

**П.Г. Полищук:** Какой размер обучающей выборки? И какое соотношение активных и неактивных соединений?

**В.С. Сухачев:** Размер выборки для машинного обучения составил 501 соединение. Из 501 соединения 360 неаритмогенны, а 141 аритмогенны.

**П.Г. Полищук:** Сколько бинарных дескрипторов использовалось?

**В.С. Сухачев:** Использовался один вид бинарных дескрипторов. Этот тип дескрипторов представлял собой экспериментальные данные по способности лекарства связываться с белками человека. Он был взят по той причине, что взаимодействия лекарственных соединений с белками человека, включая ионные каналы, являются основной причиной желудочковой тахикардии. Полученная выборка включает 501 соединение и 572 бинарных дескриптора (белковых мишеней).

**П.Г. Полищук:** Как дескрипторы PASS конвертировались в бинарные?

**В.С. Сухачев:** Дескрипторы PASS не конвертировались, а использовались “как есть”, поскольку алгоритм Random Forest не предъявляет строгих требований к однородности типов чисел.

**П.Г. Полищук:** Слайд 14. Почему была выбрана такая функция оценки и что за параметр k? И как выполнялся прогноз в рамках этой сети?

**В.С. Сухачев:** Параметр k это коэффициент, подбираемый посредством многократного скользящего контроля таким образом, чтобы точность прогноза была максимально возможной. Чем меньше этот коэффициент, тем меньше вклад в общую оценку тех мишеней, которые наиболее удалены в сети от ионных каналов, и наоборот.

Данная функция часто используется в публикациях по смежной тематике, как например, в этой работе: Yu, Hasun & Choo, Sungji & Park, Junseok & Jung, Jinmyung & Kang, Yeeok & Lee, Doheon. (2016). Prediction of drugs having opposite effects on disease genes in a directed network. BMC Systems Biology. 10. 10.1186/s12918-015-0243-2.

**П.Г. Полищук:** Более общий вопрос. Как Вы оцениваете полезность модели, включающей экспериментальные дескрипторы как бинарные в вашем случае?

**В.С. Сухачев:** Ниже приведены результаты, полученные тем же методом, но с использованием каждого типа дескрипторов по отдельности.

	AUC	Чувствительность	Специфичность	Сбал. точность
PASS	0.779	0.695	0.691	0.693
BINARY	0.796	0.678	0.803	0.740
Pass+binary	0.795	0.731	0.717	0.724

Как видно из таблицы, бинарные дескрипторы позволяют получить более высокую точность прогноза по сравнению с расчетными дескрипторами. Многие крупные фармацевтические компании тестируют разрабатываемые соединения на большое количество мишеней, не связанных с основным терапевтическим эффектом. Так что экспериментальные данные о взаимодействии с мишенями человека могут быть доступны и на ранних этапах разработки новых лекарств. Если такие данные есть, то можно использовать экспериментальные дескрипторы, если нет, то только расчетные.

К вопросу о методе прогноза с помощью сети: алгоритм измеряет сумму расстояний между каждой парой мишень -- ионный канал для каждого соединения. Полученные оценки суммируются по всем ионным каналам и всем мишеням лекарства, для которого делается прогноз.