

Модель машинного обучения для прогнозирования вентиляторных порогов

Александр Минкин

Институт прикладной математики им. М.В. Келдыша РАН

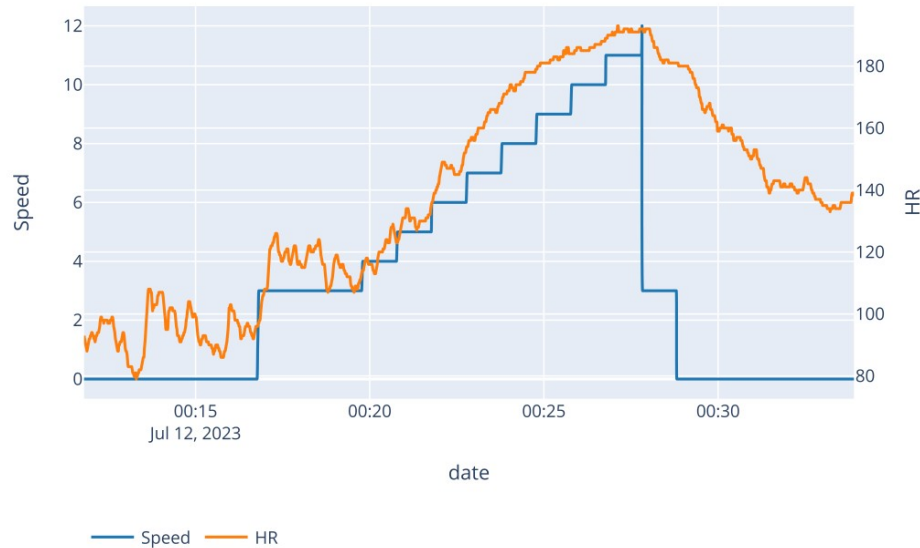


- Длительное пребывание в космосе приводит к снижению физической работоспособности космонавтов.
- Одной из важнейших задач для поддержания здоровья человека в космосе является внедрение индивидуального плана кардиотренировок в зависимости от его физической подготовки.
- Для этого, в частности, необходимо оценить индивидуальные аэробные и анаэробные пороги, которые, в целом, эквивалентны вентиляторным порогам, определяемым на основе данных газообмена.

Кардиореспираторное нагрузочное тестирование (СРЕТ)

- Традиционный способ оценки порогов аэробного и анаэробного состоит в проведении кардиореспираторного нагрузочного тестирования с газоанализатором. Такое исследование было сделано с участниками изоляционного эксперимента «Сириус-23» (<http://sirius.imbp.ru/>) на беговой дорожке. Тест был проведён со ступенчато-возрастающей физической нагрузкой до предела её переносимости.

Speed, HR by date



- «COSMED Qark CPET» gas analyzer
- «H/p/cosmos» treadmill



Photo: ГНЦ РФ – ИМБП РАН

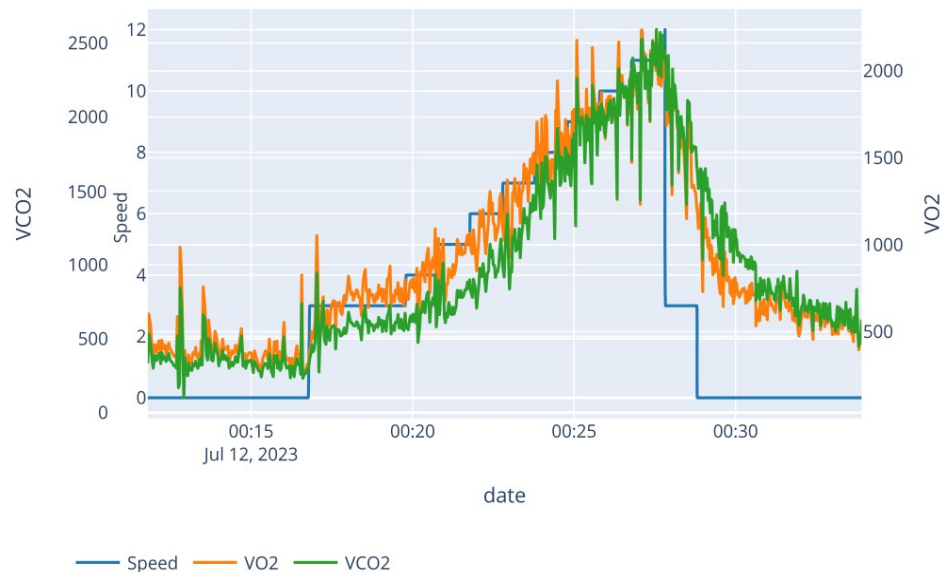
Кардиореспираторное нагрузочное тестирование (СРЕТ)

- Данные о газообмене, полученные с эргоспирометра «COSMED Qark СРЕТ», содержат ~80 характеристик таких, как поглощение O_2 , выделение CO_2 , минутная вентиляция, частота дыхания и т. д. Данные о газообмене содержат помехи и шумы!

	date	Y	Rf	VT	VE	IV
99	2023-07-12 00:17:55	0	18.1150	0.910000	16.470500	798.0
100	2023-07-12 00:17:56	0	18.5525	0.902000	16.722750	774.0
101	2023-07-12 00:17:57	0	18.9900	0.894000	16.975000	750.0
102	2023-07-12 00:17:58	0	18.7200	0.946333	17.686333	812.0
103	2023-07-12 00:17:59	0	18.4500	0.998667	18.397667	874.0
...	...	...	...	...	...	...
1503	2023-07-12 00:41:19	2	37.5000	0.408000	15.300000	1271.0
1504	2023-07-12 00:41:20	2	32.9700	1.317000	39.299500	1235.0
1505	2023-07-12 00:41:21	2	28.4400	2.226000	63.299000	1199.0
1506	2023-07-12 00:41:22	2	31.1700	1.960500	60.378500	1351.0
1507	2023-07-12 00:41:23	2	33.9000	1.695000	57.458000	1503.0
[1409 rows x 98 columns]						

IV – inspiratory volume, VE – minute ventilation, Rf – respiratory frequency, VT – tidal volume, VO₂ – oxygen uptake, VCO₂ – CO₂ production, Y – **expert** level labels

Speed, VO₂, VCO₂ by date



Визуальная оценка ВП

VT1 описывает переход от аэробного метаболизма к анаэробному при незначительном накоплении лактата. VT2 определяет непропорциональное увеличение вентиляции при резком накоплении лактата. Их прогнозирование обычно выполняется путем визуального изучения графических вентиляционных эквивалентов или методом V-образного наклона (V-slope) после предварительной обработки данных.

V-slope method

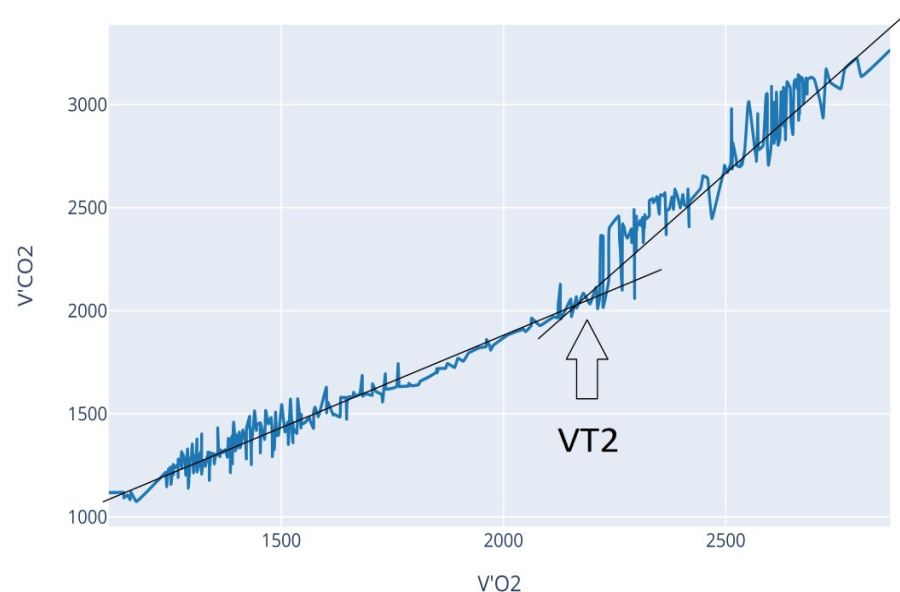
VT1 графически определяется следующим образом:

- Первое непропорциональное увеличение $\dot{V}E$;
- Увеличение $\dot{V}E/\dot{V}O_2$ без увеличения $\dot{V}E/\dot{V}CO_2$;
- Увеличение $P_{et}O_2$ (end-tidal O_2) без последующего снижения $P_{et}CO_2$.

VT2 графически определяется следующим образом:

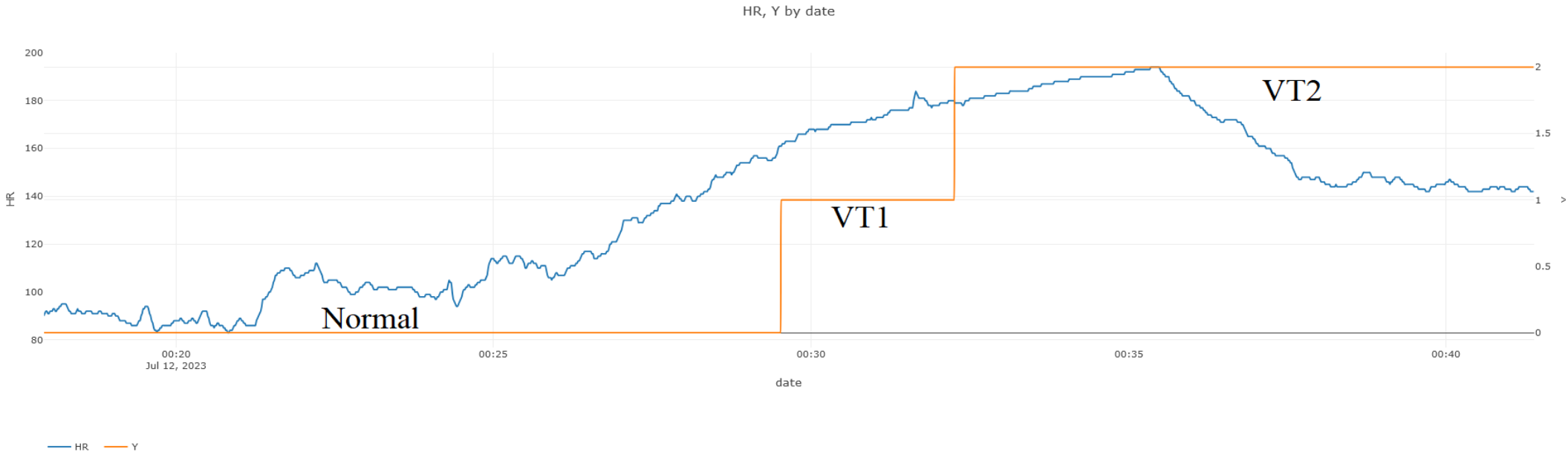
- Второе непропорциональное увеличение $\dot{V}E$;
- Первое систематическое снижение $P_{et}CO_2$ (end-tidal CO_2);
- Первое систематическое увеличение $\dot{V}E/\dot{V}CO_2$.

$\dot{V}CO_2$ by $\dot{V}O_2$



Модели машинного обучения

Поскольку данные газообмена содержат довольно много помех, оценка ВП требует значительных усилий со стороны эксперта. Эта работа может быть оптимизирована и упрощена с использованием ML.

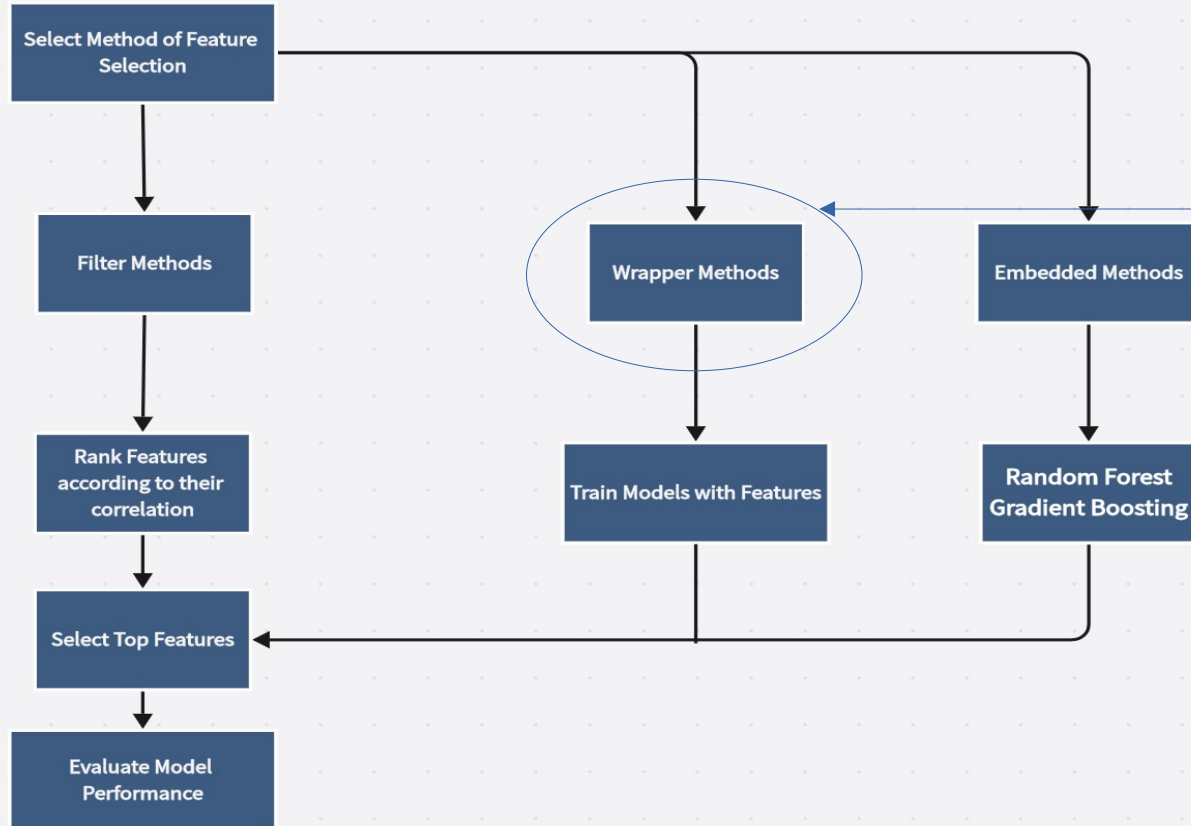


Данные СРЕТ были размечены экспертом на основе оценки результатов газоанализа. Состояние до достижения VT1 отмечено “0”, между VT1-VT2 - “1” и после превышения VT2 - “2”.

Общая схема обработки данных



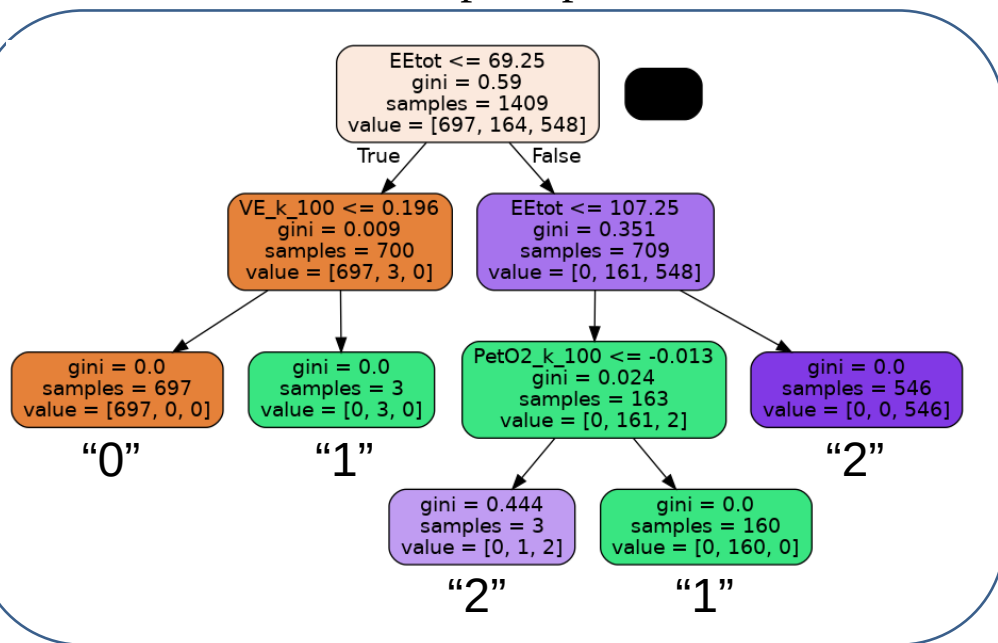
Методы отбора признаков



Отбор признаков делается на основе сопоставления их значимости с точностью классификации по ML модели

Деревья решений

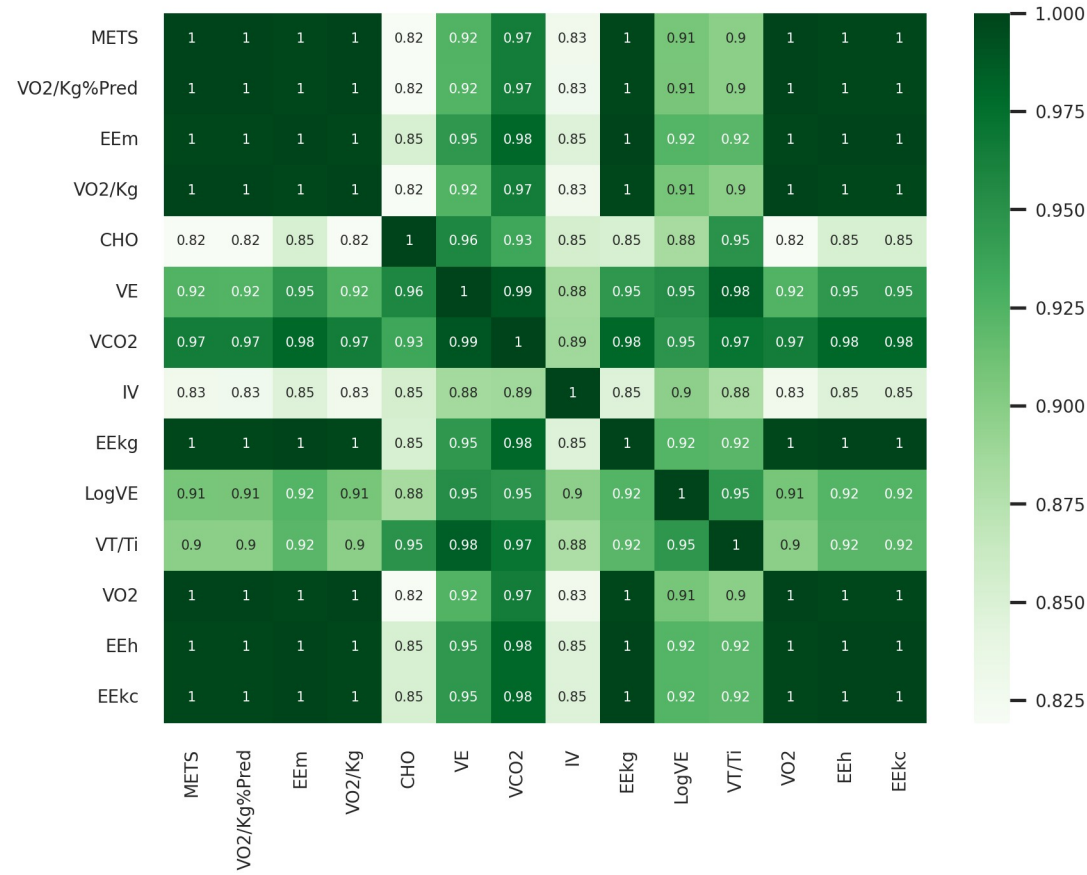
Пример:



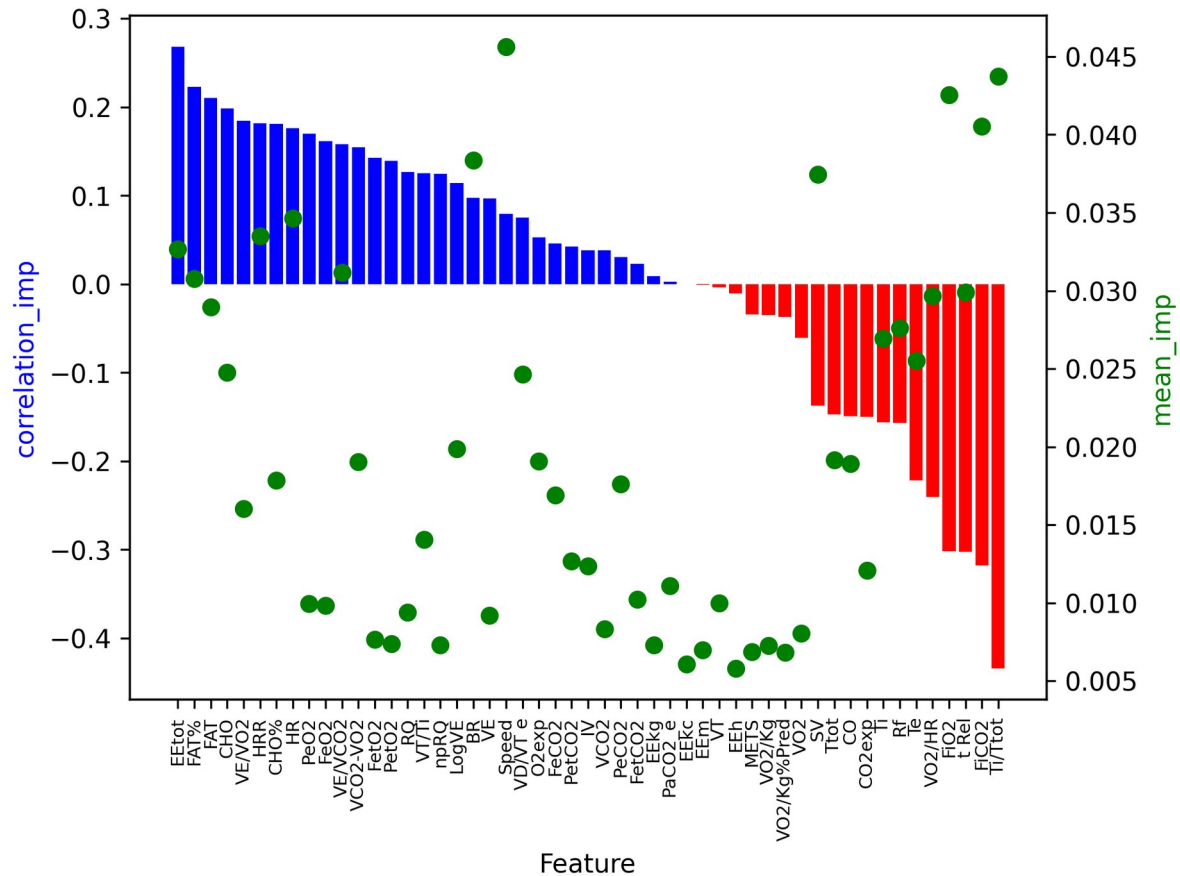
Преимущества:

- Возможность работы с коррелированными признаками.
- Быстрая обучаемость.
- Внутренняя объяснимость.
- Возможность представление обученной модели в виде каскада условных инструкций.

Корреляции между признаками



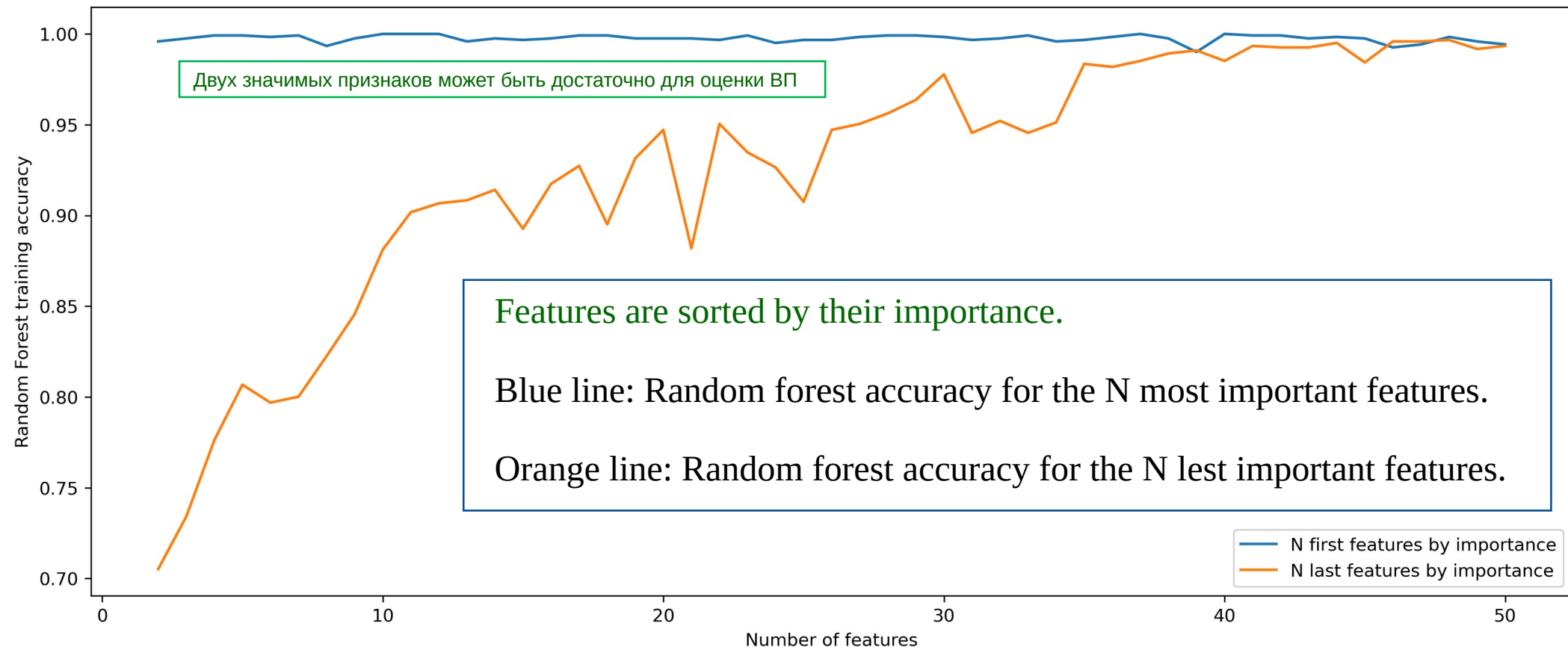
Итерационное исключение признаков



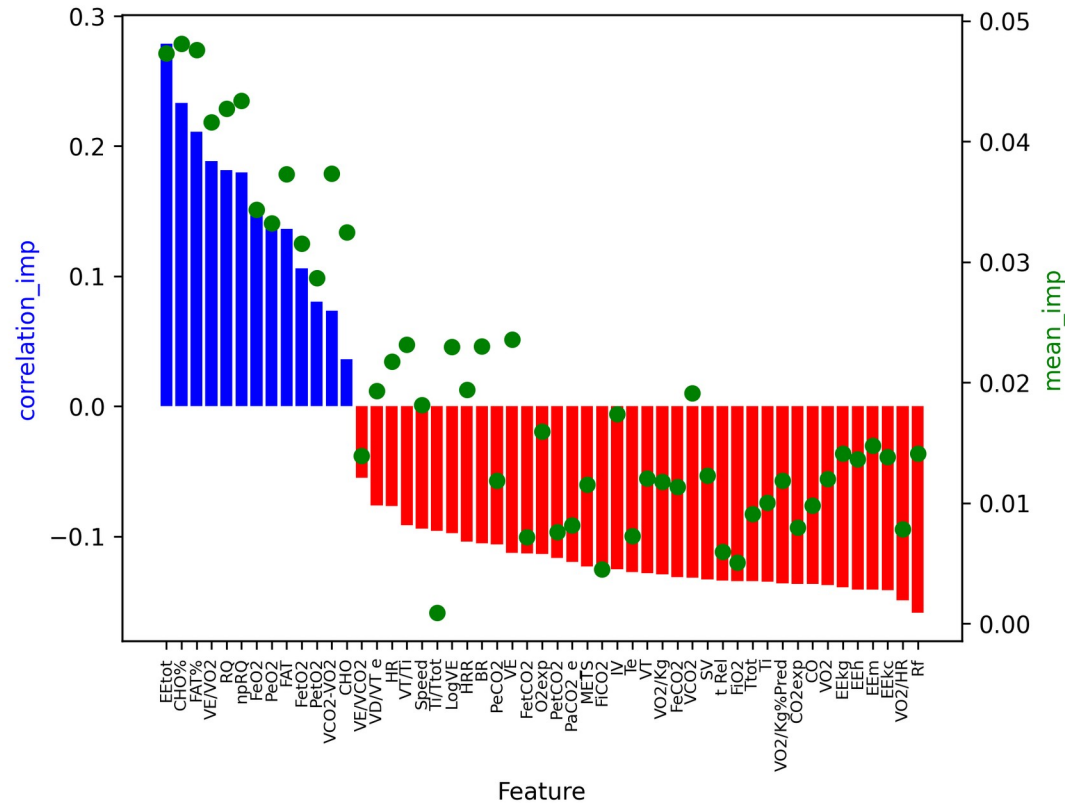
correlation_imp – корреляция между значимостью признаков (снижение неоднородности Джини) и точностью классификации на обучающей выборке.

mean_imp – средняя значимость признаков.

Точность классификации на значимых признаках



Итерационное добавление признаков



Micro average F1 score:

	Test set					
Train set	1	2	3	4	5	6
1	1.0	0.834	0.910	0.840	0.892	0.908
2	0.556	0.999	0.551	0.913	0.551	0.657
3	0.889	0.754	1.0	0.741	0.935	0.841
4	0.488	0.914	0.531	1.0	0.545	0.576
5	0.899	0.736	0.968	0.755	1.0	0.892
6	0.911	0.789	0.920	0.811	0.904	0.989

Выбранные признаки в порядке убывания значимости:
 [Eetot, CHO%, FAT%, VE/VO2, RQ, FeO2, PeO2, FetO2, PetO2, VCO2-VO2]

Выводы

- 1) Представленный классификатор на основе деревьев решений – это простая и объяснимая модель, которая быстро обучается, позволяя работать с коррелированными признаками. Это дает возможность использования деревьев решений для отбора значимых признаков. Такой отбор может проводиться путем последовательного исключения или добавления признаков.
- 2) Наиболее значимыми для классификации ВП признаками являются показатели непрямой калориметрии (Eetot, FAT, CHO) и дыхательный коэффициент (RQ).
- 3) Информация о важности признаков (снижение неоднородности Джини) может быть использована для обучения более сложных моделей.
- 4) Модель случайного леса, обученная на значимых признаках, позволяет автоматизировать прогноз ВП, что существенно снижает трудозатраты эксперта в процессе планирования индивидуальных кардиотренировок.

1) Автор выражает благодарность сотрудникам ИМБП РАН Елене Фоминой и Марие Кокуевой за помощь в подготовке данной работы.

Спасибо за внимание!