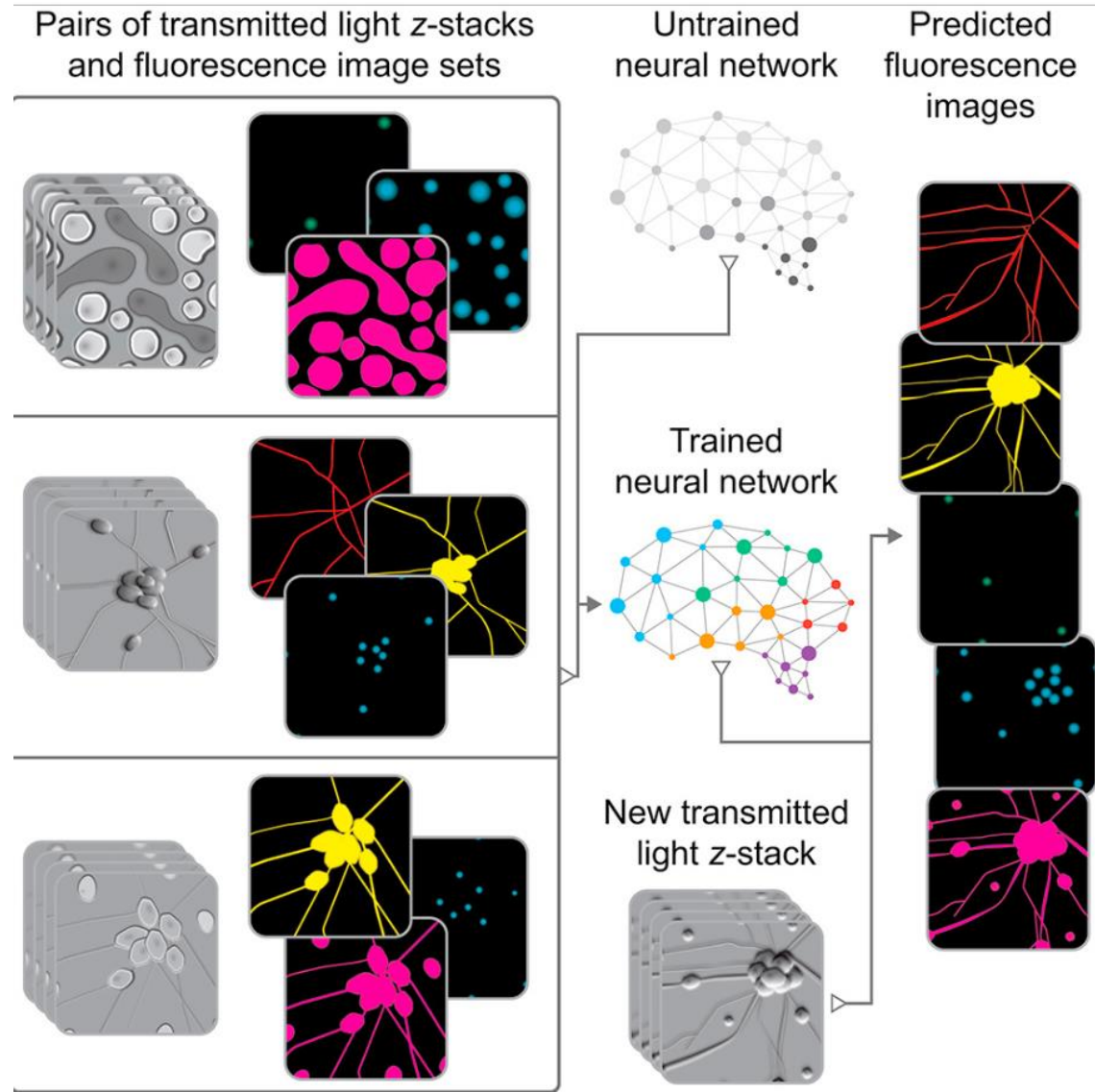


Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

Применение в микроскопии

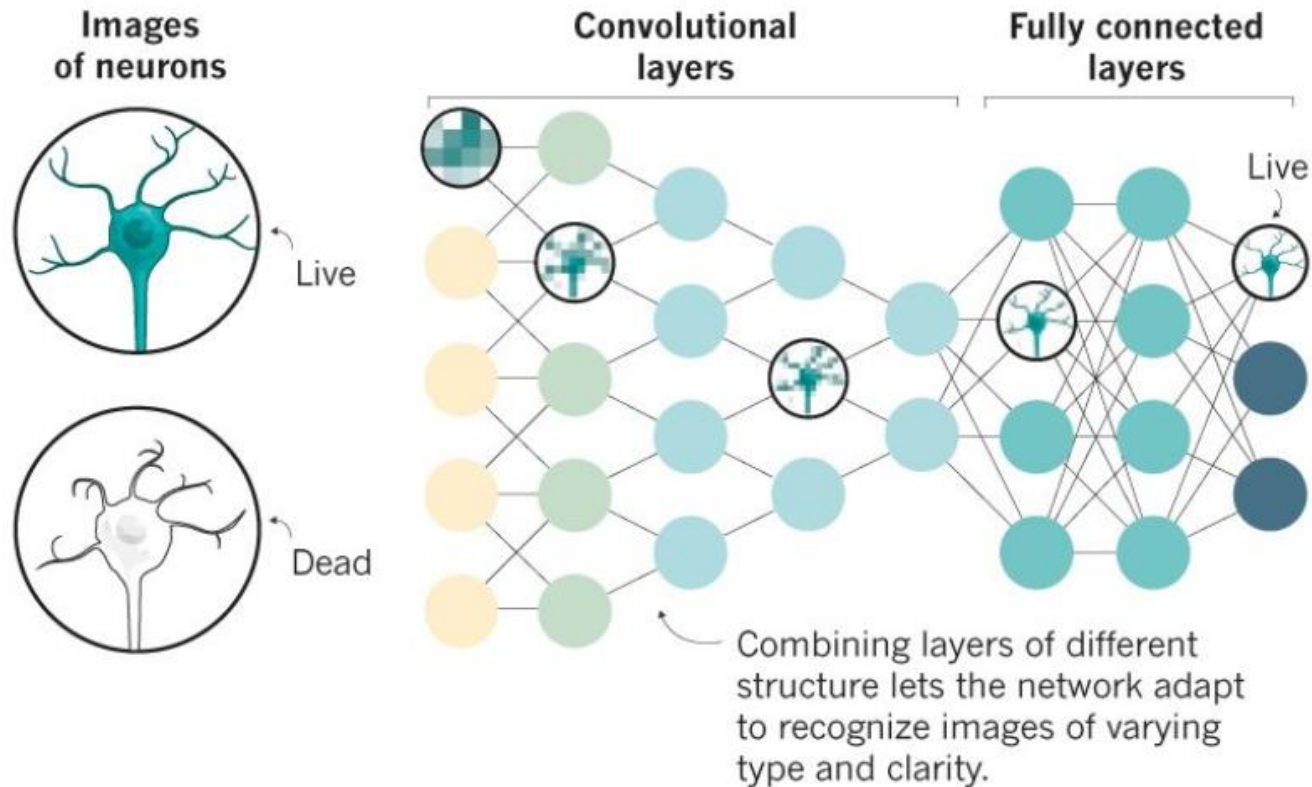
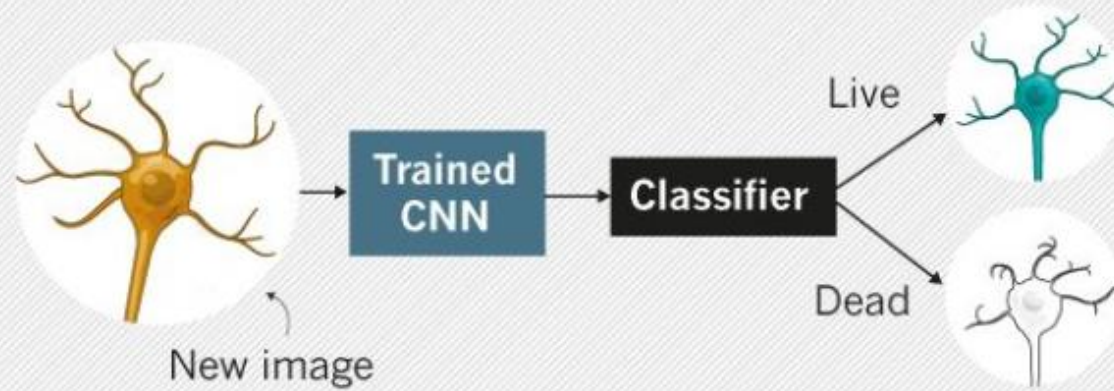


- Gladstone Institute of Neurological Disease in San Francisco

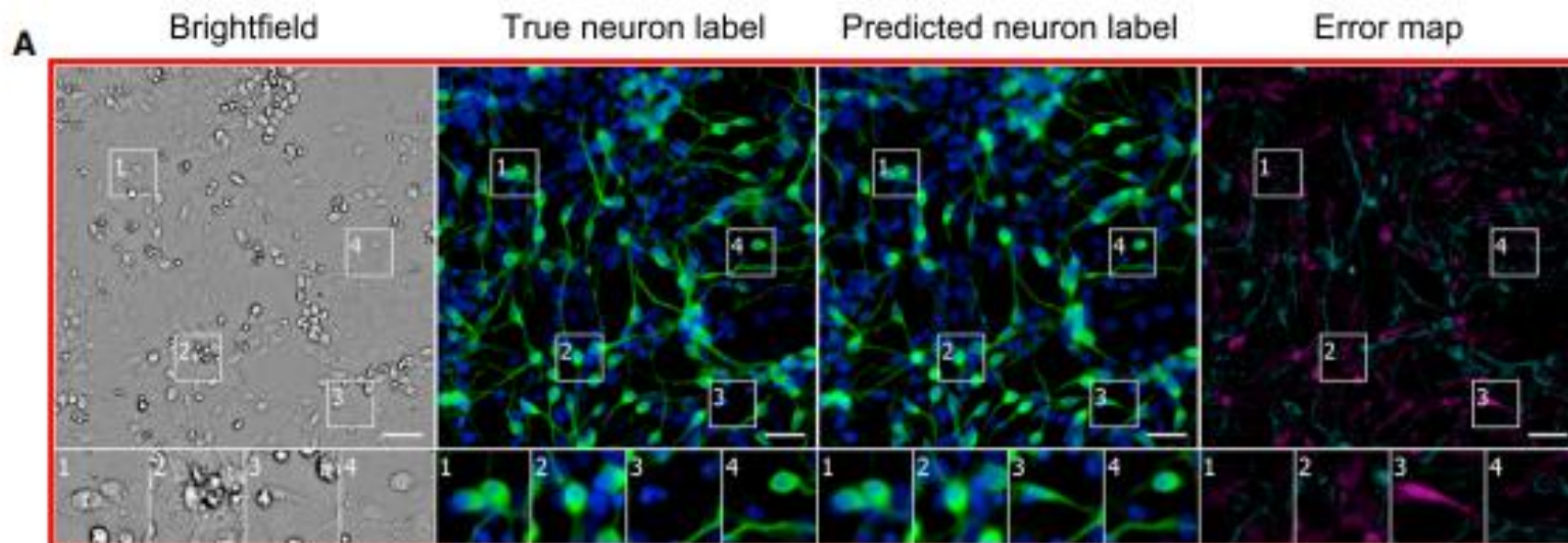
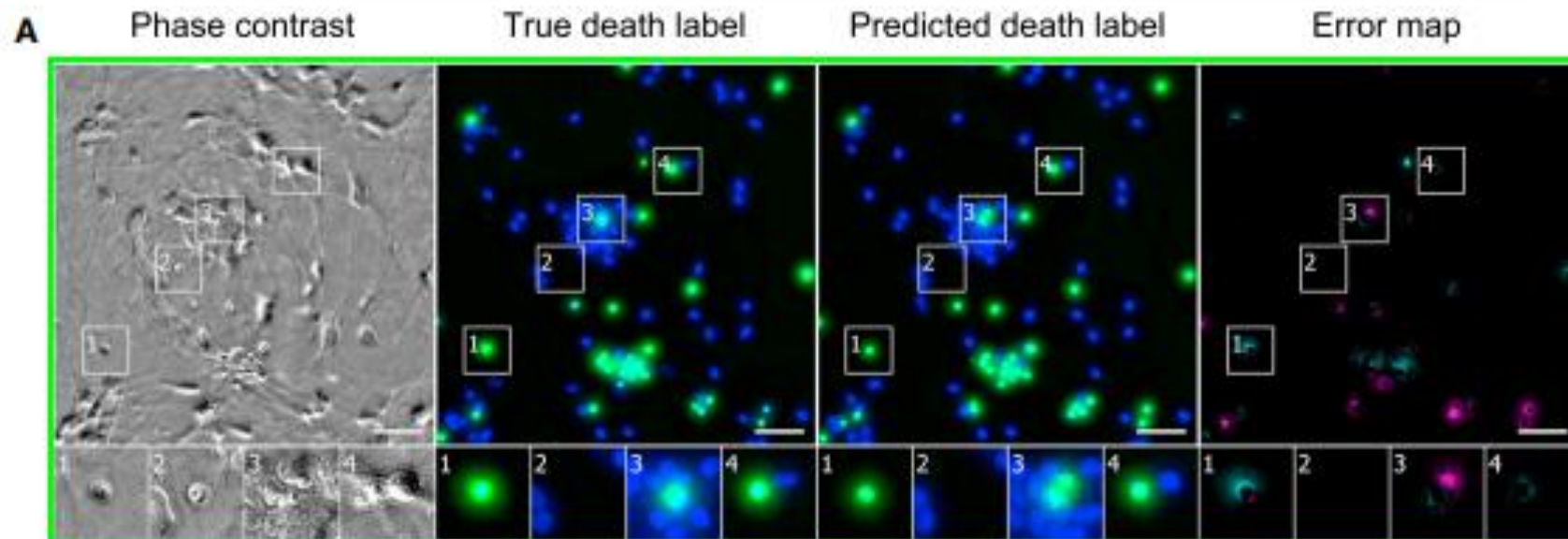


APPLICATION

Challenged with unlabelled images, the network assigns each cell as alive or dead with high accuracy.



Применение в микроскопии



Другие полезные инструменты

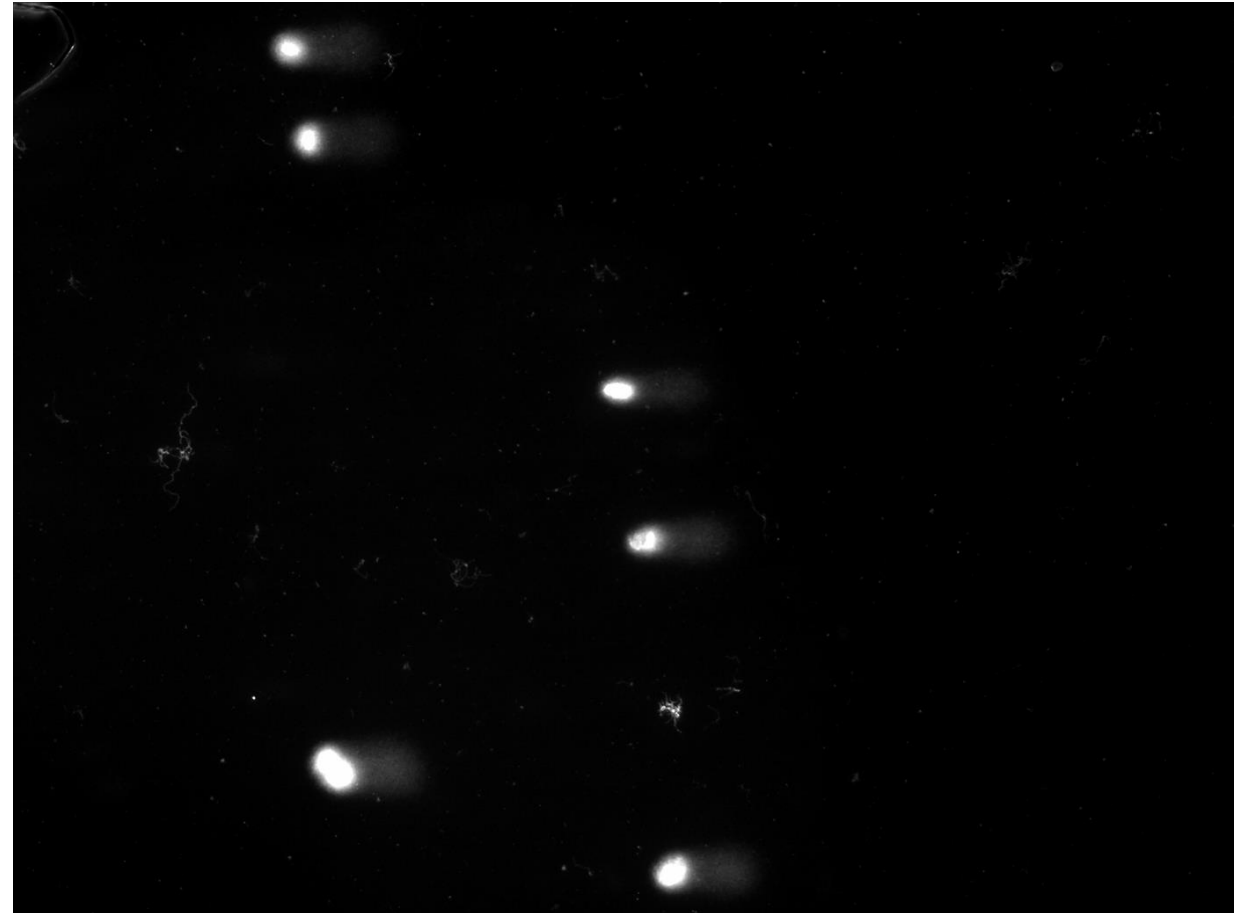
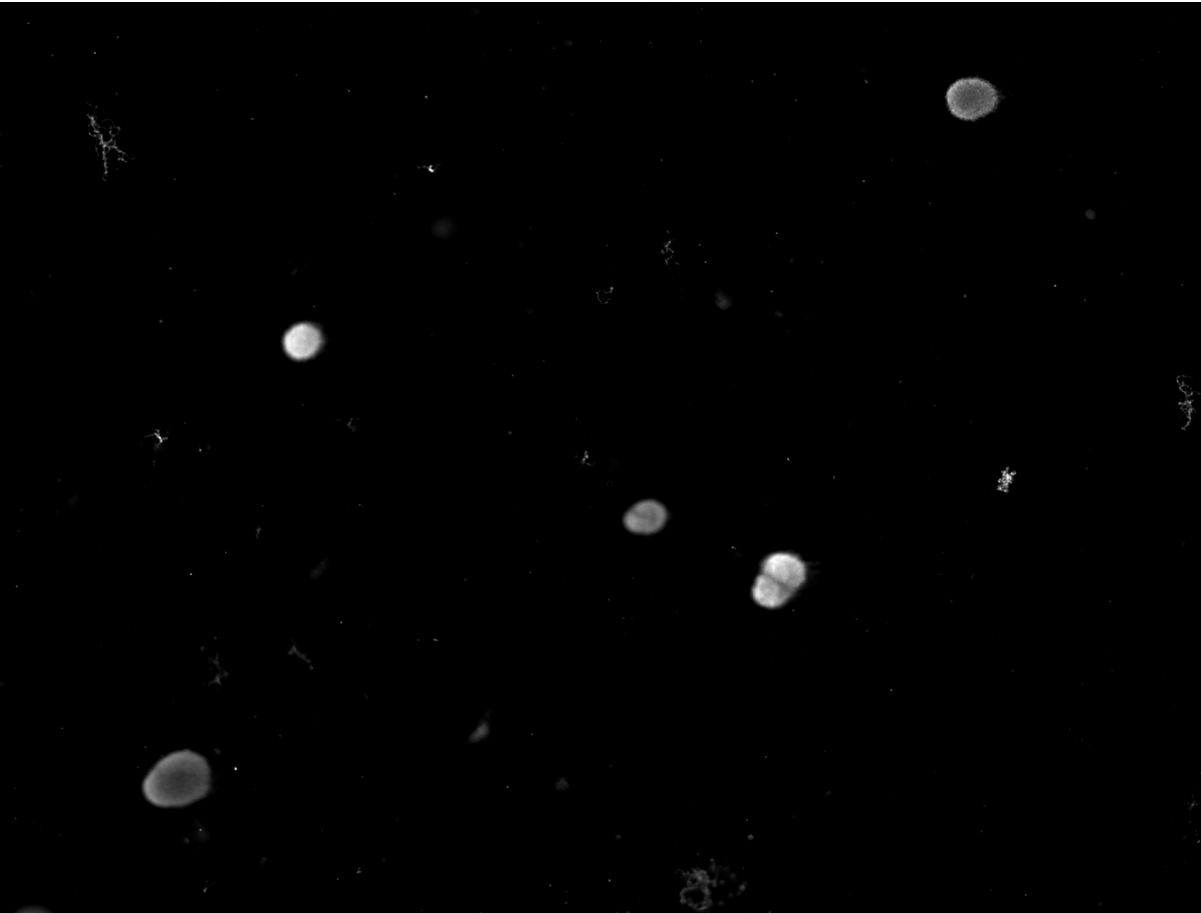
CellProfiler is free, open-source software for quantitative analysis of biological images.



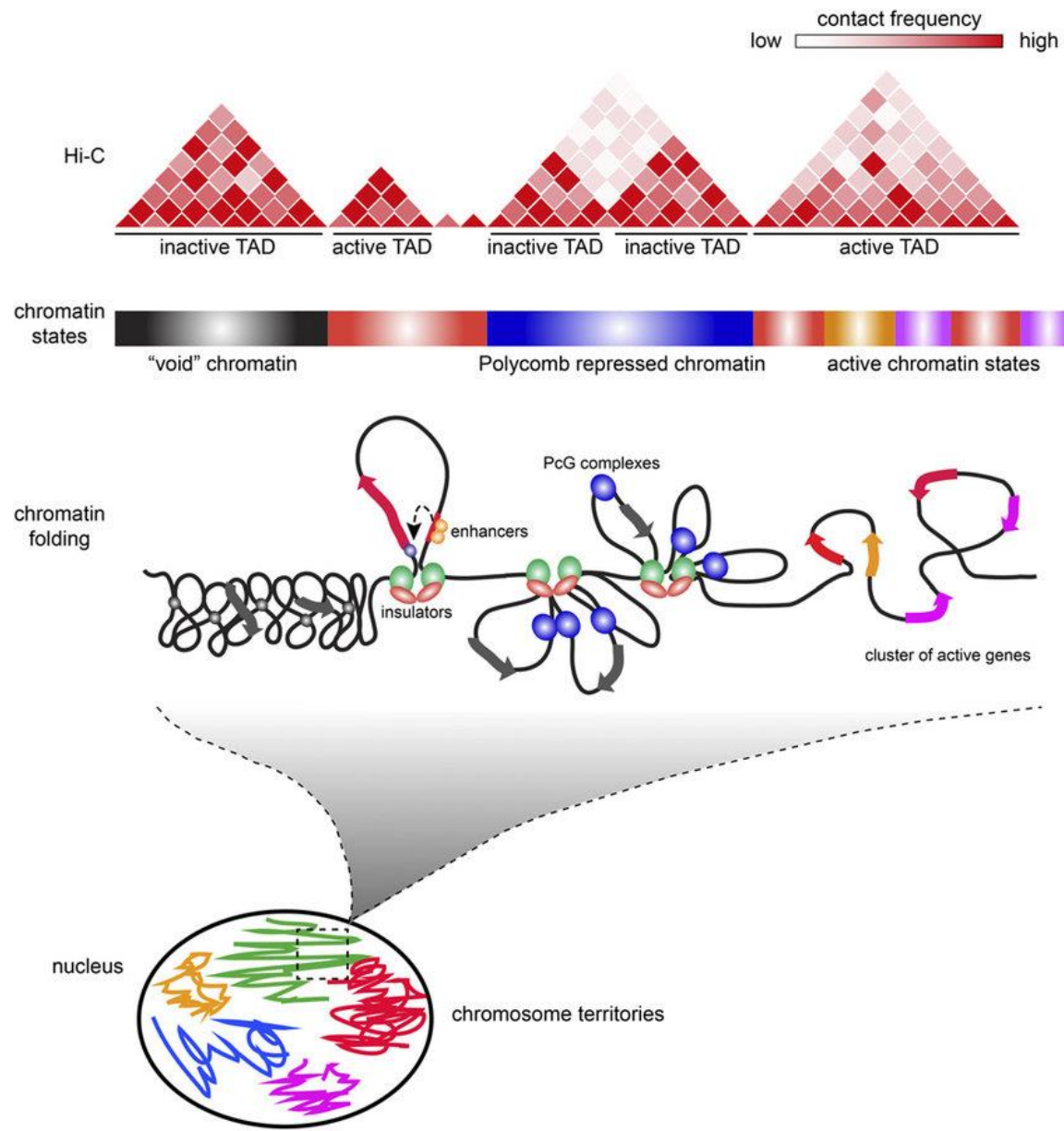
Anne Carpenter



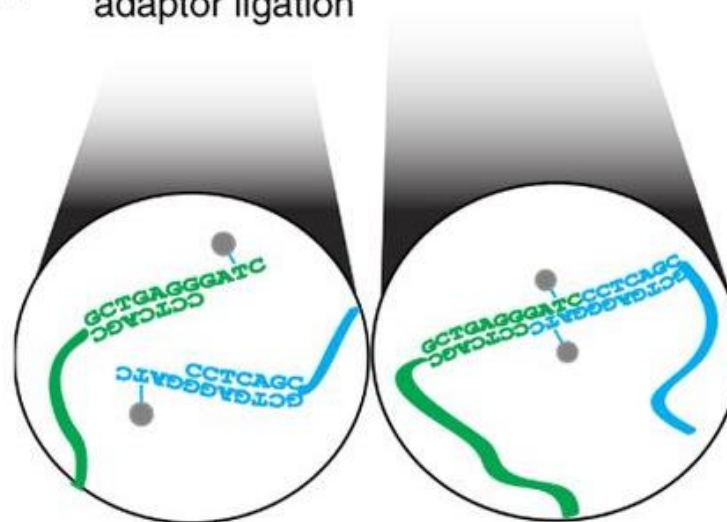
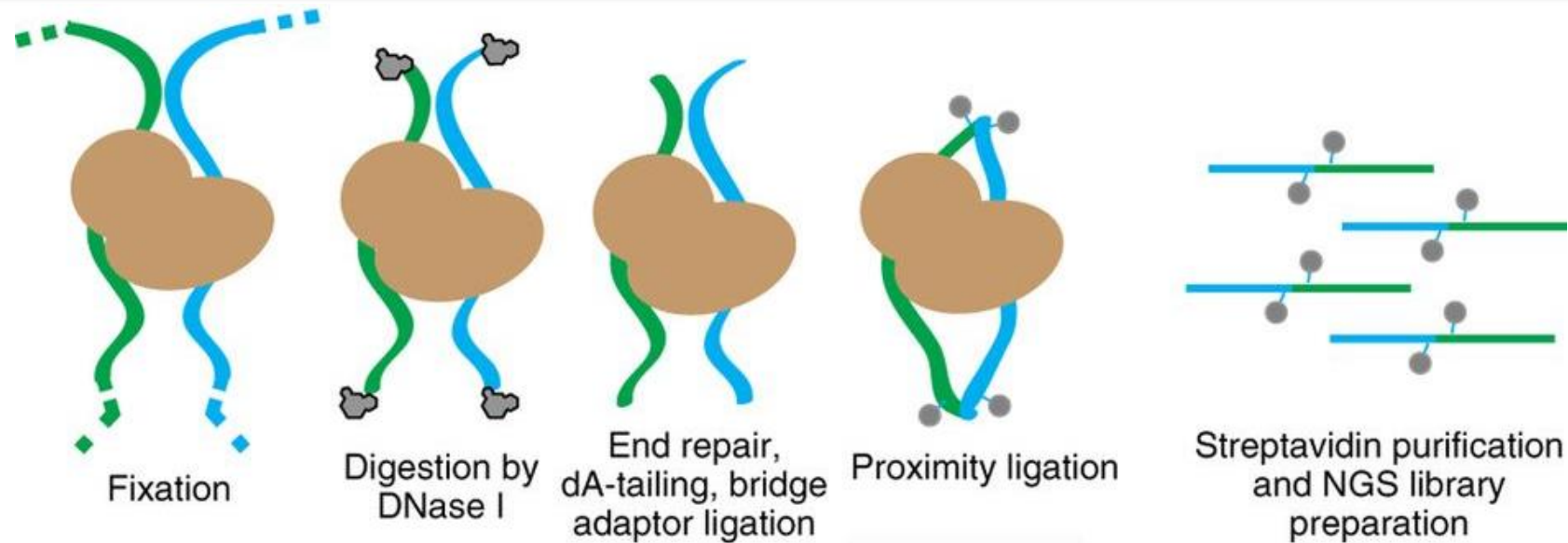
Cell profiler. Comet assay.



Предсказание трехмерной архитектуры хроматина

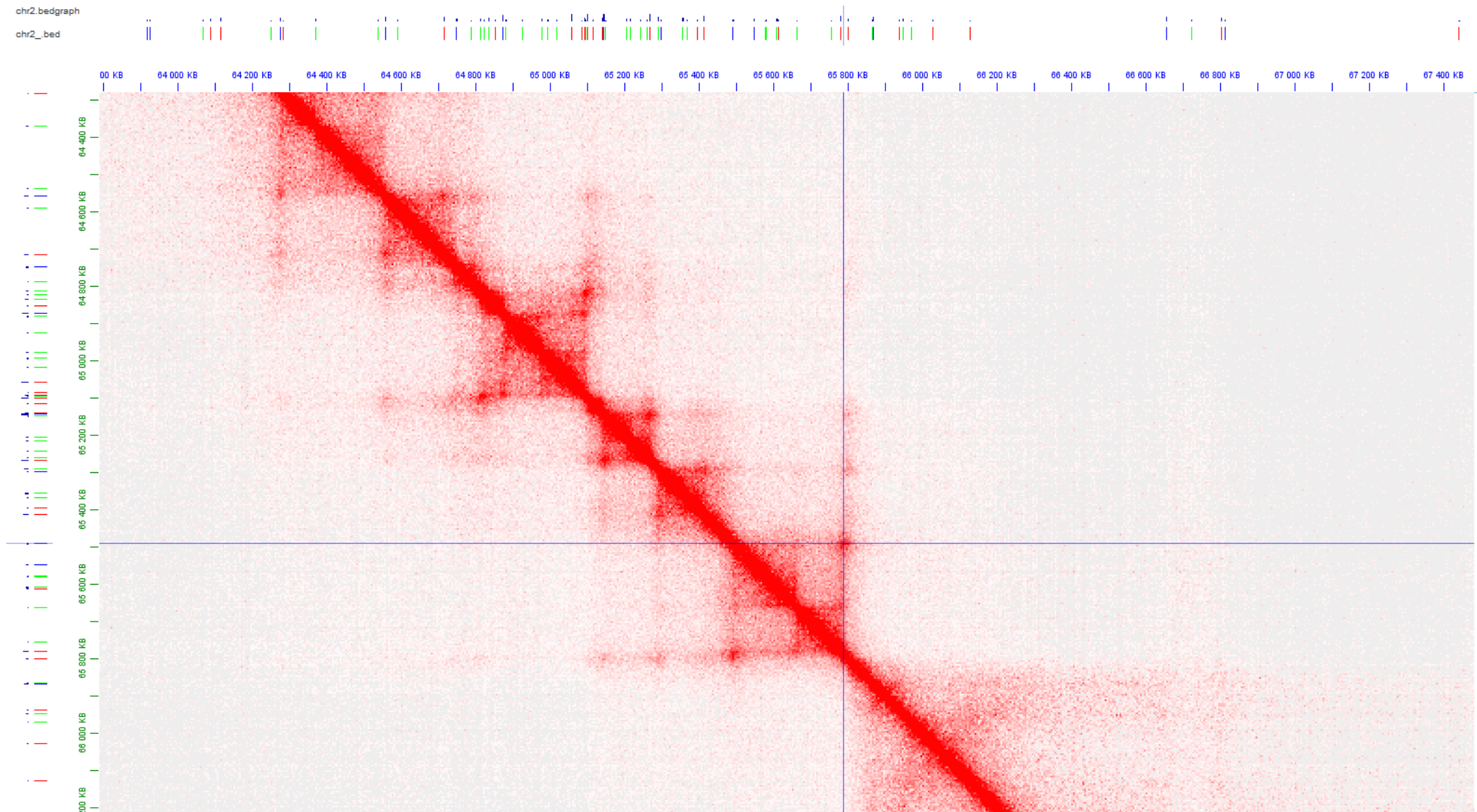


Предсказание трехмерной архитектуры хроматина

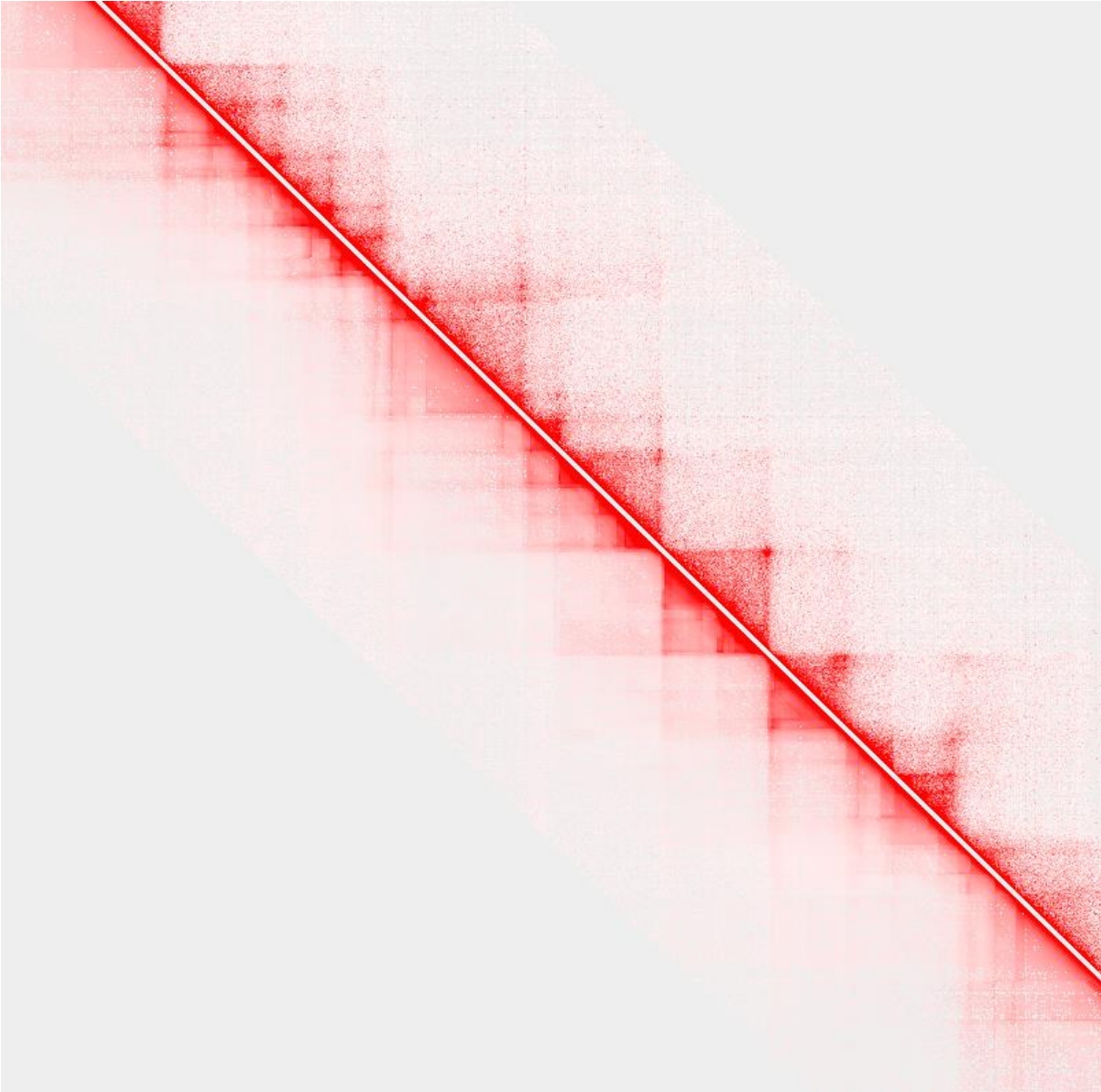


Метод Hi-C

Предсказание трехмерной архитектуры хроматина



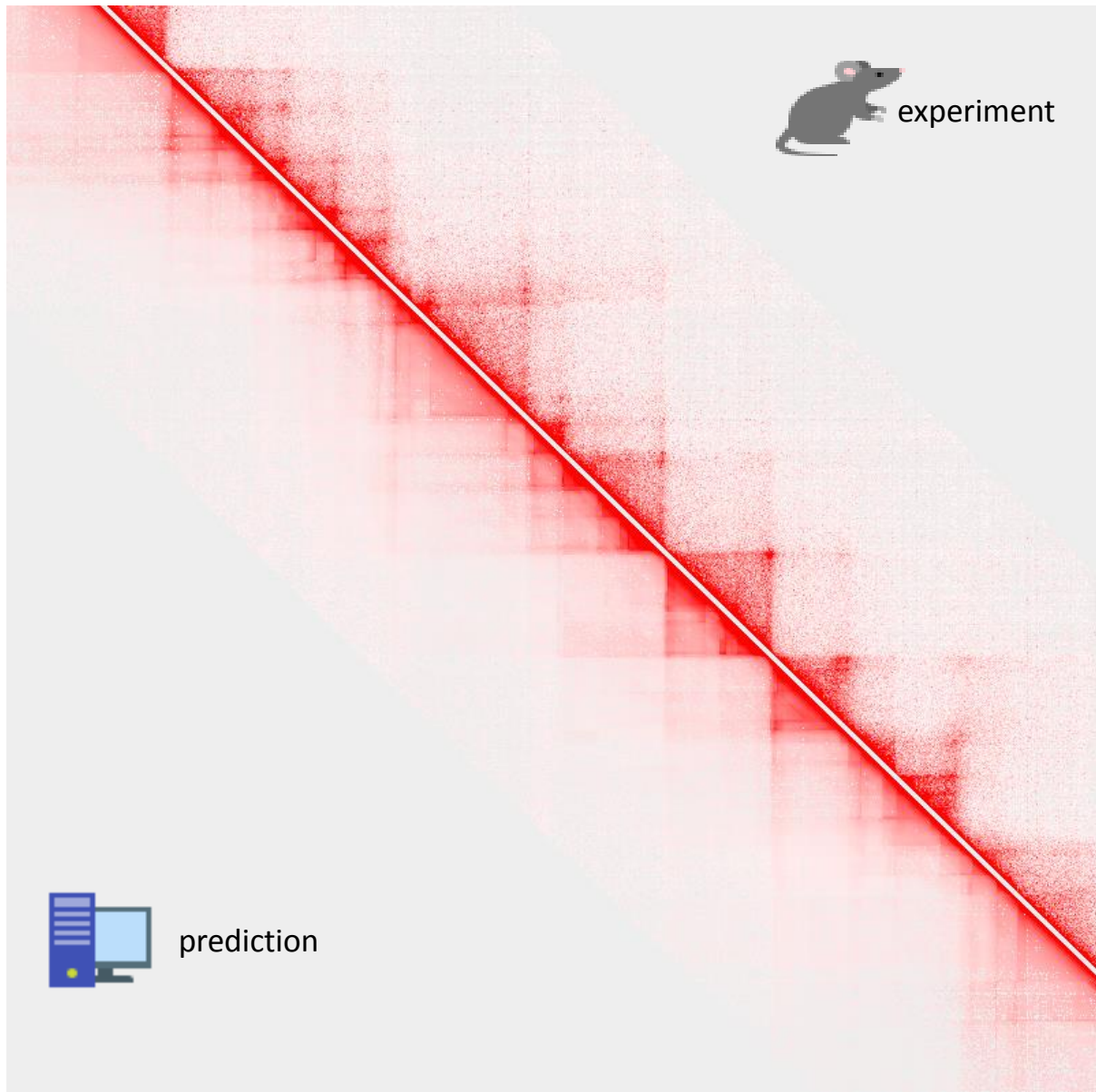
Предсказание трехмерной архитектуры хроматина



Признаки:

- ☐ Геномное расстояние
- ☐ ChIP-seq данные белка CTCF
- ☐ Ориентация CTCF
- ☐ Другие ChIP-seq данные

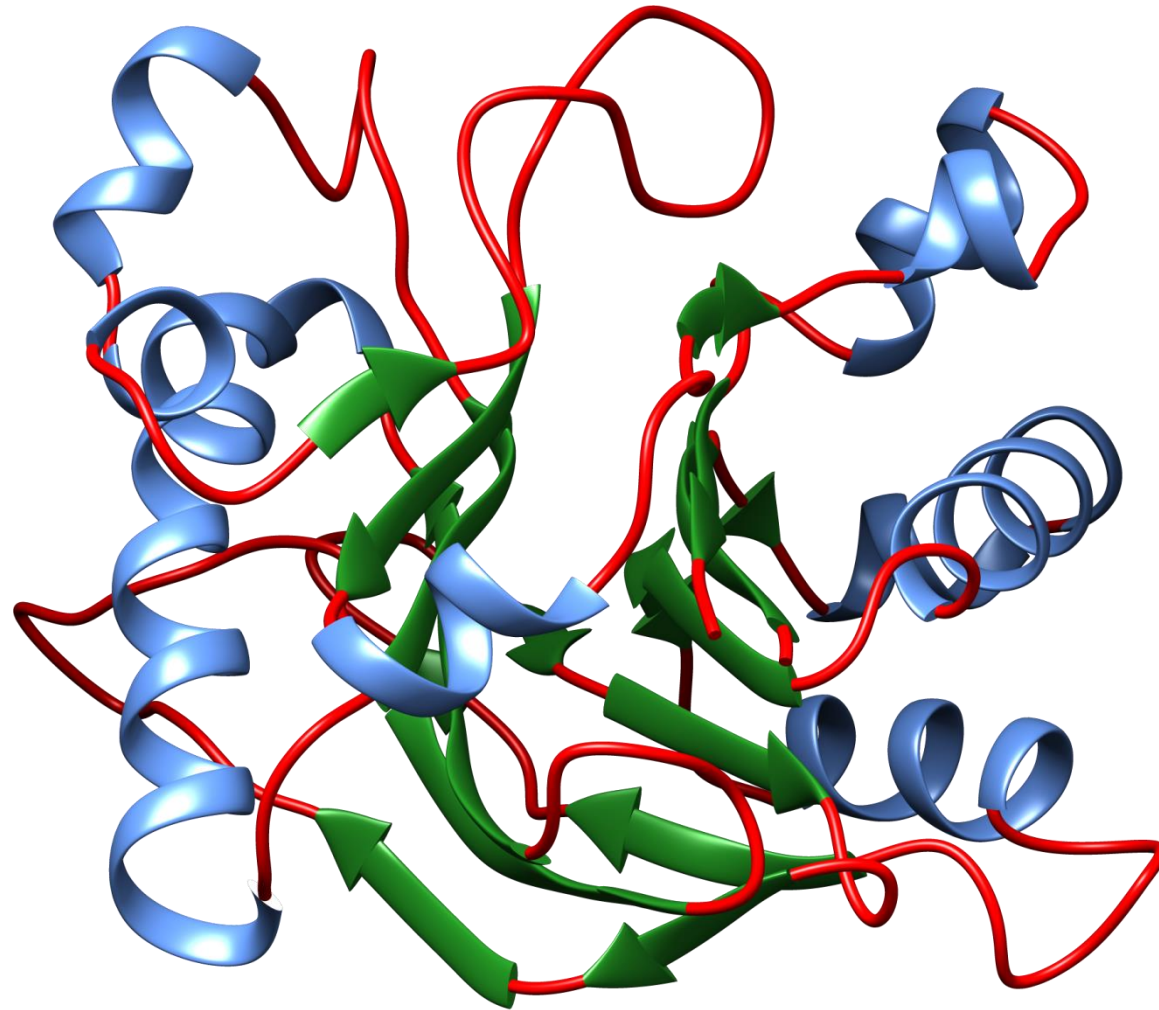
Предсказание трехмерной архитектуры хроматина



Признаки:

- ☐ Геномное расстояние
- ☐ ChIP-seq данные белка CTCF
- ☐ Ориентация CTCF
- ☐ Другие ChIP-seq данные

Предсказание вторичной структуры белка



α -спираль



β -слой



петли

Предсказание вторичной структуры белка

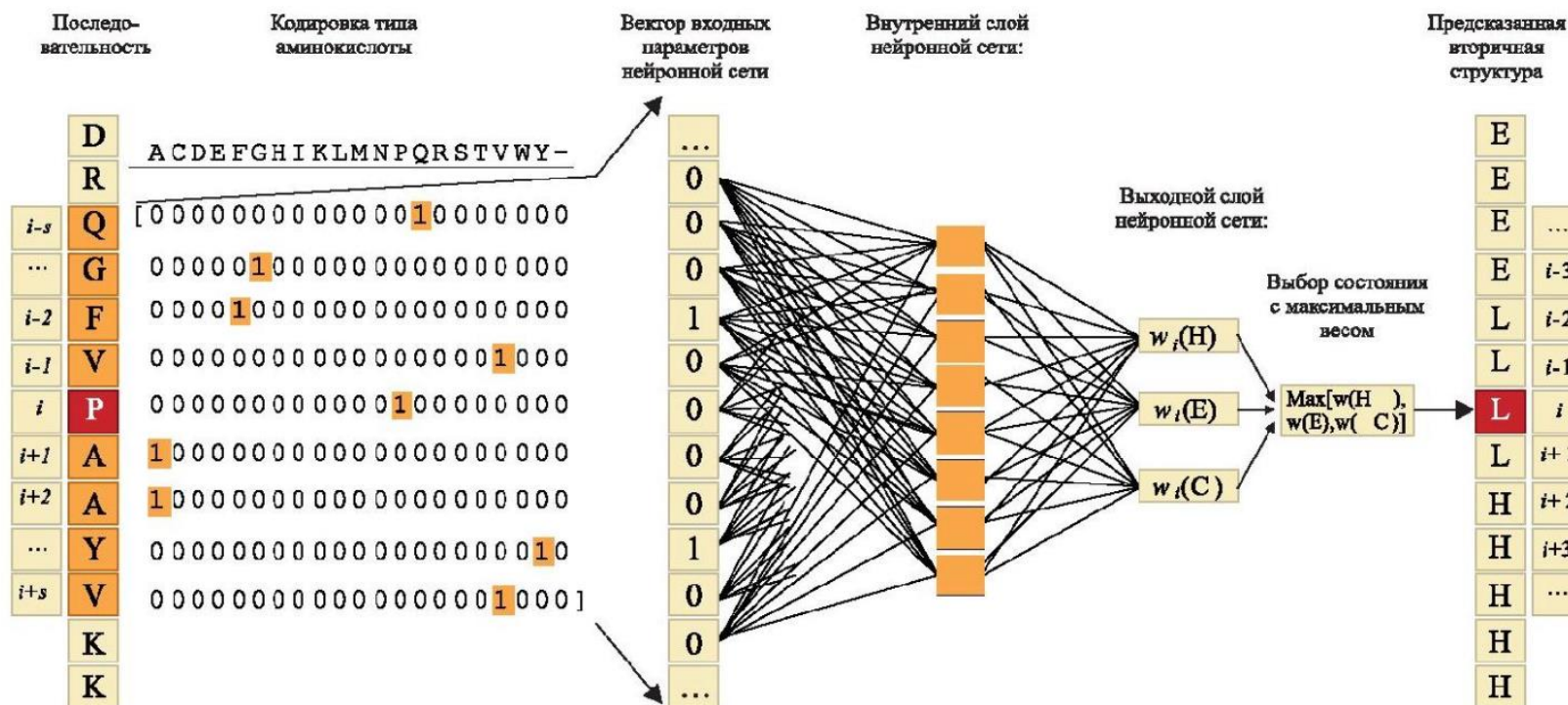


Рис. 2.33. Схема строения нейронной сети для предсказания вторичной структуры (Quian, Sejnowski, 1988)

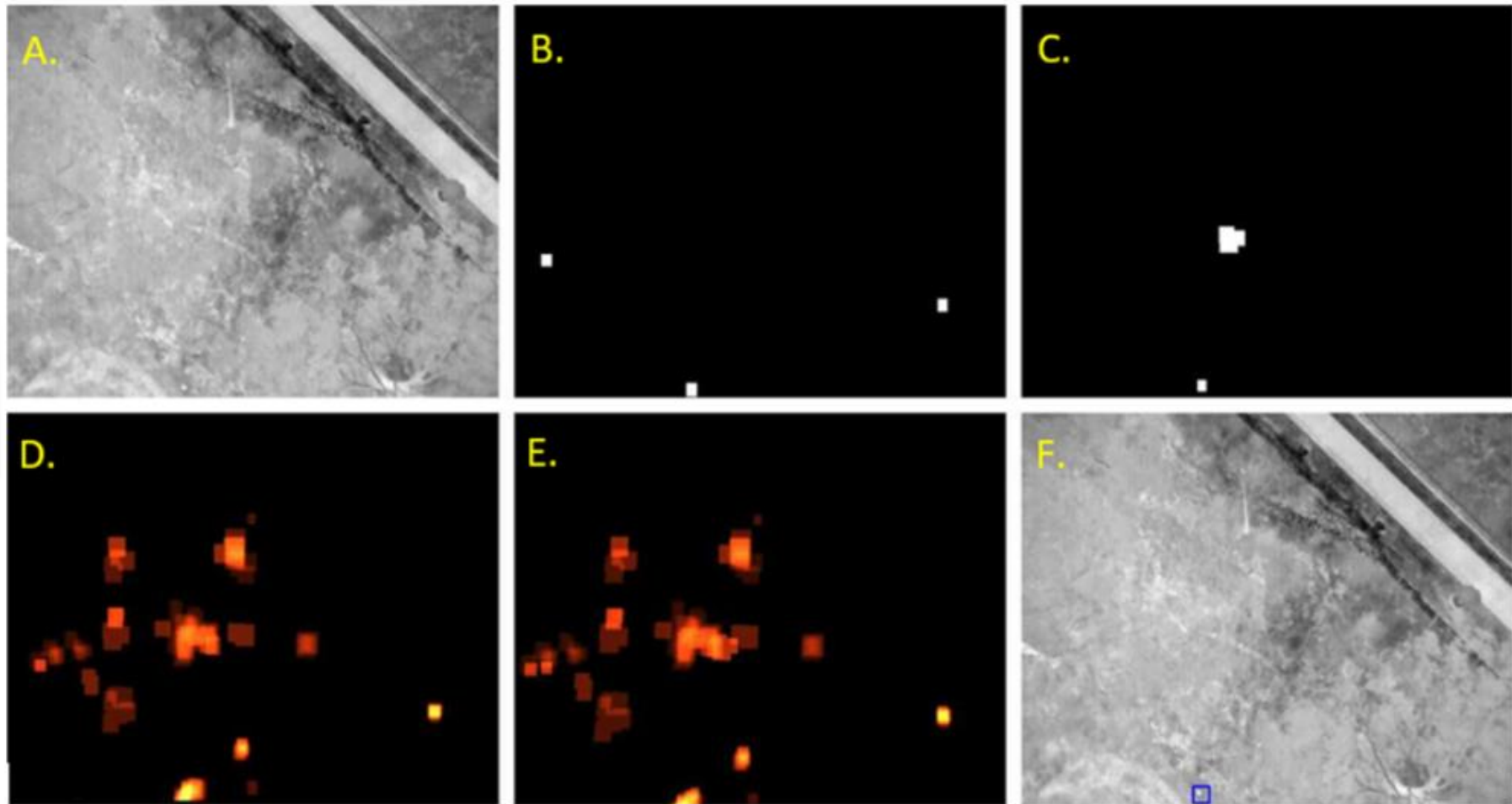
Обнаружение коал

Обычный человек 23 %
Эксперт 70 %



Обнаружение коал

918 images
732 annotated
koalas.



Example detection maps for a single image, and the combined results. (A) Input image, (B) Heat map for the F-RCNN, (C) Heat map for YOLO detectors, (D) the previous accumulator image, (E) updated accumulator incorporating the new heat map, and (F) the detected koala in the original image.

Обнаружение коал

$$P_d = \frac{K_{detected}}{K_{obs}}$$

$$PR = \frac{K_{detected}}{(K_{obs} + K_{false})}$$



	Pd	PR
автоматически	85%	60%
вручную	52%	34%



Спасибо за
внимание!