



Биометрическая идентификация личности по ЭКГ

БИОНАСК 2018



Участники



Глухов Эрнест | СПбГУ

Программирование, Биоинформатика
Python, R, Pipeline Pilot



Миночкина Александра | МФТИ

Математика, Программирование, Биоинформатика
Python



Мирошникова Анастасия |
Новосибирский государственный
университет

Биология, Биоинформатика, Химия
R, Python, bash

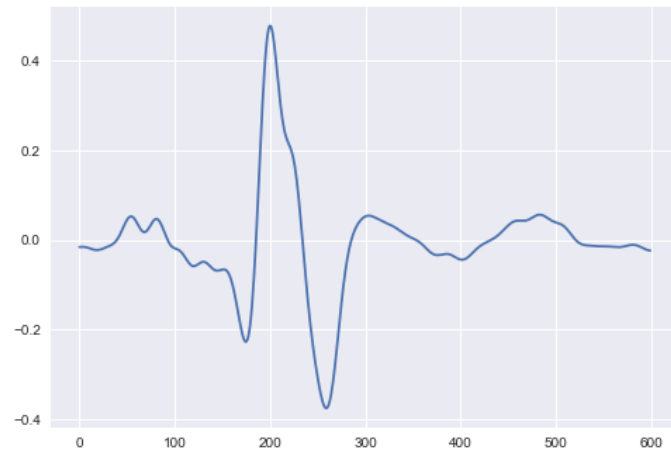


Богданов Марат
Робертович

к.б.н., доцент кафедры
вычислительной математики и
кибернетики УГАТУ, Уфа,
доцент кафедры прикладной
информатики БГПУ им.М.
Акмуллы, Уфа.

Цель проекта

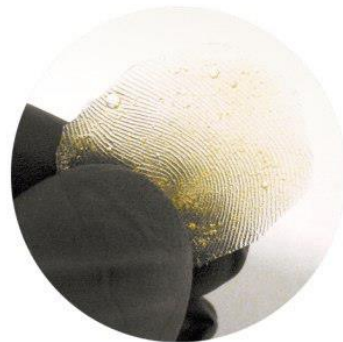
1. Определение наиболее информативных признаков ЭКГ для биометрической идентификации.
2. Выбор таких методов МО, которые обеспечивают максимальную точность.



Актуальность

Существующие методы биометрической идентификации личности (распознавание отпечатков пальцев, лица, голоса, сетчатки глаза) не устойчивы к фальсификации.

При этом потребность в достоверном методе определения индивида увеличивается.



Гипотеза

Гипотеза: человека можно точно идентифицировать по ЭКГ.

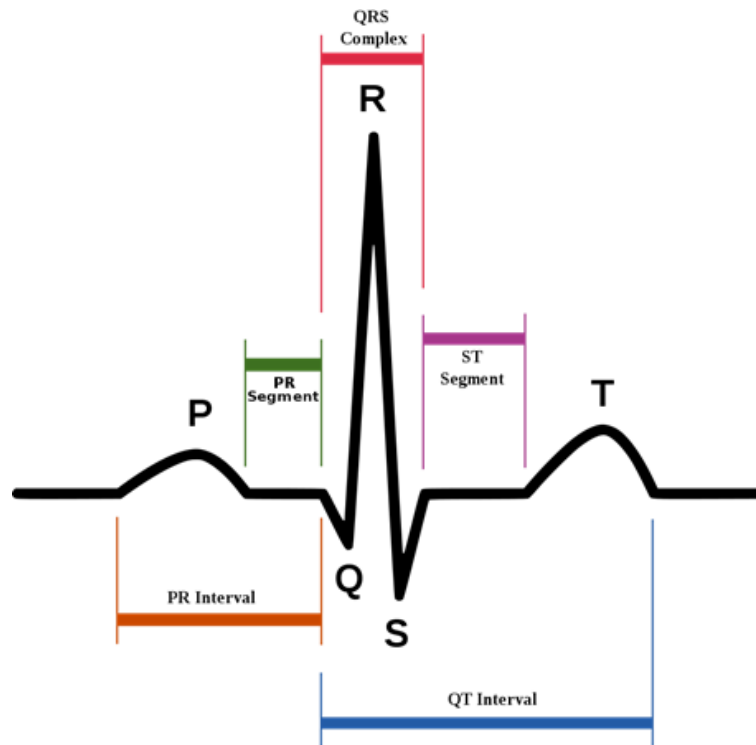
В терминах машинного обучения: задача классификации индивидов по ЭКГ решается с высокой точностью.

Обзор существующих практик

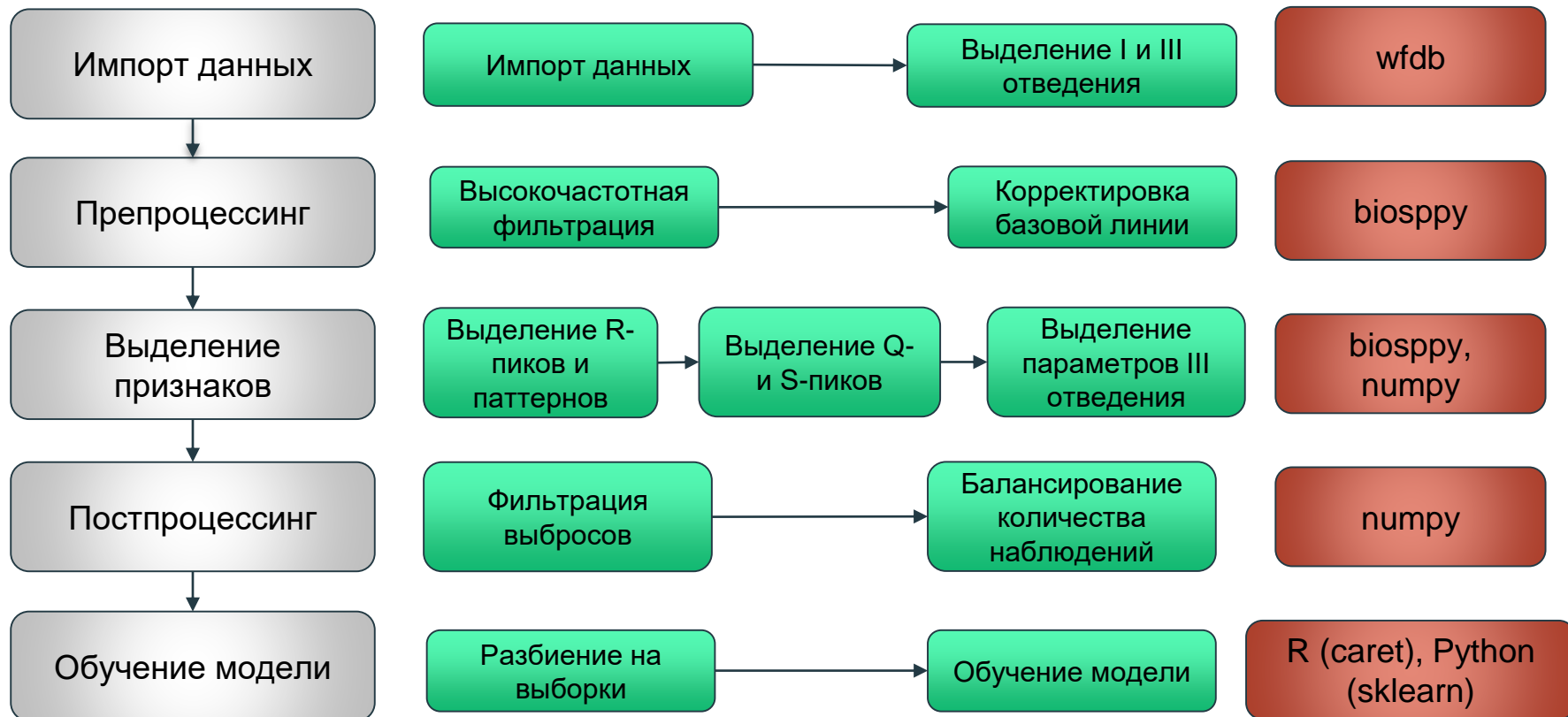
1. Анализ характерных точек ЭКГ:

- временные
- амплитудные
- морфологические

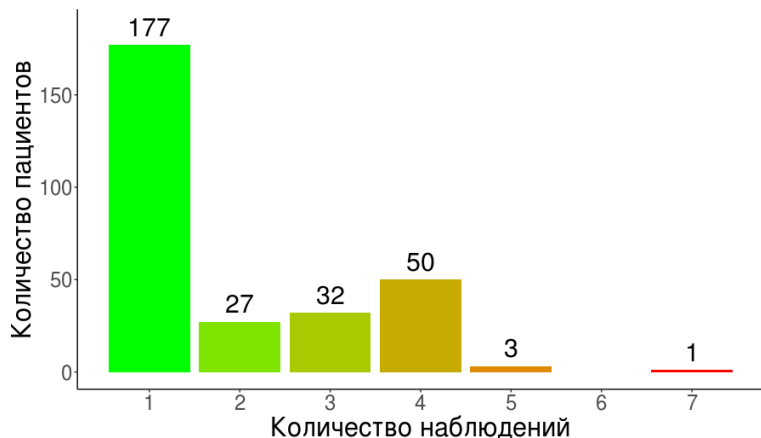
1. Анализ фазовых и частотных свойств ЭКГ



Этапы выполнения



Описание данных



Проект: PhysioNet

База: Physikalisch-Technische Bundesanstalt (PTB)

Образцов ЭКГ: 549

Испытуемых: 290

Возраст: от 17 до 87 лет

Отведений: 16

Частота дискретизации: 1кГц

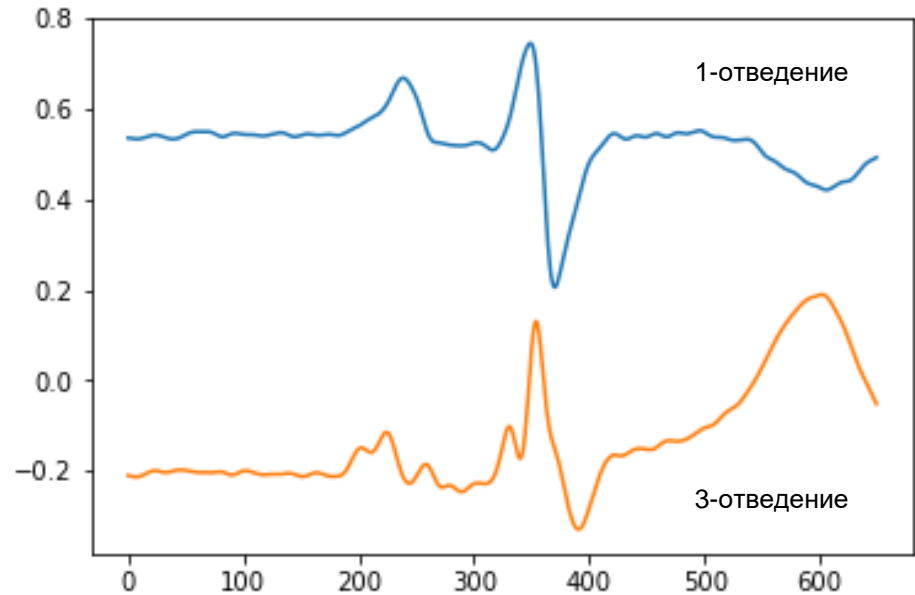
Разрядность: 16 бит

Продолжительность измерения: 180 сек.

Импорт данных

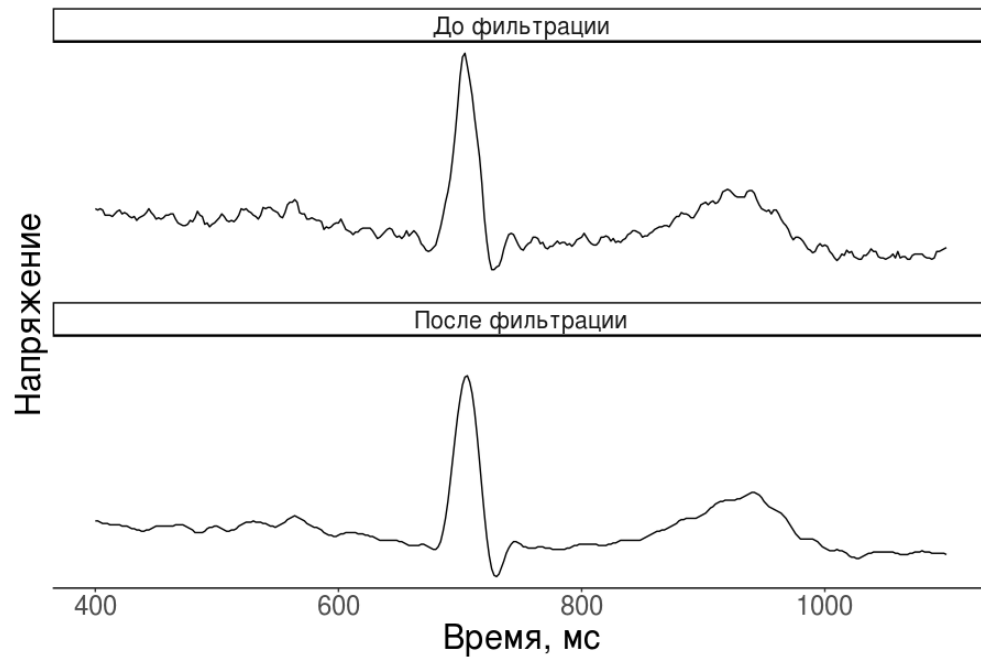
1. Выделение первого и третьего отведения из кардиограммы

Взаимное расположение первого и третьего отведения характеризуют индивидуальные особенности работы сердца индивида



Препроцессинг данных

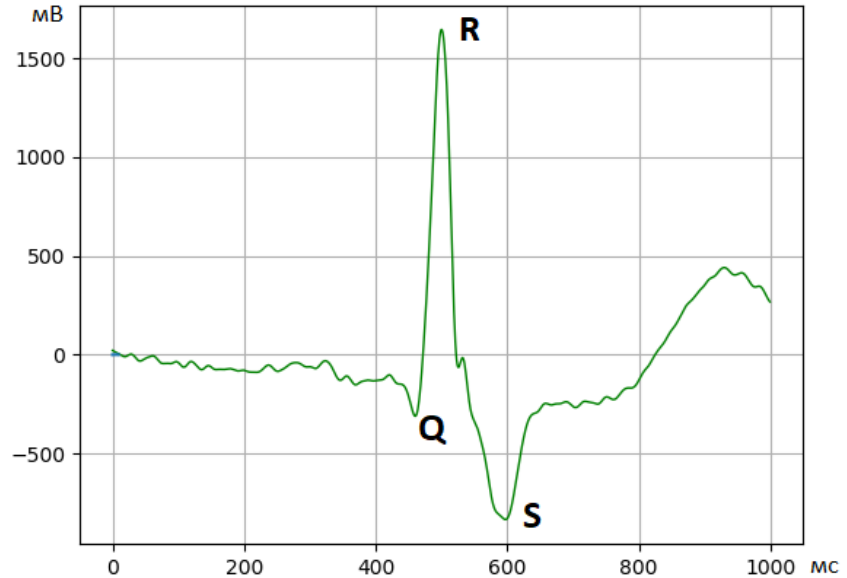
1. Высокочастотная фильтрация
2. Коррекция базовой линии



Выделение признаков

1. Обработка полной кардиограммы:

- определение R-пиков на диаграмме
- определение длины кардиоцикла

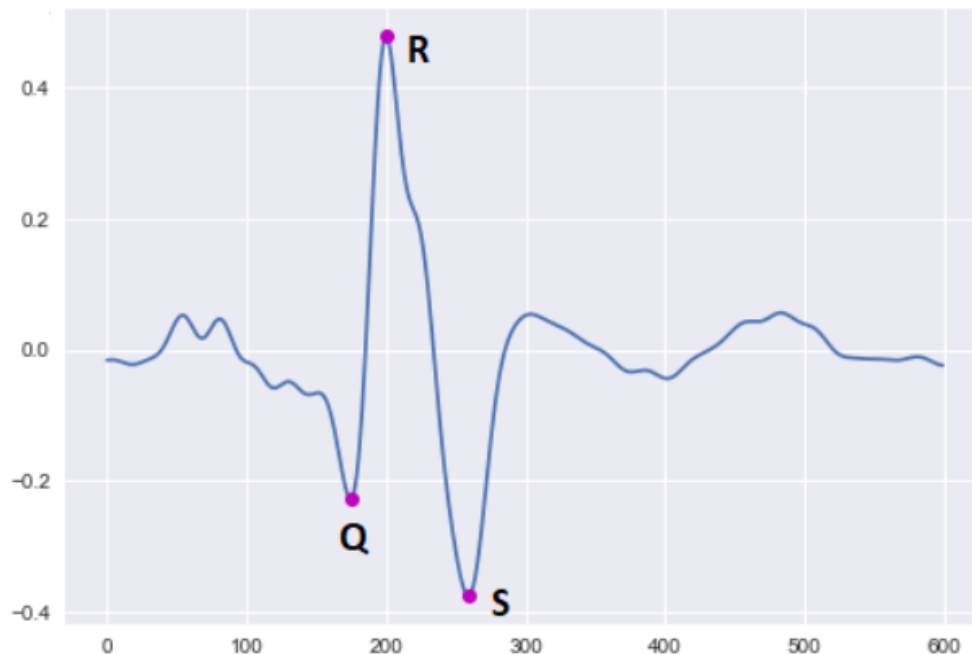


Выделение признаков

2. Создание паттернов:

- выделение кардиоциклов
- определение R-пиков

...

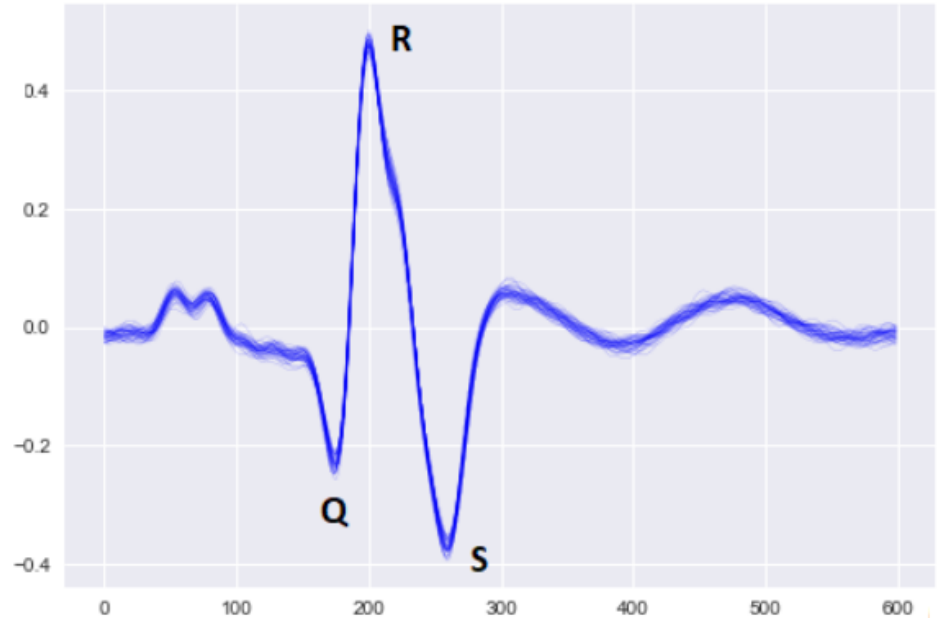


Выделение признаков

2. Создание паттернов:

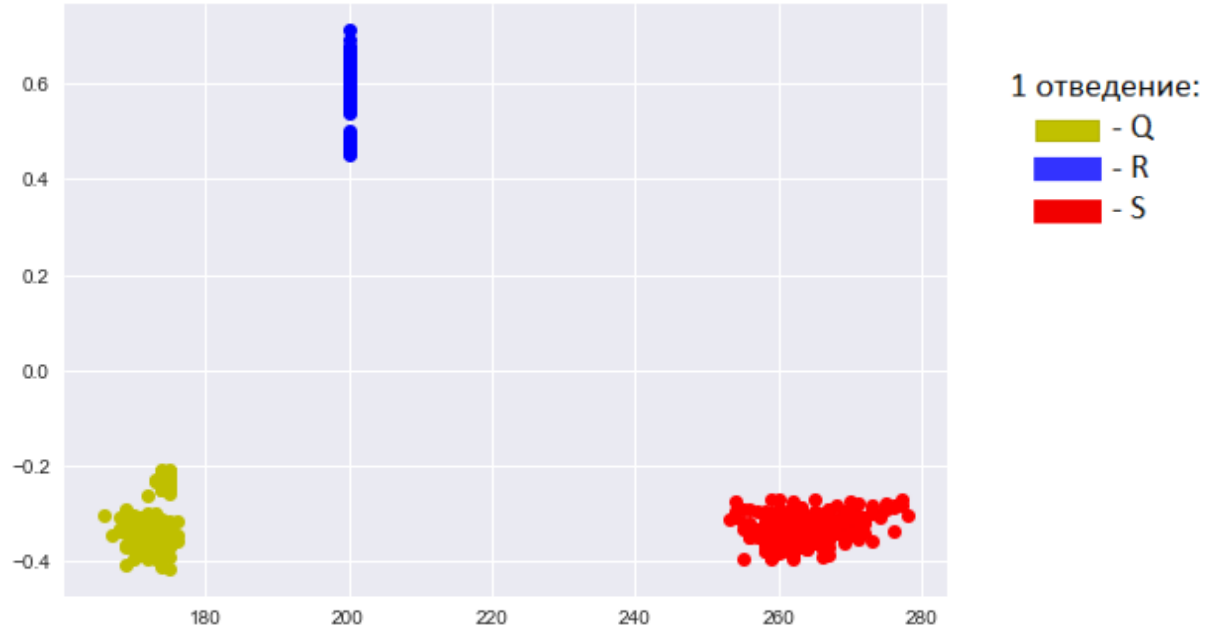
...

- синхронизация по времени R-пиков
- синхронизация по амплитуде R-пиков
- определение координат пиков Q, R, S



Выделение признаков

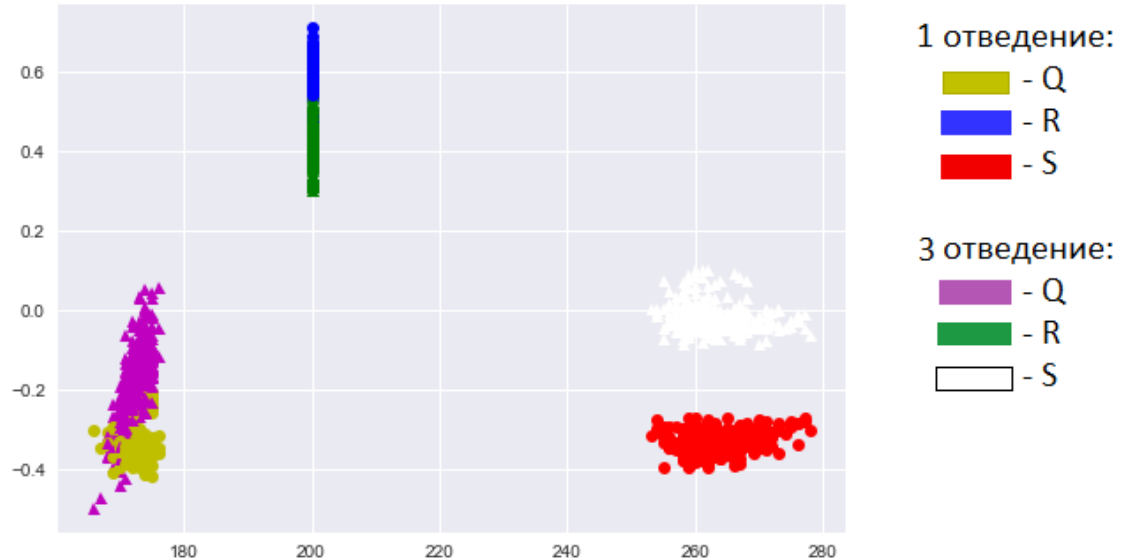
3. Выделение кластеров точек, соответствующих QRS пику первого отведения ЭКГ



Выделение признаков

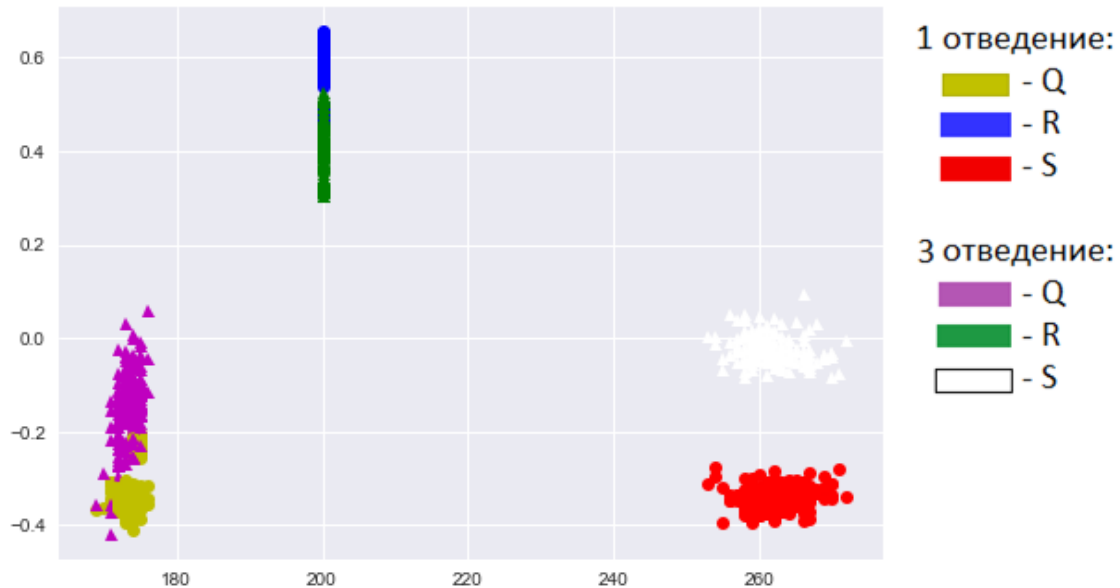
4. Выделение кластеров точек, соответствующих QRS-областям третьего отведения.

Кластеры точек 1 и 3 отведения синхронизированы по времени.



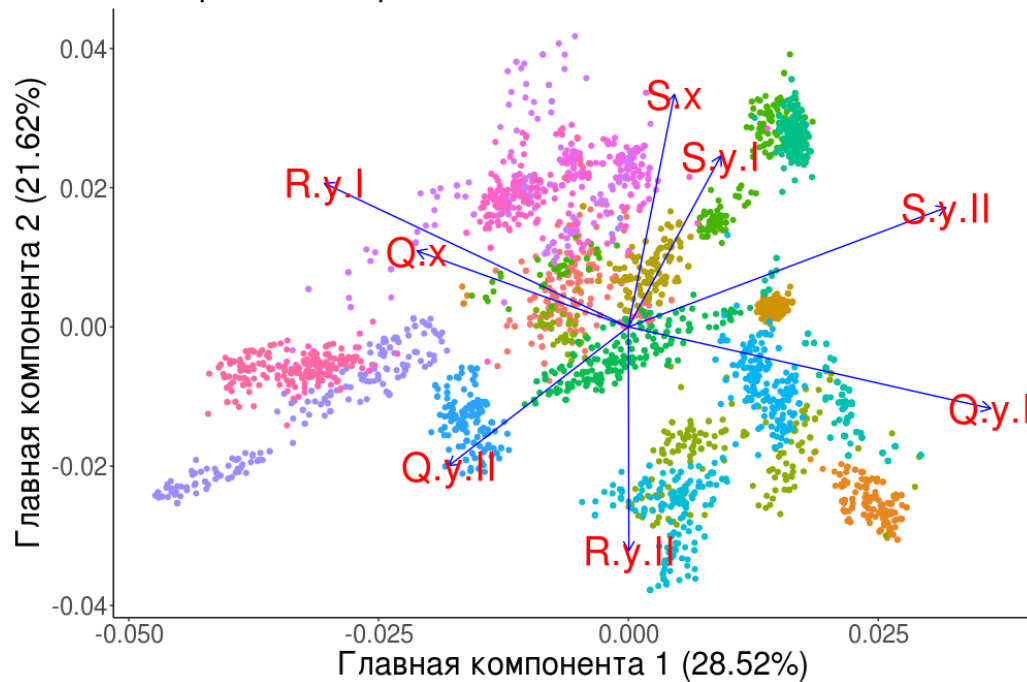
Постпроцессинг

1. Фильтрация выбросов
2. Балансирование количества наблюдений



Финальные выборки

Выборка 3 в координатах главных компонент



Конечные выборки для обучения:

- 1) 10 индивидов, 8 признаков, несбалансированные классы
- 2) 18 индивидов, 6 признаков, сбалансированные классы
- 3) 18 индивидов, 9 признаков, сбалансированные классы
- 4) 268 индивидов, 9 признаков, сбалансированные классы

Модели. Линейный ДА

Линейный дискриминантный анализ

Модели были кросс-валидированы:

- данные разбиты на 5 подвыборок
- воспроизведены в 10 повторях.

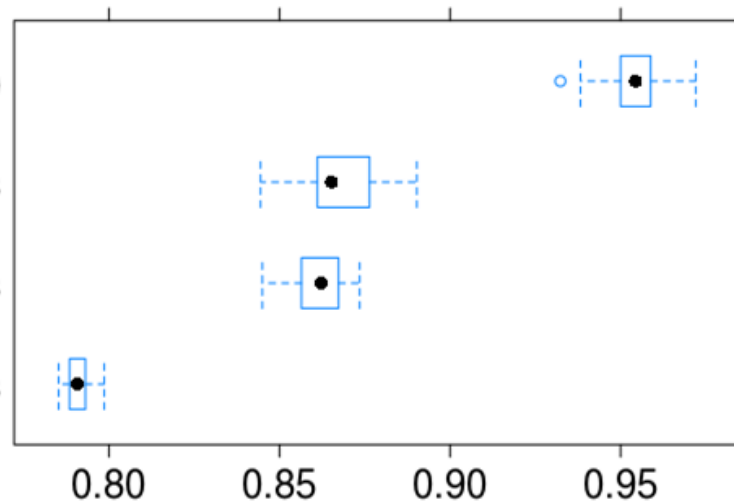
На основе кросс-валидации построены доверительные интервалы для точности.

18 индивидов (9)

18 индивидов

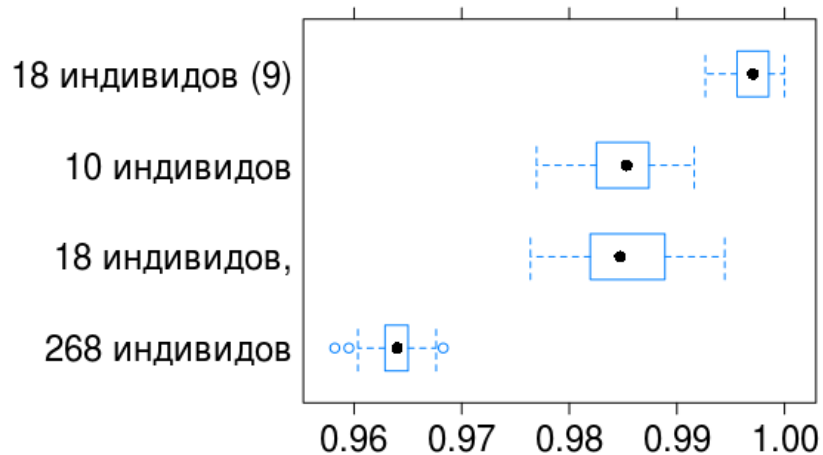
10 индивидов

268 индивидов

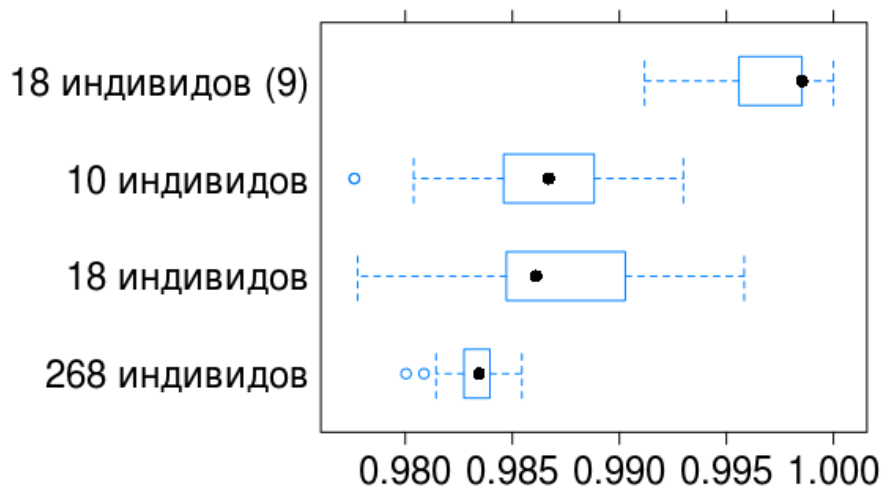


Модели. Наивный Байес и квадратичный ДА

Наивный Байесовский классификатор

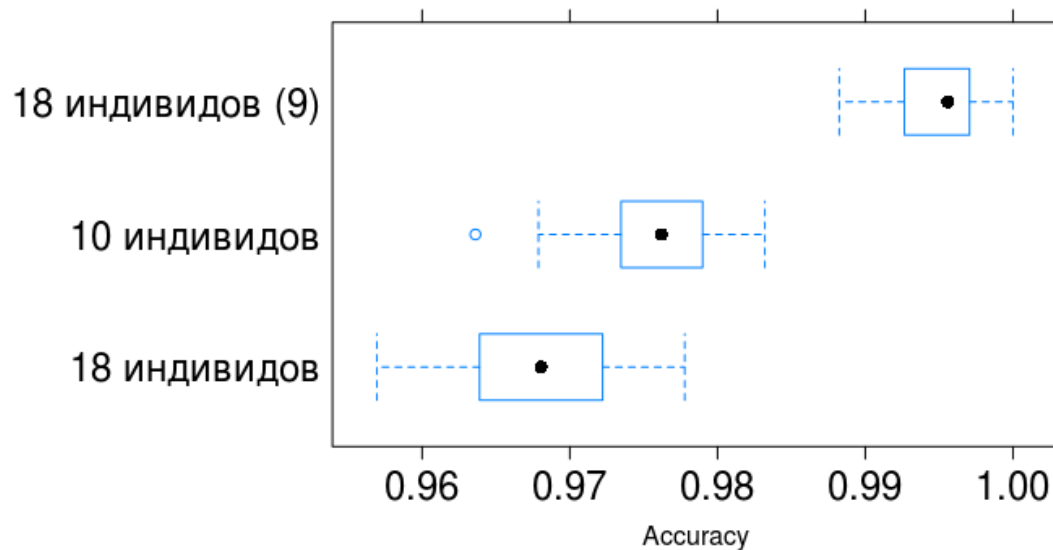


Квадратичный дискриминантный анализ



Модели. Опорные вектора

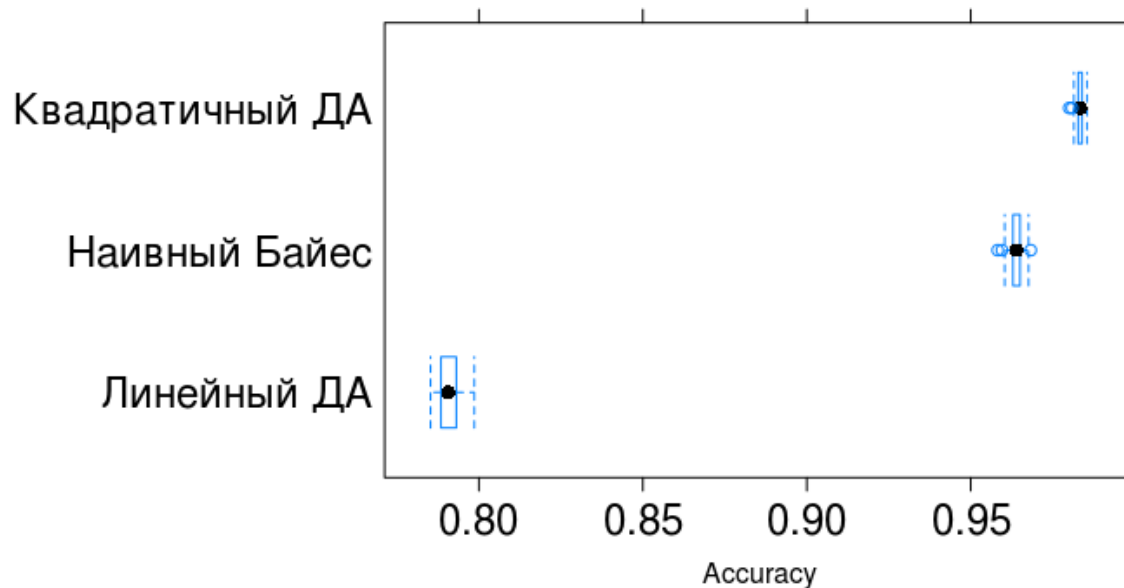
Метод опорных векторов



Показывающие
обнадеживающие
результаты на малых
выборках.

Модели. Сравнение

Качество модели для полной выборки



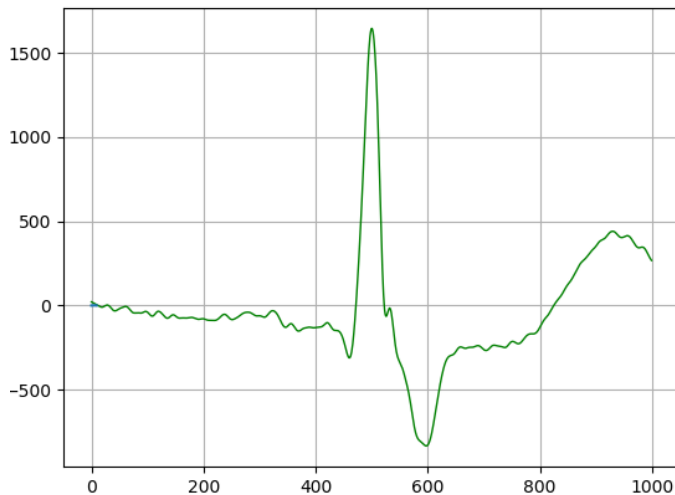
Лучшие модели:

- квадратичный дискриминантный анализ
- наивный Байесовский классификатор.

Модели. LSTM RNN

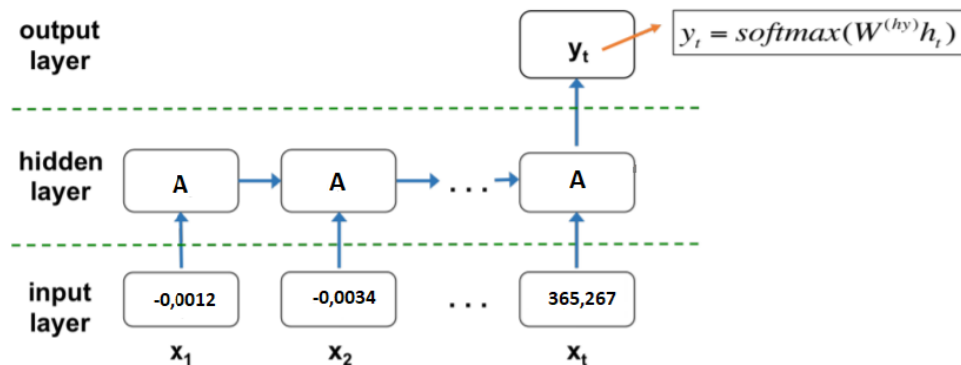
Преимущества:

- не нужно выделять признаки
- используется информация о взаимосвязи последовательных значений ЭКГ



Модели. LSTM RNN

x, MC	z, MB
1	-0,0012
2	-0,0034
...	...
1000	365,267



Accuracy: 0,723

Выводы

- Выделены воспроизводимые характеристики ЭКГ, устойчивые в течение жизни и позволяющие идентифицировать индивида.
- Построены модели, решающие задачу классификации с высокой точностью.

Биометрическая идентификация личности по ЭКГ

БИОНАСК 2018