

ПРОГНОЗИРОВАНИЕ НЕБЛАГОПРИЯТНЫХ ЭФФЕКТОВ МЕЖЛЕКАРСТВЕННЫХ ВЗАИМОДЕЙСТВИЙ НА СЕРДЕЧНО-СОСУДИСТУЮ СИСТЕМУ НА ОСНОВЕ АНАЛИЗА СВЯЗЕЙ «СТРУКТУРА–АКТИВНОСТЬ»

© 2023 В.С. Сухачёв*, С.М. Иванов, А.В. Дмитриев

Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича,
119121 Москва, Россия; электронная почта: withstanding@yandex.ru

Поступила в редакцию 21.02.2023

После доработки 17.03.2023

Принята к публикации 17.03.2023

Совместное применение лекарственных средств часто приводит к межлекарственным взаимодействиям, которые могут сопровождаться различными нежелательными лекарственными реакциями (НЛР), представляющими угрозу для жизни и здоровья пациентов. Воздействие, обусловленное НЛР на сердечно-сосудистую систему, является одним из наиболее значимых проявлений межлекарственных взаимодействий (МЛВ). Клиническая оценка НЛР, возникающих в результате МЛВ между всеми используемыми в терапевтической практике парами лекарств, невозможна. Цель данной работы заключалась в построении при помощи анализа «структура–активность» (Q)SAR-моделей для прогнозирования НЛР со стороны сердечно-сосудистой системы, опосредованных взаимодействиями между парами лекарств при их совместном приёме. Данные о неблагоприятных эффектах, возникающих в результате МЛВ, были получены из базы данных DrugBank. Необходимые для построения точных (Q)SAR-моделей данные о парах лекарств, которые не вызывают соответствующие эффекты, получены из базы данных TwoSides, содержащей результаты анализа спонтанных сообщений. Для описания пары структур лекарств были использованы два типа дескрипторов: RoSMNA-дескрипторы и вероятностные оценки прогноза биологических активностей, полученные при помощи программы PASS. Связи «структура–активность» были установлены с использованием метода «Random Forest». Точность прогноза была вычислена посредством пятикратного скользящего контроля. Наибольшие значения точности были получены с использованием вероятностных оценок PASS в качестве дескрипторов. Площадь под ROC-кривой составила 0,94 – для брадикардии, 0,96 – для тахикардии, 0,90 – для аритмии, 0,90 – для удлинения QT-интервала на ЭКГ, 0,91 – для гипертонии и 0,89 – для гипотонии.

КЛЮЧЕВЫЕ СЛОВА: межлекарственные взаимодействия; (Q)SAR; нежелательные лекарственные реакции.

DOI: 10.31857/S0320972523050068, **EDN:** AXVAFQ

ВВЕДЕНИЕ

Приём лекарственных средств (ЛС) сопряжён с риском возникновения нежелательных лекарственных реакций (НЛР), некоторые из которых могут нести серьёзные риски для жизни и здоровья пациентов [1, 2]. Зачастую возникновение НЛР связано с нежелательным взаимодействием лиганда со сторонней, не связанной с терапевтическим эффектом, мишенью в организме человека как напрямую, так и опосредованно через метаболиты, обра-

зующиеся в результате биотрансформации. Данная проблема усугубляется в случае, когда осуществляется приём нескольких ЛС одновременно [3]. Поскольку число потенциальных мишеней и число связанных с ними сигнальных и метаболических путей при совместном приёме становится значительно больше, чем в случае независимого друг от друга приёма лекарств, в ряде случаев создаются условия для существенного роста риска возникновения НЛР. О большинстве клинически значимых межлекарственных взаимодействий (МЛВ)

Принятые сокращения: ЛС – лекарственное средство; МЛВ – межлекарственные взаимодействия; НЛР – нежелательные лекарственные реакции; AUC – площадь под ROC-кривой; MNA – многоуровневые окрестности атомов; PRR – пропорциональные коэффициенты отчётности.

* Адресат для корреспонденции.

становится известно, как правило, постфактум [4]. В частности, пункт 4 статьи 64 Федерального закона от 12 апреля 2010 г. № 61-ФЗ «Об обращении лекарственных средств» регламентирует обязанность производителей лекарственных препаратов собирать данные о МЛВ и принимать меры в случае, если они вызывают угрожающие здоровью последствия. В свою очередь, недостаточная информированность врачей о рисках МЛВ также является причиной большого числа НЛР у пациентов. Для иллюстрации данного факта достаточно указать на то, что большинство врачебных рекомендаций составляются под одну нозологию, а рекомендации на случаи мультиморбидности зачастую отсутствуют [5]. Данная проблема особенно актуальна для педиатрии, где среднее количество нозологий в зависимости от возраста может варьироваться от 4 до 10 [6].

Одним из важнейших источников знаний о НЛР, и в особенности о МЛВ, является анализ спонтанных сообщений [7]. Спонтанное сообщение (англ. spontaneous report) — это извещение о возникновении НЛР, не задокументированной на стадии клинических испытаний лекарств, поданное в контрольные органы. Главным же недостатком метода спонтанных сообщений можно назвать тот факт, что не все медицинские работники пользуются им по ряду причин [8].

Алгоритмы анализа спонтанных сообщений в том или ином виде основаны на нахождении соотношения частот встречаемости (отношения шансов) для поиска ассоциаций между лекарствами или их комбинациями и НЛР [9].

По состоянию на 2013 г., НЛР, связанные с МЛВ, составили 19,8% всех спонтанных сообщений, полученных Росздравнадзором [10]. В Управление по санитарному надзору за качеством пищевых продуктов и медикаментов (англ. Food and Drug Administration, FDA) ежегодно поступает приблизительно 0,25 миллиона сообщений, среди которых на МЛВ приходится порядка 30% [11].

Вследствие этого возникает необходимость в прогнозировании МЛВ на как можно более ранней стадии разработки ЛС и на ранней стадии применения препарата в клинической практике [12]. Это позволит избежать как ущерба для здоровья пациентов, так и экономических издержек для представителей фармацевтической промышленности. Метод прогноза МЛВ должен предоставить возможность предсказывать конкретные НЛР при приёме парных комбинаций ЛС, что, в свою очередь, позволит быстрее и эффективнее отслеживать НЛР в целом. На данный момент существуют

методы предсказания МЛВ, опирающиеся на различные биомедицинские данные, такие как профили экспрессии генов [13], молекулярные сети белок-белковых взаимодействий [14] или же данные молекулярной динамики [15]. Используемые в подобных работах методы построения моделей также отличаются, начиная с методов анализа сети взаимодействия белков [14] и заканчивая методами машинного обучения [16]. Все эти подходы объединяет повышенная трудоёмкость как при сборе данных, так и непосредственно при построении моделей, заключающаяся в поиске данных по белок-белковым взаимодействиям и адекватном описании системы.

Модели для предсказания НЛР, вызванных МЛВ, уже существуют, но имеют ряд недостатков. В частности, найденные в литературе модели не способны выполнить прогноз для новых соединений, поскольку они основаны на данных об известных мишенях [16]. Другие модели используют данные, полученные из спонтанных сообщений без предварительной обработки [17], что снижает точность прогноза из-за неточного представления связей между лекарствами и НЛР.

Цель данной работы заключалась в построении моделей для прогнозирования НЛР со стороны сердечно-сосудистой системы, обусловленных МЛВ между парами лекарств, на основе анализа связей «структура–активность» (англ. (Quantitative) Structure-Activity Relationships, (Q)SAR) с использованием данных из открытых источников и результатов анализа спонтанных сообщений, в частности.

МАТЕРИАЛЫ И МЕТОДЫ

Получение данных об эффектах межлекарственных взаимодействий. Данные о неблагоприятных эффектах МЛВ были получены из базы данных DrugBank [18] и приведены к табличному виду с помощью процедуры парсинга с использованием регулярных выражений на языке R: 4.2.1 (2022-06-23) (<https://www.r-project.org>) в среде RStudio 2021.09.0+351 «Ghost Orchid». Названия НЛР были заменены на синонимичные названия в соответствии с медицинским словарём для регуляторной деятельности MedDRA (<https://www.meddra.org/>). Таким образом было получено 127 294 пары лекарств, состоящих из 1632 индивидуальных соединений и ассоциированных с шестью сердечно-сосудистыми НЛР: аритмией, тахикардией, брадикардией, удлинением интервала QT на электрокардиограмме, гипертензией и гипотензией.

Получение отрицательных примеров. Полученные данные представлены парами лекарств, взаимодействие которых приводит к возникновению шести НЛР. Для построения (Q)SAR-моделей этого недостаточно, поскольку обучающая выборка в равной мере должна включать как связанные с НЛР пары лекарств («положительные примеры»), так и те пары, которые однозначно не связаны с рассматриваемым эффектом («отрицательные примеры»).

В рамках данной работы применялся подход, согласно которому отрицательными примерами считались те пары, для которых имеется информация о способности вызывать какие-либо эффекты, кроме исследуемого. Наличие такой информации для пары лекарств свидетельствует об изученности её профиля эффектов МЛВ и повышает вероятность отсутствия исследуемого эффекта в соответствующем профиле. Информация о парах лекарств, которые с высокой вероятностью не вызывают исследуемые НЛР, была получена из базы данных TwoSides [19]. TwoSides предоставляет результаты анализа диспропорциональности, примененного к данным из спонтанных сообщений [20]. Каждое спонтанное сообщение содержит информацию о лекарствах, которые принимал пациент, и о возникших НЛР. Сущность анализа диспропорциональности заключается в сравнении доли сообщений по лекарству, где упоминается НЛР, к доле соответствующих сообщений в базе по всем лекарствам с последующим расчётом величины Proportional Reporting Ratio (PRR). Эта величина отражает силу и вероятность связи между лекарством и НЛР [19]. TwoSides предоставляет информацию о величинах PRR, вычисленных для индивидуальных лекарств и их пар, по отношению ко всем НЛР. Ранее было показано, что соответствующая информация может быть использована для предсказания пар лекарств, которые способны и, наоборот, не способны вызывать конкретные НЛР [7]. Информация о НЛР для пар лекарств из TwoSides была дополнена информацией о НЛР одиночных лекарств из базы данных OffSides. Первая содержит информацию о PRR и его стандартной ошибке для пар лекарств и НЛР, вторая содержит те же параметры, но уже для индивидуальных лекарств, а не для пар. Обе выборки были объединены таким образом, что конечная выборка содержала следующие величины:

PRR, PRR1, PRR2 — величины, отражающие силу и вероятность связи с НЛР, относящиеся к паре и к отдельным лекарствам соответственно.

Поскольку спонтанные сообщения могут содержать ошибки и неточности, авторы баз данных TwoSides и OffSides предложили оценивать статистическую значимость величин PRR по логарифмической формуле (1):

$$\log(\text{PRR} \dots) - 1,96 \times \text{PRRError} \dots > \log 2, \quad (1)$$

где PRRError, PRRError1, PRRError2 — стандартные ошибки вычисления PRR.

Ассоциации пар лекарств с НЛР, удовлетворяющие данному неравенству, считаются значимыми, а прочие — незначимыми. Более подробное описание способов вычисления указанных выше величин можно получить из публикации базы данных TwoSides [19].

В рамках данной работы были выделены две группы ассоциаций пар с НЛР.

1. Ассоциации, в которых обе величины, PRR1 и PRR2, удовлетворяют неравенству (1). Такие ассоциации могут являться примерами аддитивного взаимодействия.

2. Ассоциации, в которых величина PRR для пары удовлетворяет неравенству (1) и превышает максимальную из двух величин PRR1 и PRR2. Такие ассоциации могут являться примерами синергетического взаимодействия.

Для дальнейшего анализа был составлен список пар, принадлежащих как минимум к одной из двух групп и ассоциированных хотя бы с двумя НЛР, относящимися к нарушениям сердечного ритма и артериального давления соответственно. Эти пары были исключены из дальнейшего рассмотрения, поскольку отрицательные примеры не должны содержать пар лекарств, ассоциированных с искомой НЛР. Оставшиеся пары, упоминающиеся хотя бы в 100 спонтанных сообщениях, были использованы в качестве отрицательных примеров (рис. 1).

Генерация дескрипторов. Для компьютерного описания структур каждой пары лекарств применялись два типа дескрипторов: биологические активности, предсказанные с помощью программы PASS [21], и PoSMNA-дескрипторы [22].

Дескрипторы на основе спектра биологической активности. Программа PASS позволяет прогнозировать 7017 видов биологической активности по структурной формуле соединения со средней точностью 0,96 и минимальной точностью 0,85. Спектр биологической активности PASS включает фармакотерапевтические эффекты, молекулярные механизмы действия, взаимодействие с ферментами метаболизма и транспортёрами, побочные и токсические эффекты, влияние на транскрипцию генов.

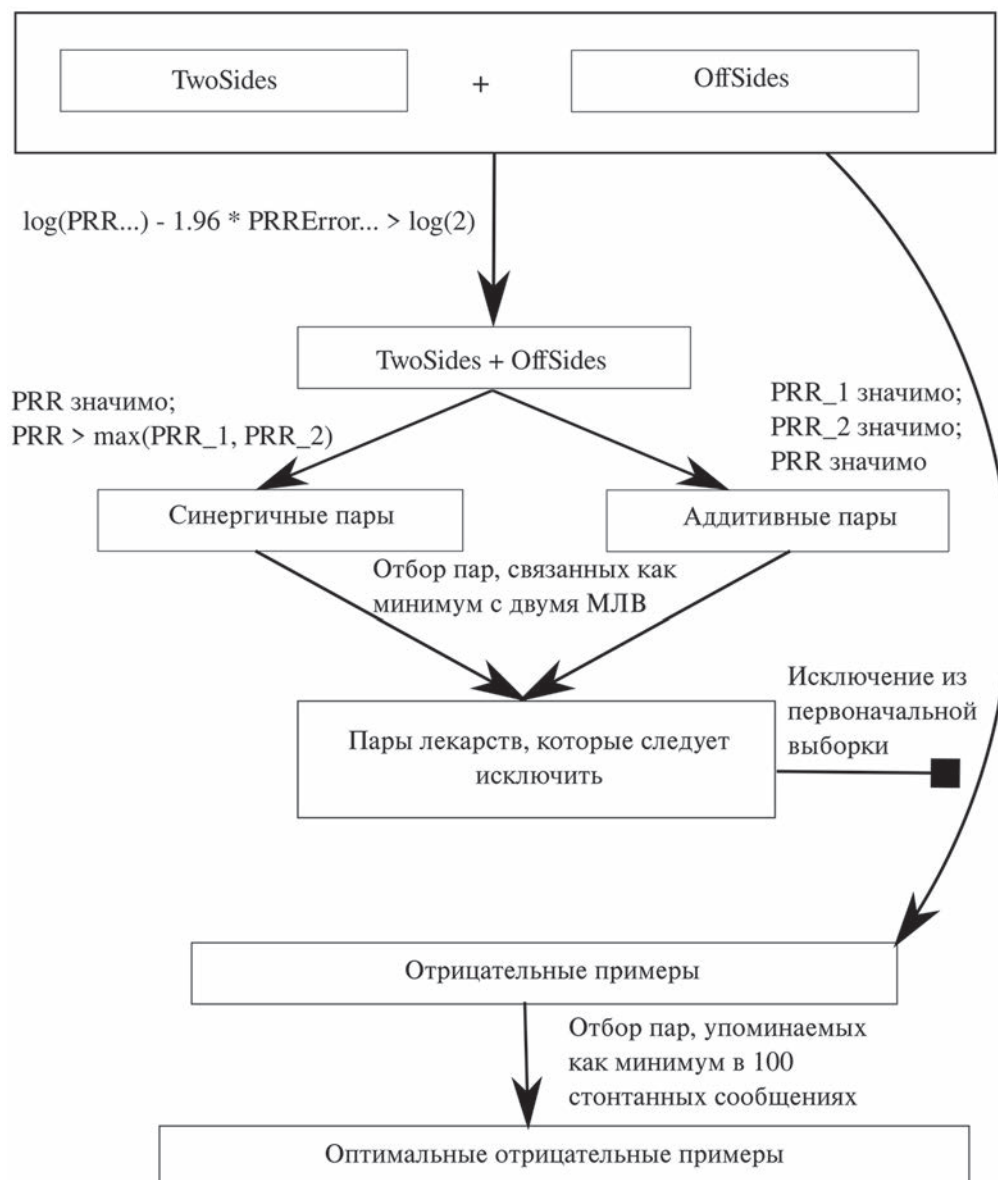


Рис. 1. Схема генерации отрицательных примеров с использованием баз данных TwoSides и OffSides

PASS вычисляет две оценки для каждой биологической активности: P_a — вероятность наличия активности и P_i — вероятность отсутствия активности. Для дальнейших вычислений была использована величина P , полученная на основе этих оценок, по формуле (2):

$$P = \frac{P_a}{P_a + P_i}, \quad (2)$$

где P_a — вероятность наличия активности, P_i — вероятность отсутствия активности, P — объединение P_a и P_i .

Возможные значения полученной величины P лежат в интервале от 0 до 1. Для вычисления числовых дескрипторов, описывающих пару соединений, были найдены нормализованные суммы и модули разностей величин P ,

полученных для индивидуальных соединений (P_1 и P_2) (формула (3)):

$$D_1 = \frac{P_1 + P_2}{2}; \quad D_2 = \frac{|P_1 - P_2|}{2}, \quad (3)$$

где P_1 и P_2 — величины, отражающие вероятность наличия активности для индивидуальных соединений в составе пары. D_1 и D_2 — это дескрипторы пары, представляющие собой суммы и модули разностей величин P_1 и P_2 соответственно. Таким образом, для каждой пары соединений было получено 14 032 числовых дескриптора, значения которых лежат в интервале от 0 до 1.

Структурные дескрипторы PoSMNA. PoSMNA-дескрипторы являются линейными комбинациями фрагментных MNA-дескрипторов вто-

рого порядка [23], которые, в свою очередь, представляют описание структуры молекулы ЛС с точки зрения ближайшего окружения каждого атома. Таким образом, MNA-дескрипторы второго порядка состоят из уникальных строк, соответствующих каждому атому и его окружению через 1 или 2 связи. MNA-дескрипторы позволяют однозначно описать структуру молекулы ЛС. Поскольку каждый MNA-дескриптор представляет собой строку символов, то для генерации PoSMNA-дескрипторов эти строки «склеивают» между собой [23]. Если соединение А описывается двенадцатью MNA-дескрипторами, соединение Б описывается шестью, то пара соединений будет описана 72 PoSMNA-дескрипторами, представляющими собой комбинации MNA-дескрипторов.

Для того чтобы полученные PoSMNA-дескрипторы можно было использовать для построения моделей, применяя метод машинного обучения, каждый элемент PoSMNA был конвертирован в целое положительное число через операцию хэширования с помощью пакета digest (<https://CRAN.R-project.org/package=digest>).

При данном подходе всё многообразие структурных фрагментов пар лекарств в выборке описывается вектором конечной длины, каждое значение в котором указывает на наличие или отсутствие определённого структурного фрагмента в данной паре лекарств. Поскольку при высокой длине вектора дескрипторы получаются слишком разреженными для прогноза, а при слишком короткой — теряется часть данных о структуре, в рамках данной работы длина вектора подбиралась эмпирическим путём. Наилучшие значения точности прогноза были достигнуты при длине вектора в 15 000 значений. Таким образом каждой паре лекарств была поставлена в соответствие строка нулей и единиц длиной 15 000 знаков (так называемый молекулярный отпечаток). В итоге было достигнуто две цели: 1) дескрипторы приняли форму, приемлемую для машинного обучения и 2) за счёт незначительных потерь данных была достигнута значительная экономия вычислительных ресурсов и существенно уменьшено время, необходимое для вычислений.

Построение и оценка моделей. Классификационные модели были построены при помощи метода Random Forest, имплементированного в пакете R ranger [24, 25] со значениями гиперпараметров по умолчанию (число деревьев (ntree) — 500, число дескрипторов, отбираемых для построения каждого дерева (mtry) — корень квадратный от общего числа дескрипторов) с использованием двух

типов описания пар лекарств: дескрипторов, сгенерированных на основе результатов прогноза PASS, и PoSMNA-дескрипторов. Точность моделей была определена с использованием пакетов ROCR [26] и caret (<https://topepo.github.io/caret/>) при помощи процедуры пятикратного скользящего контроля.

Результатом применения алгоритмов классификации является вероятностная оценка принадлежности к определённому классу. Полученные оценки могут как совпадать, так и отличаться от известных значений класса, представленных в контрольной выборке. Исходя из этого, возможны четыре типа оценок результатов прогноза.

1) Истинно позитивный (англ., True Positive, TP) — случай, при котором модель определила класс как положительный, и он является таковым на самом деле.

2) Ложноположительный (англ., False Positive, FP) — случай, при котором модель ошибочно определила класс как положительный, в то время как он является отрицательным.

3) Истинно негативный (англ., True Negative, TN) — случай, при котором модель определила класс как отрицательный, и он является таковым на самом деле.

4) Ложноотрицательный (англ., False Negative, FN) — случай, при котором модель ошибочно определила класс как отрицательный, в то время как он является положительным.

На базе этих оценок рассчитываются такие показатели точности прогноза, как площадь под ROC-кривой (англ., area under the ROC curve, AUC) [26], чувствительность и специфичность.

AUC — это площадь под ROC-кривой. ROC-кривая строится следующим образом: на оси ординат откладывается доля правильно классифицированных положительных примеров, на оси абсцисс — доля ложноположительных предсказаний. Таким образом, исходя из свойства непрерывности функции вероятности, следует, что ROC-кривая выражает точность прогноза для каждого порога отсечения вероятности.

Чувствительность (TPR, от англ., true positive rate) — точность предсказания наличия НЛР. В данном случае это вероятность того, что пара, вызывающая конкретную НЛР, будет верно распознана. Рассчитывается по формуле (4):

$$TPR = \frac{TP}{TP + FP}, \quad (4)$$

где TP — истинно позитивные случаи, FP — ложноположительные случаи.

Специфичность (TNR, от англ., true negative rate) – точность предсказания отсутствия НЛР. В данном случае это вероятность того, что модель не обнаружит НЛР там, где её нет. Рассчитывается по формуле (5):

$$TNR = \frac{TN}{FN + TN}, \quad (5)$$

где FN – ложно отрицательные случаи, TN – истинно отрицательные случаи.

Этот показатель так же важен, как и чувствительность, поскольку позволяет избежать так называемой ошибки второго рода, при ко-

торой потенциально безвредная пара лекарств может классифицироваться как опасная для здоровья.

Сбалансированная точность (BA, от англ., Balanced accuracy) – среднее арифметическое от значений специфичности и чувствительности, как и AUC, является объективным показателем точности прогноза модели (формула (6)).

$$BA = \frac{TPR + TNR}{2}, \quad (6)$$

где TPR – чувствительность, TNR – специфичность.

Таблица 1. Характеристики выборок «положительных» и «отрицательных» примеров для шести НЛР

НЛР	Количество пар (положительные примеры)	Количество пар (отрицательные примеры)	Количество лекарств (положительные примеры)	Количество лекарств (отрицательные примеры)
Аритмия	9231	9231	366	685
Удлинение интервала QT	19 371	19 371	484	740
Брадикардия	4848	4848	186	634
Тахикардия	6506	6506	216	666
Гипертензия	24 040	24 040	501	752
Гипотензия	22 890	22 890	925	750

Таблица 2. Результаты прогноза для двенадцати моделей

Оценки точности модели	Брадикардия	Аритмия	Удлинение интервала QT	Тахикардия	Гипертензия	Гипотензия
PoSMNA						
AUC	0,91	0,88	0,88	0,93	0,94	0,89
Чувствительность	0,71	0,69	0,69	0,73	0,75	0,67
Специфичность	0,97	0,99	0,91	0,98	0,95	0,91
Сбалансированная точность	0,84	0,84	0,80	0,86	0,85	0,79
PASS						
AUC	0,94	0,90	0,90	0,96	0,91	0,89
Чувствительность	0,65	0,55	0,76	0,70	0,85	0,77
Специфичность	0,97	0,96	0,86	0,97	0,82	0,85
Сбалансированная точность	0,81	0,76	0,81	0,83	0,83	0,81

Примечание. Даны результаты прогноза для шести НЛР с использованием двух типов дескрипторов.

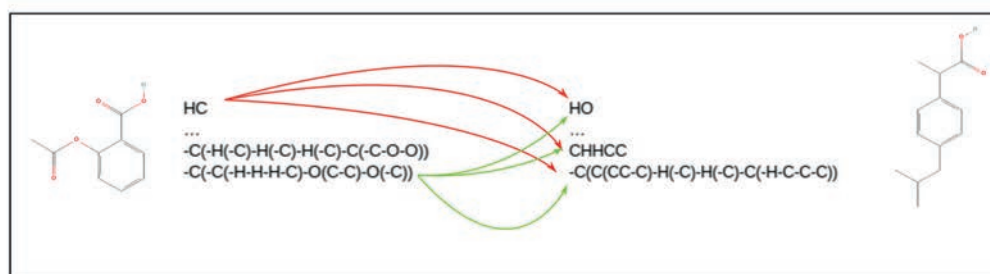
РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Характеристики полученных выборок. На основе анализа информации, полученной из баз данных DrugBank и TwoSides, были созданы выборки пар ЛС, которые в результате взаимодействия вызывают и не вызывают шесть сердечно-сосудистых НЛР (табл. 1). Четыре НЛР (аритмия, тахикардия, брадикардия и удлинение интервала QT) связаны с нарушениями сердечного ритма и две НЛР (гипертензия и гипотензия) связаны с изменением артериального давления. Выборки

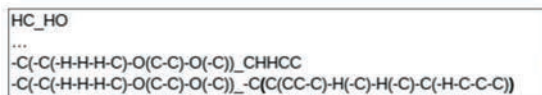
были опубликованы в общем доступе и свободно доступны по ссылке (https://github.com/allaroundddi/DDI_samples).

Оценки полученных моделей. На основе полученных выборок, с использованием двух типов дескрипторов (см. раздел «Материалы и методы») и метода Random Forest, были построены классификационные модели для прогноза соответствующих НЛР. Схема получения PoSMNA-дескрипторов представлена на рис. 2, а схема перекрёстного контроля — на рис. 3. В результате были получены следующие показатели точности прогноза (табл. 2).

MNA-дескрипторы двух соединений объединялись таким образом, что m дескрипторов одного соединения и n другого давали $m \cdot n$ PoSMNA дескрипторов для пары.



Таким образом получилось $m \cdot n$ текстовых строк, описывающих пару через комбинацию структурных фрагментов двух отдельных соединений.



Каждой строке PoSMNA дескриптора ставилось в соответствие целое положительное число.

HC_HO	→	207167035
...	...	...
-C-(C(H-H-H-C)-O(C-C)-O(C))_CHHCC	→	727650521
-C-(C(H-H-H-C)-O(C-C)-O(C))_C(C(C-C-H)-(C)-(C)-(C-C-C))	→	262076202

2035
521
...
2620

	2035	521	...	2620
Аспирин + ибупрофен	1	1	...	1
Аспирин + флурбипрофен	1	0	...	1
...			...	
Изониазид + транилципромин	0	0	...	

Таким образом была составлена таблица, в которой каждая строка — это пара лекарств, а каждый столбец — наличие или отсутствие определённого структурного фрагмента в этой паре. Таким образом каждой паре поставлена в соответствие битовая строка.

Рис. 2. Иллюстрация алгоритма получения хэшированных PoSMNA-дескрипторов

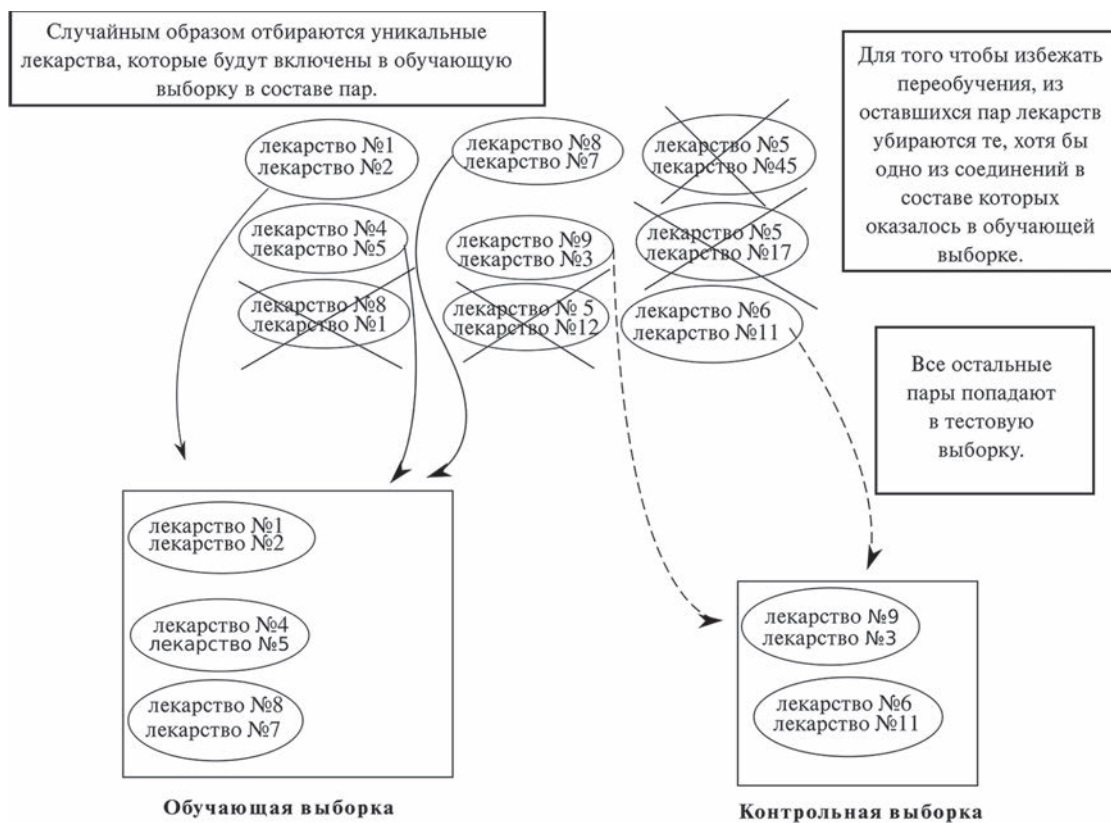


Рис. 3. Иллюстрация метода скользящего контроля, примененного в исследовании

С использованием моделей, построенных на двух типах дескрипторов, был выполнен консенсусный прогноз с различными вариантами комбинирования оценок вероятностей. Наибольшее увеличение показателей точности прогноза было достигнуто для отбора оценок по максимальному значению: в среднем по выборке AUC увеличилась на 1,8%, а сбалансированная точность — на 4,1%. Результаты представлены в табл. 3.

Ретроспективный анализ. Для демонстрации применимости разработанного подхода был выполнен прогноз для нескольких пар лекарств. Из базы данных DrugBank и сайта Drugs.com (<https://www.drugs.com/>) случайным образом были отобраны шесть пар лекарств, три из которых вызывают удлинение интервала QT на электрокардиограмме, и три — нет. Данная НЛР была выбрана потому, что для неё имеется больше всего данных по МЛВ. Три пары лекарств были выбраны таким образом, чтобы факт наличия у них НЛР был подтверждён литературными источниками, а оставшиеся три пары (отрицательные примеры) были выбраны произвольным образом. После этого данные пары были исключены на этапе построения и валидации моделей, и с использованием перестроенных моделей был осуществлён прогноз. Результаты приведены в табл. 4.

Как видно по последним двум столбцам, известная способность пар лекарств вызывать удлинение интервала QT на ЭКГ предсказана верно.

Сравнение с имеющимися работами. В рамках данной работы были построены (Q)SAR-модели, которые позволяют предсказывать конкретные неблагоприятные эффекты МЛВ, в то время как большая часть работ посвящена прогнозированию факта наличия МЛВ.

В работе Liu et al. [28] выборки были получены с использованием базы данных по взаимодействиям лекарство—белок STITCH, которые послужили признаками отдельных лекарств и ранее упомянутой в текущей работе базы данных МЛВ TwoSides. В качестве отрицательных примеров применялись те пары, для которых нет данных по их способности вызывать конкретную НЛР. Построение моделей производилось методом, основанным на статистике Байеса. Полученные модели обладали следующими средними показателями точности прогноза: чувствительность — 63%, специфичность — 90%, AUC — 86%. Соответствующие значения для нашей работы равны 70%, 90% и 90%.

В работе Shankar et al. [17] использовались данные о МЛВ из TWOSIDES и данные о влиянии лекарств на экспрессию генов

Таблица 3. Консенсус-прогноз для шести НЛР с усреднением оценок вероятности, выбором наибольшей оценки, а также оценками без использования консенсуса

НЛР	Оценка точности модели	Max	Mean	PASS	PoSMNA
Аритмия	AUC	0,95	0,95	0,90	0,88
	сбалансированная точность	0,89	0,81	0,76	0,84
	чувствительность	0,82	0,69	0,55	0,69
	специфичность	0,96	0,99	0,96	0,99
Брадикардия	AUC	0,94	0,94	0,94	0,91
	сбалансированная точность	0,88	0,82	0,81	0,84
	чувствительность	0,81	0,64	0,65	0,71
	специфичность	0,95	0,99	0,97	0,97
Гипертензия	AUC	0,95	0,95	0,91	0,94
	сбалансированная точность	0,88	0,89	0,84	0,85
	чувствительность	0,92	0,83	0,85	0,75
	специфичность	0,83	0,94	0,82	0,95
Гипотензия	AUC	0,92	0,92	0,89	0,89
	сбалансированная точность	0,85	0,83	0,81	0,79
	чувствительность	0,85	0,74	0,77	0,67
	специфичность	0,85	0,91	0,85	0,91
Тахикардия	AUC	0,96	0,97	0,96	0,93
	сбалансированная точность	0,91	0,86	0,84	0,86
	чувствительность	0,84	0,72	0,70	0,73
	специфичность	0,97	0,99	0,97	0,98
QT	AUC	0,92	0,92	0,90	0,88
	сбалансированная точность	0,85	0,84	0,81	0,75
	чувствительность	0,82	0,76	0,76	0,69
	специфичность	0,88	0,91	0,86	0,80

из LINC S L1000 с построением моделей посредством искусственных нейронных сетей. В отличие от нашей работы, данные из TWOSIDES в работе Shankar et al. используются в полном объёме без предварительного отбора и контроля качества. К сожалению, авторы не приводят средних значений оценки прогноза, тем не менее для перечисленных в

работе моделей среднее значение AUC равняется 70%, в то время как в нашей работе оно равно 90%.

Область применимости метода. Разработанный нами метод позволяет предсказывать НЛР при совместном приёме нескольких лекарственных средств. Тем не менее у метода существуют определённые ограничения применимости метода.

Таблица 4. Ретроспективный анализ точности модели для прогноза удлинения интервала QT

Лекарство № 1	Лекарство № 2	Характеристика пары	Вызывает удлинение интервала QT, согласно известным данным	Спрогнозированная вероятность удлинения интервала QT
Рамиприл	Ранолазин	кардиопротектор + + антиангинальное средство	нет	0,38
Варденафил	Энкорafenиб	ингибитор фосфодиэстеразы + + ингибитор киназ	да ¹	0,68
Такролимус	Ломефлоксацин	иммуносупрессор + + антибиотик	да [27]	0,69
Псевдоэфедрин	Серталиин	адрено- и симпатомиметик + + антидепрессант	нет	0,08
Розувастин	Трандолаприл	гиполипидемический препарат + ингибитор ангиотензинпревращающего фермента	нет	0,02
Гранисетрона	Гареноксацин	противорвотное средство + + диагностическое средство	да ²	0,84

Примечание. Нет — означает, что для данной комбинации лекарств не удалось найти информации о риске возникновения удлинения интервала QT, в ином случае указано слово «да» и дана ссылка на источник.

¹ «Product Information. Bravado (encorafenib)» Array BioPharma Inc., Boulder, CO.

² «Product Information. Definity (perflutren)» Lantheus Medical Imaging Inc., North Billerica, MA.

Во-первых, это программные ограничения: программа PASS, с помощью которой были получены соответствующие дескрипторы, имеет программное ограничение по молекулярной массе рассматриваемых веществ (1250 кДа), а также не различает оптические изомеры и супрамолекулярные структуры. Во-вторых, это ограничение в доступности данных. Несмотря на то что наш метод не требует каких-либо данных, кроме взаимосвязи пар лекарств с МЛВ, не для всех НЛР удаётся собрать достаточное число данных. Так, в данной работе не проводилось построения моделей для оценки инфаркта миокарда и мерцательной аритмии, поскольку для них крайне мало доступных данных. В-третьих, это ограниченность метода структурой соединений, входящих в состав обучающей выборки. Вследствие этого данные модели не смогут привести адекватный прогноз для лекарств с принципиально новой химической структурой.

ЗАКЛЮЧЕНИЕ

Построены модели для предсказания шести наиболее важных сердечно-сосудистых НЛР, обусловленных МЛВ, на базе двух видов дескрипторов. Показатели точности, вычислен-

ные при помощи пятикратного скользящего контроля, превышают 80%. Также был выполнен консенсус-прогноз для двух типов моделей, что привело к небольшому увеличению оценок точности прогноза. Средние значения величины AUC (~ 90%) для шести эффектов, связанных с сердечно-сосудистой системой, свидетельствуют о высокой точности бинарной классификации, а близость значений чувствительности и специфичности говорит о правильно выбранном подходе к созданию отрицательных примеров. Кроме того, использование выбранных дескрипторов позволяет добиться приемлемой точности без использования анализа дополнительных биомедицинских данных, включая анализ профилей экспрессии генов, методы сетевой фармакологии и молекулярной динамики [13–15]. С использованием построенных моделей нами был выполнен прогноз для случайно отобранных пар ЛС с известным риском возникновения конкретной НЛР. Результаты прогноза полностью подтверждаются ретроспективным анализом, подкреплённым литературными данными. Таким образом, полученные модели могут быть использованы для прогнозирования НЛР, возникающих при совместном приёме произвольной пары лекарств. Высокие значения точности прогноза говорят о потенциальной возможности приме-

нения аналогичного алгоритма при построении моделей для прогнозирования НЛР МЛВ, относящихся к другим системам органов.

Вклад авторов. Иванов С.М. — концепция и руководство работой; Сухачёв В.С. — выполнение работы и написание текста статьи; Дмитриев А.В. — консультация, помощь в подготовке рукописи.

Финансирование. Исследование выполнено при финансовой поддержке Российского научного фонда (грант № 17-75-20250).

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Соблюдение этических норм. Настоящая статья не содержит каких-либо исследований с участием людей или с использованием животных в качестве объектов.

СПИСОК ЛИТЕРАТУРЫ

1. Rekić, D., Reynolds, K. S., Zhao, P., Zhang, L., Yoshida, K., Sachar, M., Piquette, M. M., Huang, S. M., and Zineh, I. (2017) Clinical drug–drug interaction evaluations to inform drug use and enable drug access, *J. Pharm. Sci.*, **106**, 2214–2218, doi: 10.1016/j.xphs.2017.04.016.
2. Carpenter, M., Berry, H., and Pelletier, A. L. (2019) Clinically relevant drug–drug interactions in primary care, *Am. Fam. Physician*, **99**, 558–564.
3. Kim, J., and Parish, A. L. (2017) Polypharmacy and medication management in older adults, *Nurs. Clin. North Am.*, **52**, 457–468, doi: 10.1016/j.cnur.2017.04.007.
4. Triaridis, S., Tsiropoulos, G., Rachovitsas, D., Psillas, G., and Vital, V. (2009) Spontaneous haematoma of the pharynx due to a rare drug interaction, *Hippokratia*, **13**, 175–177.
5. Костылева М. Н., Строк А. Б., Постников С. С., Грацианская А. Н., Ермилин А. Е. (2022) Фармакотерапия в многопрофильном педиатрическом стационаре: полипрагмазия и риск лекарственных взаимодействий на примере клинического случая, *Безопасность и риск фармакотерапии*, **10**, 302–314, doi: 10.30895/2312-7821-2022-10-3-302-314.
6. Ключников С. О. (2014) Полипрагмазия: пути решения проблемы, *Детские инфекции*, **13**, 36–41, doi: 10.22627/2072-8107-2014-13-4-36-41.
7. Ivanov, S., Lagunin, A., Filimonov, D., and Poroikov, V. (2019) Assessment of the cardiovascular adverse effects of drug–drug interactions through a combined analysis of spontaneous reports and predicted drug–target interactions, *PLoS Comput. Biol.*, **15**, e1006851, doi: 10.1371/journal.pcbi.1006851.
8. Зырянов С. К., Затолочина К. Э., Казаков А. С. (2022) Актуальные вопросы обеспечения безопасности пациентов: роль фармаконадзора, *Общественное здоровье*, **2**, 25–34, doi: 10.21045/2782-1676-2021-2-3-25-34.
9. Noguchi, Y., Tachi, T., and Teramachi, H. (2020) Comparison of Signal detection algorithms based on frequency statistical model for drug–drug interaction using spontaneous reporting systems, *Pharm. Res.*, **37**, 86, doi: 10.1007/s11095-020-02801-3.
10. Казаков А. С., Лепяхин В. К., Астахова А. В. (2013) Осложнения фармакотерапии, связанные с взаимодействием лекарственных средств, *Рос. мед. биол. вестн. им. акад. И.П. Павлова*, **21**, 70–76, doi: 10.17816/pavlovj2013370-76.
11. Chen, Z., Elizabeth, R., Lin, L., Nicole, P., and Jiuyong, L. (2020) Detecting high-quality signals of adverse drug–drug interactions from spontaneous reporting data, *JBI*, **112**, 103603, doi: 10.1016/j.jbi.2020.103603.
12. Strandell, J., Bate, A., Lindquist, M., and Edwards, I. R. (2008) Drug–drug interactions – a preventable patient safety issue? *Br. J. Clin. Pharmacol.*, **65**, 144–146, doi: 10.1111/j.1365-2125.2007.02981.x.
13. Taguchi, Y., and Turki, T. (2021) Novel method for the prediction of drug–drug interaction based on gene expression profiles, *Eur. J. Pharm. Sci.*, **160**, 105742, doi: 10.1016/j.ejps.2021.105742.
14. Huang, J., Niu, C., Green, C. D., Yang, L., Mei, H., and Han, J. D. (2013) Systematic prediction of pharmacodynamic drug–drug interactions through protein–protein–interaction network, *PLoS Comput. Biol.*, **9**, e1002998, doi: 10.1371/journal.pcbi.1002998.
15. Varma, M. V., Pang, K. S., Isoherranen, N., and Zhao, P. (2015) Dealing with the complex drug–drug Interactions: towards mechanistic models, *Biopharm. Drug Dispos.*, **36**, 71–92, doi: 10.1002/bdd.1934.
16. Kastrin, A., Ferk, P., and Leskošek, B. (2018) Predicting potential drug–drug interactions on topological and semantic similarity features using statistical learning, *PLoS One*, **13**, e0196865, doi: 10.1371/journal.pone.0196865.
17. Shankar, S., Bhandari, I., Okou, D. T., Srinivasa, G., and Athri, P. (2021) Predicting adverse drug reactions of two-drug combinations using structural and transcriptomic drug representations to train an artificial neural network, *Chem. Biol. Drug. Des.*, **97**, 665–673, doi: 10.1111/cbdd.13802.
18. Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A.,

- Chin, L., Cummings, R., Le, D., Pon, A., Knox, C., and Wilson, M. (2018) DrugBank 5.0: A major update to the DrugBank database for 2018, *Nucleic Acids Res.*, **46**, D1074-D1082, doi: 10.1093/nar/gkx1037.
19. Tatonetti, N. P., Ye, P. P., Daneshjou, R., and Altman, R. B. (2012) Data-driven prediction of drug effects and interactions, *Sci. Transl. Med.*, **4**, 125, doi: 10.1126/scitranslmed.3003377.
20. Hazell, L., and Shakir, S. A. W. (2006) Under-reporting of adverse drug reactions: a systematic review, *Drug Saf.*, **29**, 385-396, doi: 10.2165/00002018-200629050-00003.
21. Filimonov, D. A., and Poroikov, V. V. (2008) *Probabilistic Approaches in Activity Prediction. Chemoinformatics Approaches to Virtual Screening*, RSC Publishing, Cambridge, pp. 182-216, doi: 10.1039/97818475558879-00182.
22. Dmitriev, A., Filimonov, D., Lagunin, A., Karasev, D., Pogodin, P., Rudik, A., and Poroikov, V. (2019) Prediction of severity of drug-drug interactions caused by enzyme inhibition and activation, *Molecules*, **24**, E3955, doi: 10.3390/molecules24213955.
23. Filimonov, D., Poroikov, V., Borodina, Y., and Glorizova, T. (1999) Chemical similarity assessment through multilevel neighborhoods of atoms: definition and comparison with the other descriptors, *J. Chem. Inf. Comput. Sci.*, **39**, 666-670, doi: 10.1021/ci980335o.
24. Breiman, L. (2001) Random forests, *Mach. Learn.*, **45**, 5-32, doi: 10.1023/A:1010933404324.
25. Wright, M. N., and Ziegler, A. (2017) Ranger: a fast implementation of Random Forests for high dimensional data in C++ and R, *J. Stat. Softw.*, **77**, 1-17, doi: 10.18637/jss.v077.i01.
26. Sing, T., Sander, O., Beerenwinkel, N., and Lengauer, T. (2005) ROCr: visualizing classifier performance in R, *Bioinformatics*, **21**, 3940-3941, doi: 10.1093/bioinformatics/bti623.
27. Witchel, H. J., Hancox, J. C., and Nutt, D. J. (2003) Psychotropic drugs, cardiac arrhythmia, and sudden death, *J. Clin. Psychopharmacol.*, **23**, 58-77, doi: 10.1097/00004714-200302000-00010.
28. Liu, R., AbdulHameed, M. D. M., Kumar, K., Yu, X., Wallqvist, A., and Reifman, J. (2017) Data-driven prediction of adverse drug reactions induced by drug-drug interactions, *BMC Pharmacol. Toxicol.*, **18**, 44, doi: 10.1186/s40360-017-0153-6.

PREDICTION OF ADVERSE EFFECTS OF DRUG–DRUG INTERACTIONS ON THE CARDIOVASCULAR SYSTEM BASED ON THE ANALYSIS OF STRUCTURE–ACTIVITY RELATIONSHIPS

V. S. Sukhachev*, S. M. Ivanov, and A. V. Dmitriev

Institute of Biomedical Chemistry, 119121 Moscow, Russia; e-mail: withstanding@yandex.ru

The co-administration of drugs often leads to drug–drug interactions, which may be accompanied by various adverse drug reactions that pose a threat to the life and health of patients. The effect caused by adverse drug reactions from the cardiovascular system is one of the most significant manifestations of drug–drug interaction. Clinical assessment of adverse drug reactions resulting from drug–drug interaction between all drug pairs used in therapeutic practice is not possible. The purpose of this work was to build models using structure-activity analysis to predict the effects of adverse drug reactions on the cardiovascular system, mediated by interactions between drug pairs when they are taken together. Data on adverse effects resulting from drug–drug interaction were obtained from the DrugBank database. The data on drug pairs that do not cause the corresponding effects, necessary for building accurate structure–activity models, were obtained from the TwoSides database, which contains the results of the analysis of spontaneous reports. Two types of descriptors were used to describe a pair of drug structures: PoSMNA descriptors and probabilistic estimates of the prediction of biological activities obtained using the PASS program. Structure–activity relationships were established using the Random Forest method. Prediction accuracy was calculated by means of a five-fold cross-validation. The highest accuracy values were obtained using PASS probabilistic estimates as descriptors. The area under the ROC curve was 0.94 for bradycardia, 0.96 for tachycardia, 0.90 for arrhythmia, 0.90 for ECG QT prolongation, 0.91 for hypertension, 0.89 for hypotension.

Keywords: drug–drug interaction, (Q)SAR, adverse drug reactions