Д.И. Осолодкин: В начальной части работы количественно не охарактеризованы обучающие выборки (слайды 5-6). Не могли бы вы представить данные о количестве положительных и отрицательных примеров, а также о химическом разнообразии соединений выборок?
В.С. Сухачёв: Размер выборки для машинного обучения составил 501 соединение.
Из 501 соединения 360 неаритмогенны (-), а 141 аритмогенны(+).
Для иллюстрации химического разнообразия, было подсчитано количество терминов АТС отдельно для аритмогенных и неаритмогенных соединений. На 141 аритмогенное соединение приходится 76 терминов АТС 4-го уровня, которые отражают показания к применению лекарства, и, часто, химический класс соединений.
На 360 неаритмогенных соединений приходится 182 термина АТС 4-го уровня.

Д.И. Осолодкин: При комбинировании дескрипторов возникает проблема их распределения: если бинарные дескрипторы могут принимать значения либо 0, либо 1, то дескрипторы PASS Targets варьируют от -1 до 1 с максимумом распределения где-то посередине. Учитываются ли эти дескрипторы как равноценные и обрабатываются ли они моделью одинаково? Не является ли причиной низкого качества прогноза смешивание дескрипторов различных типов?
В.С. Сухачёв: Что касается PASS и бинарных дескрипторов, то они комбинировались следующим образом. Если из эксперимента известно, что соединение действует на мишень, то мы ставили 1 (бинарные дескрипторы). Если не известно (чаще всего просто не тестировали), то мы заменяем «0» на результат прогноза PASS, но только в том случае, если Pa-Pi>0. В противном случае оставляем 0. В таблице не было отрицательных значений, кроме того Random Forest как правило хорошо работает с гетерогенными дескрипторами. Однако, чтобы исключить возможность влияния распределения дескрипторов на точность прогноза, мы в ближайшее время попробуем использовать другие способы их комбинации, например, консенсус прогноза двух моделей, построенных на двух разных типах дескрипторов.

P09917	P34913	P20292	Q15722	Q9NS75	P09960	P41145	P41143	P35372	P03372	Q92731	Q02153	Q9NP56
1.0000	1.0000	1.0000	1.0000	1	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0	0.0000
0.0000	0.0000	0.0000	0.0000	0	0.0000	1.0000	1.0000	1.0000	0.0820	0.0913	0	0.0000
0.0000	0.0000	0.0000	0.0000	0	0.0000	0.0000	0.1769	0.0000	1.0000	1.0000	0	0.0000
0.0000	0.0000	0.0000	0.0000	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1	1.0000
0.0000	0.0000	0.0000	0.0000	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0	0.0000
0.0000	0.0000	0.0000	0.0000	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0	0.0000
0.0000	0.0000	0.0000	0.0000	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0	0.0000
0.0000	0.0000	0.0000	0.0391	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0	0.0000
0.0000	0.0000	0.0000	0.0000	0	0.0000	0.0675	0.0000	0.0039	0.0427	0.1039	0	0.0000
0.0483	0.0000	0.0055	0.0026	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0	0.0372
0.0000	0.0000	0.0000	0.0000	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0	0.0000
0.0000	0.0000	0.0000	0.0000	0	0.0000	0.1156	0.0000	0.0651	0.0000	0.0000	0	0.0000
0.0000	0.0000	0.0000	0.0000	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0	0.0000
0.0000	0.0000	0.0000	0.0000	0	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000	0	0.0000
0.0000	0.0000	0.0000	0.0000	0	0.0000	0.2568	0.0000	0.0868	0.0000	0.0358	0	0.1020
0.0000	0.0000	0.0000	0.0000	0	0.0000	0.0959	0.0000	0.0520	0.0000	0.0000	0	0.0000
0.0000	0.0000	0.0000	0.0000	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0	0.0661