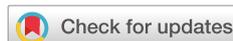


УДК 615.015.11:544.165:[615.065+615.033/034]:575.112:[519.23+004.032.26+004.852]
<https://doi.org/10.30895/2312-7821-2023-11-4-390-408>



Обзорная статья / Review

Прогноз *in silico* токсикологических и фармакокинетических характеристик лекарственных соединений

П.М. Васильев[✉], А.В. Голубева, А.Р. Королева, М.А. Перфильев, А.Н. Кочетков

Федеральное государственное бюджетное образовательное учреждение высшего образования «Волгоградский государственный медицинский институт»
Министерства здравоохранения Российской Федерации,
Площадь Павших борцов, д. 1, Волгоград, 400131, Российская Федерация

✉ Контактное лицо: Васильев Павел Михайлович pvassiliev@mail.ru

РЕЗЮМЕ

Актуальность. Изучение токсикологических и фармакокинетических свойств лекарственных соединений является важным этапом доклинических исследований и в случае неудовлетворительных результатов может привести к невозможности дальнейшей разработки лекарственного препарата. Поэтому создание компьютерных методов для предварительного скрининга свойств химических соединений является актуальной задачей.

Цель. Обзор современных подходов к компьютерному прогнозу ADMET (Absorption, Distribution, Metabolism, Excretion, Toxicity) свойств фармакологически активных веществ, в особенности наиболее важных токсикологических и фармакокинетических показателей, и изложение результатов собственных исследований в этой области.

Обсуждение. В результате проведенного обзора моделей для прогноза токсикологических характеристик химических соединений (острой токсичности, канцерогенности, мутагенности, генотоксичности, эндокринологической токсичности, цитотоксичности, кардиотоксичности, гепатотоксичности, иммунотоксичности) установлено, что точность прогноза указанных свойств колеблется от 74,0 до 98,0%. Результаты обзора моделей для прогноза фармакокинетических характеристик химических соединений (всасываемости в желудочно-кишечном тракте (ЖКТ), биодоступности при пероральном приеме, объема распределения, клиренса общего, почечного и печеночного, времени полувыведения) свидетельствуют о том, что коэффициент детерминации при прогнозе указанных свойств колеблется от 0,265 до 0,920. Проведенный обзор литературы показал, что в настоящий момент наиболее часто используемыми методами оценки *in silico* параметров ADMET фармакологически активных соединений являются метод случайного леса и метод опорных векторов. Собственные результаты по компьютерному моделированию консенсусным методом в системе IT Microcosm и методом искусственных нейронных сетей двенадцати токсикологических и фармакокинетических характеристик химических соединений показали, что в сравнении с данными литературы IT Microcosm лучше прогнозирует два токсикологических свойства: канцерогенность и проникновение через гематоэнцефалический барьер (точность прогноза до 93,4%); а нейросетевые модели – четыре токсикологических свойства: острую токсичность, канцерогенность, генотоксичность, проникновение через гематоэнцефалический барьер (точность прогноза до 93,8%) и три фармакокинетических свойства: всасываемость в ЖКТ, объем распределения и печеночный клиренс (коэффициент детерминации 0,825).

Выводы. Наиболее перспективным и практически значимым направлением в разработке систем *in silico* прогноза ADMET-характеристик новых лекарственных препаратов является использование технологии искусственных нейронных сетей.

Ключевые слова: ADMET; компьютерный прогноз; *in silico*; токсикологические параметры; фармакокинетические параметры; химические соединения; лекарственные вещества; консенсусный метод; искусственные нейронные сети

Для цитирования: Васильев П.М., Голубева А.В., Королева А.Р., Перфильев М.А., Кочетков А.Н. Прогноз *in silico* токсикологических и фармакокинетических характеристик лекарственных соединений. *Безопасность и риск фармакотерапии*. 2023;11(4):390–408. <https://doi.org/10.30895/2312-7821-2023-11-4-390-408>

Финансирование. Работа выполнена без спонсорской поддержки.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов, требующего раскрытия в данной статье.

In Silico Prediction of Toxicological and Pharmacokinetic Characteristics of Medicinal Compounds

Pavel M. Vassiliev✉, Arina V. Golubeva, Anastasia R. Koroleva, Maksim A. Perfilov, Andrey N. Kochetkov

Volgograd State Medical University,
1 Pavshikh Bortsov Sq., Volgograd 400131, Russian Federation

✉ Corresponding author: Pavel M. Vassiliev pvassiliev@mail.ru

ABSTRACT

Scientific relevance. Studies of the toxicological and pharmacokinetic properties of medicinal compounds are a crucial stage of preclinical research; unsatisfactory results may invalidate further drug development. Therefore, the development of *in silico* methods for a preliminary pre-experimental assessment of toxicological and pharmacokinetic properties is a relevant and crucial task.

Aim. The study aimed to review current approaches to *in silico* prediction of the absorption, distribution, metabolism, excretion, and toxicity (ADMET) parameters of pharmacologically active compounds, in particular, the most important toxicological and pharmacokinetic parameters, and to present the results of the authors' own research in this area.

Discussion. According to the review of models for predicting the toxicological properties of chemical compounds (acute toxicity, carcinogenicity, mutagenicity, genotoxicity, endocrine toxicity, cytotoxicity, cardiotoxicity, hepatotoxicity, and immunotoxicity), the accuracy of predictions ranged from 74.0 to 98.0%. According to the review of models for predicting the pharmacokinetic properties of chemical compounds (gastrointestinal absorption; oral bioavailability; volume of distribution; total, renal, and hepatic clearance; and half-life), the coefficient of determination for the predictions ranged from 0.265 to 0.920. The literature review showed that the most widely used methods for *in silico* assessment of the ADMET parameters of pharmacologically active compounds included the random forest method and the support vector machines method. The authors compared the literature data with the results they obtained by modelling 12 toxicological and pharmacokinetic properties of chemical compounds using the consensus method in the IT Microcosm system and artificial neural networks. IT Microcosm outperformed the models described in the literature in terms of predicting 2 toxicological properties, including carcinogenicity and blood–brain barrier penetration (the prediction accuracy reached 93.4%). Neural network models were superior in predicting 4 toxicological properties, including acute toxicity, carcinogenicity, genotoxicity, and blood–brain barrier penetration (the prediction accuracy reached 93.8%). In addition, neural network models were better in predicting 3 pharmacokinetic properties, including gastrointestinal absorption, volume of distribution, and hepatic clearance (the coefficient of determination reached 0.825).

Conclusions. The data obtained suggest that artificial neural networks are the most promising and practically significant direction for the development of *in silico* systems for predicting the ADMET characteristics of new medicinal products.

Keywords: ADMET; computer prediction; *in silico*; toxicological parameters; pharmacokinetic parameters; chemical compounds; medicinal compounds; consensus method; artificial neural networks

For citation: Vassiliev P.M., Golubeva A.V., Koroleva A.R., Perfilev M.A., Kochetkov A.N. *In silico* prediction of toxicological and pharmacokinetic characteristics of medicinal compounds. *Safety and Risk of Pharmacotherapy*. 2023;11(4):390–408. <https://doi.org/10.30895/2312-7821-2023-11-4-390-408>

Funding. The study was performed without external funding.

Disclosure. The authors declare having no conflict of interest requiring disclosure in this article.

Введение

Под аббревиатурой ADMET понимается совокупность характеристик фармакологически активного вещества, отражающих такие его свойства, как абсорбция, распределение, метаболизм, выведение и токсичность (Absorption, Distribution, Metabolism, Excretion, Toxicity) [1]. Изучение токсикологических свойств лекарственных препаратов является одним из самых важных этапов доклинических исследований [2]. При этом изучение фармакокинетических свойств соединений не менее значимо, поскольку во многих случаях определяет также и токсические эффекты новых молекул [3].

В настоящее время компьютерные методы прогноза биологической активности химических соединений, обозначаемые общим термином *in silico*, используются на одном из первоначальных этапов поиска новых веществ, обладающих фармакологической активностью [4]. Особый интерес вызывает возможность оценки *in silico* различных параметров ADMET новых молекул, поскольку это позволяет прогнозировать их ожидаемые токсикологический и фармакокинетический профили [5]. Вклад компьютерных методов в оценку показателей ADMET отражен в законе, принятом в США в 2022 г., который позволяет использовать методы *in silico* в качестве предварительных тестов в доклинических исследованиях новых лекарственных веществ¹. Как обязательный элемент методы *in silico* имплементированы в реализуемый в США масштабный токсикологический проект «Токсикология в XXI веке» (Toxicology in the 21st Century, Tox21)².

Цель работы – обзор современных подходов к компьютерному прогнозу ADMET свойств лекарственных соединений, в особенности наиболее важных токсикологических и фармакокинетических показателей, и изложение результатов собственных исследований в этой области.

Методы компьютерного прогноза ADMET-характеристик

Проблема построения зависимостей «химическая структура – биологическая активность»

является одной из актуальных фундаментальных проблем современных медико-биологических и химических наук. В общем виде это выражается соотношением:

$$A=F(S), \quad (1)$$

где A – показатель биологической активности (непрерывный или дискретный); F – математическая функция, аппроксимирующая зависимость; S – многомерное параметрическое описание химической структуры.

Если показатель A является количественной переменной (например, LD_{50}), для построения функции F используется регрессионный подход; если показатель A является дискретным (например, наличие/отсутствие канцерогенности), для построения функции F используются методы классификации [6]. Описание химической структуры S может быть представлено топологическими дескрипторами [7], структурными дескрипторами [8], физико-химическими параметрами [9], квантово-химическими параметрами [10], показателями энергий связи между исследуемой молекулой и целевым белком [11].

В настоящее время в оценке *in silico* свойств ADMET химических соединений используется большое количество разнообразных методов машинного обучения – как регрессионных, так и классификационных [12]. Поэтому целесообразно рассмотреть наиболее апробированные и часто используемые подходы.

Регрессионные методы. К числу наиболее популярных компьютерных способов оценки ADMET-характеристик (особенно фармакокинетических) относится построение количественных соотношений структура–активность QSAR (Quantitative Structure – Activity Relationship). Вид зависимости QSAR F в формуле (1) определяется на основе метода наименьших квадратов, реализованного в форме различных статистических или эвристических моделей регрессии [13]. Наиболее часто используемыми являются множественная линейная регрессия MLR (Multiple Linear Regression), частичный метод наименьших квадратов PLS (Partial Least

¹ S.5002. FDA Modernization Act 2.0 2D Session of 117th Congress. 29.12.2022. <https://www.congress.gov/117/bills/s5002/BILLS-117s5002es.pdf>

² National Toxicology Program: Toxicology in the 21st Century (Tox21). <https://ntp.niehs.nih.gov/whatwestudy/tox21/>

Squares) и регрессия на главных компонентах PCR (Principal Components Regression). Метод MLR представляет собой обобщение простой линейной регрессии для нескольких независимых переменных; метод PLS применяется в ситуациях, когда дисперсия остатков не постоянна; метод PCR реализует MLR на ненаблюдаемых главных компонентах, объясняющих максимальный уровень дисперсии. Уравнение регрессии рассчитывается на основе обучающей выборки, включающей соединения с известными количественными значениями биологической активности и молекулярных дескрипторов.

Точность регрессионных моделей во многом определяется точностью представления биологических данных, но также зависит от выбора молекулярных дескрипторов и используемого метода регрессии. Следует особо подчеркнуть, что подавляющее большинство прогнозируемых показателей ADMET характеризуется высокой вариабельностью, поэтому экстраполяция регрессионных QSAR зависимостей на новые химические классы или биологические объекты требует осторожности и тщательной интерпретации результатов [14].

Классификационные методы. Эти методы обычно используются для решения задачи классификации объектов [6] и широко используются для прогноза характеристик ADMET. В отличие от регрессионных методов – обучающая выборка содержит не количественные, а дискретные значения биологической активности, чаще всего бинарные (активно / не активно). Такое представление дает возможность нивелировать ошибки, связанные с высокой вариабельностью биологических данных, что существенно повышает достоверность получаемых классификационных моделей. Описание химической структуры соединений может быть представлено в виде набора различных молекулярных дескрипторов.

В этих граничных условиях для построения классификационных зависимостей ADMET используются методы «обучения с учителем», наиболее известными из которых являются следующие.

1. Линейный дискриминатный анализ LDA (Linear Discriminant Analysis) выполняет построение линейной разделяющей гиперплоскости на основе функций плотности распределения вероятности двух классов [15].

2. «Наивный» Байесовский классификатор NBC (Naive Bayes Classifier) основан на вычислении по формуле Байеса максимальной условной

вероятности принадлежности прогнозируемого объекта к одному из двух классов [16].

3. Метод k-ближайших соседей kNN (k-Nearest Neighbours) позволяет отнести объект к тому классу, в котором находится k наиболее сходных с ним представителей обучающей выборки [17].

4. Метод опорных векторов SVM (Support Vector Machine) производит нелинейное преобразование пространства описания обучающей выборки в пространство большей размерности таким образом, чтобы объекты двух классов стали линейно разделимы путем максимизации расстояния (зазора) между гиперплоскостью и ближайшими к ней точками из каждого класса (опорными векторами) [18].

5. Случайный лес RF (Random Forest) – использует для классификации ансамбль деревьев решений, каждое из которых обучено на случайном подмножестве данных и случайном подмножестве входных признаков (бутстрепных выборках) [19, 20].

6. Искусственные нейронные сети ANN (Artificial Neural Networks) – эмулируют сети нервных клеток живых организмов и представляют собой множество переменных (нейронов), организованных в несколько различных уровней (входные, скрытые, выходные слои) и объединенных между собой связями (синапсами). Обучение ANN происходит путем итеративной аппроксимации весов синапсов в соответствии с минимизацией ошибки классификации. В настоящее время все многообразие ANN различной архитектуры принято определять как методы искусственного интеллекта, которые интенсивно разрабатываются. Особое внимание уделяется сверточным нейросетям CNN (Convolutional Neural Network) и нейросетям глубокого обучения DNN (Deep Neural Network) [21–23].

В целом, выбор метода машинного обучения зависит от конкретного прогнозируемого свойства ADMET, доступных данных и доступных вычислительных ресурсов.

Молекулярный докинг. Этот метод является одним из самых используемых подходов к виртуальному поиску фармакологически активных соединений, он основан на вычислении энергии связывания молекулы химического соединения с белком-мишенью [24]. Докинг включает построение 3D-структур белка и лиганда, а затем моделирование их взаимодействия методами молекулярной механики. Для выполнения докинга разработано большое число компьютерных программ, наиболее известными из которых являются AutoDock, GOLD, Glide, SurflexDock,

каждая имеет свои преимущества и ограничения [25]. В качестве примера наиболее часто и успешно применяемой можно привести программу AutoDock Vina [26].

Докинг используется для прогноза характеристик ADMET фармакологически активных соединений путем оценки их аффинности к различным релевантным белкам-мишеням – например, кардиотоксичности через докинг в hERG канал [27]. Этот метод, однако, не может быть применен для оценки неспецифических видов активности, каковыми являются большинство свойств ADMET. Кроме того, из-за недостаточной точности и надежности оценочных функций, сформированных на основе принципов молекулярной механики, полученные в результате докинга показатели аффинности далеко не всегда соответствуют экспериментальным данным [28].

Примеры ресурсов для *in silico* прогноза ADMET.

Описанные выше подходы использованы разными авторами для построения решающих правил, позволяющих выполнять *in silico* прогноз различных показателей ADMET. Эти многочисленные прогностические модели обычно объединяются в один интегральный продукт – компьютерную программу или on-line ресурс. Наиболее известными из таких ADMET-прогностических инструментов являются следующие.

I. Программа DruLiTo³ – популярный инструмент оценки лекарственного подобия (drug-likeness) химических соединений [29]. Использует в качестве входной информации *.mol или *.sdf файлы. Рассчитывает 14 видов молекулярных дескрипторов, включая физико-химические параметры, такие как липофильность, молекулярная рефракция, площадь полярной поверхности и др. Эти данные используются для оценки лекарственного подобия по различным фильтрам: по правилу Липински [30], MDDR (Modern Drug Data Report) [31], Вебера [32], Гоуза [33], CMC-50 (Comprehensive Medicinal Chemistry) [33], BBB (Blood-Brain Barrier) [34], и вычисляются два количественных показателя лекарственного подобия (Quantitative Estimate of Drug-likeness, QED), которые могут принимать значения от нуля до единицы [35]. Кроме того, определяется число молекулярных фрагментов SAlerts (structure alerts) [36], которые могут

отвечать за потенциальную токсичность соединения. Для расчета применяемых в этих фильтрах молекулярных дескрипторов используются методы регрессии и NBC.

II. Платформа SwissADME⁴ позволяет прогнозировать десять ADME параметров, преимущественно для оценки лекарственного подобия и фармакокинетических свойств [37]. Химическая структура вводится в виде строки SMILES (Simplified Molecular Input Line Entry System – система упрощенного представления молекул в строке ввода). Для описания структуры химического соединения применяются фрагментные и физико-химические дескрипторы, рассчитанные программой OpenBabel [38]. Прогноз осуществляется с использованием методов MLR и SVM.

III. Программа PASS⁵ – известная система прогноза спектра различных видов биологической активности [16]. В качестве входных данных также использует *.mol и *.sdf файлы. Прогнозирует наличие/отсутствие более 7,5 тыс. видов биологической активности, в том числе более 500 побочных и токсических эффектов. Для описания структуры химического соединения используются специальные фрагментные MNA-дескрипторы (Multilevel Neighbourhoods of Atoms). Прогноз осуществляется с использованием метода NBC.

IV. Платформа ProTox-II⁶ – это веб-инструмент, который позволяет прогнозировать наличие/отсутствие 33 видов токсических свойств, в том числе воздействие на токсикологические биомишени и сигнальные пути Tox21 [39]. В качестве входных данных используется SMILES описание химической структуры. Для описания структуры химического соединения применяются смешанные фрагментные и молекулярные «отпечатки» (fingerprints). Прогноз осуществляется с использованием методов NBC, kNN, RF, SVM.

V. Платформа pkCSM⁷ – прогнозирует 28 ADMET-параметров, преимущественно фармакокинетических, 12 количественно и 16 категориально [40]. Химическая структура вводится в виде SMILES нотации. Для рабочего описания применяются основанные на графах фрагментные дескрипторы. Прогноз осуществляется с использованием метода MLR, а для бинарных показателей – одной из модификаций ANN.

³ https://niper.gov.in/pi_dev_tools/DruLiToWeb/DruLiTo_index.html

⁴ <http://www.swissadme.ch/>

⁵ <http://www.way2drug.com/passonline/>

⁶ https://tox-new.charite.de/protox_II/

⁷ <https://biosig.lab.uq.edu.au/pkcsml/>

VI. Платформа ADMETlab⁸ – позволяет оценивать различные показатели лекарственного подобия и 31 ADMET-параметр, преимущественно в категориальной форме, по шестибальной шкале от трех плюсов до трех минусов; 12 из этих характеристик отражают воздействие на токсикологические биомеханизмы и сигнальные пути Tox21 [41]. В качестве входных данных используется SMILES описание. Для описания химической структуры применяются 11 типов фрагментных и молекулярных дескрипторов (всего 403 вида). Расчет решающих правил и прогноз производятся методами PLS, NBC, RF, SVM, некоторые из них в модификации.

VII. Платформа admetSAR⁹ – предоставляет пользователю больше всего возможностей из всех перечисленных ресурсов. Ее использование позволяет оценивать 50 ADMET-показателей, вычисляемых в бинарной форме (наличие/отсутствие), с указанием вероятности оценки [42]. Химическая структура вводится в виде SMILES. Для рабочего описания применяются смешанные фрагментные и молекулярные «отпечатки» (fingerprints). Прогноз осуществляется с использованием методов kNN, RF, SVM, CNN.

Рассмотрим подробнее применение описанных методов для прогноза наиболее существенных токсикологических и фармакокинетических показателей.

Методы компьютерного прогноза токсикологических ADMET-характеристик

Острая токсичность. Исследования острой токсичности являются первоочередными для оценки токсических свойств лекарственных веществ *in vivo*¹⁰. За последнее время в ряде стран, например в США¹¹, были предприняты значительные усилия по дополнению испытаний на животных альтернативными методами *in vitro* и *in silico*. Накопление в открытом доступе результатов экспериментальных исследований позволило использовать эти данные для компьютерного прогноза острой токсичности.

В работе [43] представлены мультиклассовые модели для прогноза уровней токсичности по GHS¹². Для описания молекулярных структур применены фрагментные дескрипторы ISIDA (*In silico* Design and Data Analysis – проектирование и анализ данных *in silico*) [44]. В машинном обучении использовались методы NBC, SVM, RF. Генеральная обучающая выборка содержала значения LD₅₀ для крыс при пероральном пути введения препарата в организм для ~15 000 веществ из различных доступных баз данных. Максимальная сбалансированная точность прогноза (полусумма значений чувствительности *Sens* и специфичности *Spec*) составила BA=74%.

В исследовании [45] сопоставляли эффективность нескольких методов машинного обучения в сочетании с различными типами дескрипторов на основе набора данных о ~87 000 соединений из регистра токсических эффектов химических соединений (Registry of Toxic Effects of Chemical Substances, RTECS¹³). Показано, что использование многозадачных DNN обеспечивает значительно более высокую точность прогнозирования острой токсичности в сравнении с моделями с одним выходом на основе DNN, RF и kNN.

Канцерогенность. Механизмы формирования канцерогенной активности весьма многочисленны и мало изучены. Модели *in silico* для прогноза канцерогенных свойств пока не получили всеобщего признания регулирующих органов, в основном из-за не очень высокой прогностической точности и ограничений в области применения [46].

По данным о 852 соединениях, канцерогенов и не канцерогенов, с помощью пошагового MLR построена смешанная классификационная модель на основе молекулярных дескрипторов и показателей мутагенности [47]. На внешнем тестовом наборе получена точность прогнозирования *Ac* = 80,1%, чувствительность *Sens* = 77,1% и специфичность *Spec* = 82,2%.

Мутагенность. Традиционно первый этап оценки мутагенности включает эксперименты *in vitro*, в частности метод регистрации

⁸ <https://admetmesh.scbdd.com/>

⁹ <http://lmmd.ecust.edu.cn/admetSar1>

¹⁰ Руководство по проведению доклинических исследований лекарственных средств. Ч. 1. М: Гриф и К; 2012. Раздел I. Доклинические исследования безопасности лекарственных средств.

¹¹ S.5002. FDA Modernization Act 2.0 2D Session of 117th Congress. 29.12.2022. <https://www.congress.gov/117/bills/s5002/BILLS-117s5002es.pdf>

¹² Globally Harmonized System of Classification and Labelling of Chemicals (GHS, Rev. 10, 2023). Chapter 3.1: Acute toxicity. <https://unece.org/info/Transport/Dangerous-Goods/pub/381478>

¹³ Registry of Toxic Effects Of Chemical Substances (RTECS®): comprehensive guide to the RTECS®. DHHS (NIOSH) Publication Number 97-119. <https://www.cdc.gov/niosh/docs/97-119/default.html>

обратных генных мутаций на бактериях — тест Эймса. Использование в тесте специальных бактериальных штаммов *Salmonella typhimurium* и *Escherichia coli* позволяет выявлять мутагены разного механизма действия. По некоторым оценкам, межлабораторная воспроизводимость данных теста на сальмонеллах составляет 85% [48], что свидетельствует об ограничении данного *in vitro* метода. Поэтому поднимается вопрос о необходимости разработки качественной модели для предсказания мутагенности *in silico* вместо тестов *in vitro*.

В работе [49] была разработана методика *in silico* для прогнозирования химической мутагенности на основе базы данных из 7617 соединений (4252 мутагена и 3365 немутагена). Были построены модели с использованием методов SVM, RF, ANN, kNN, NBC. Для описания химической структуры применяли различные виды молекулярных «отпечатков» (fingerprints). Эффективность оценивали пятикратной перекрестной проверкой и на внешнем тестовом наборе из 831 соединений. Точность прогноза Acc для пяти лучших моделей в пятикратном перекрестном контроле составила от 80,8 до 84,1%. На внешнем наборе точность Acc изменялась от 90,4 до 98,0%, что превосходит точность платформы ToxTree [50].

Генотоксичность. Под генотоксичностью химических соединений понимается их способность взаимодействовать с ДНК, повреждать ее и нарушать работу генетического аппарата клеток [51]. Чтобы адекватно оценить наличие/отсутствие генотоксического потенциала вещества, необходимо исследовать его способность индуцировать генные мутации и повреждение хромосом. В руководстве Европейского агентства по лекарственным средствам (European Medicines Agency, EMA) по тестированию на генотоксичность и интерпретации результатов для лекарственных препаратов, предназначенных для использования человеком¹⁴, рекомендовано сочетать бактериальные тесты на мутагенность с тестами на генотоксичность *in vitro* и/или *in vivo* на клетках млекопитающих. В дополнение к этим тестам применяют методы *in silico* [52, 53].

Описано несколько моделей *in silico* для прогнозирования генотоксичности, определяемой

в микроядерном анализе *in vitro* [54]. Лучшей признана модель [55], полученная на подструктурных фрагментах, рассчитанных с помощью платформы SARpy¹⁵. Было выявлено более 100 структурных признаков, включая связанные с отсутствием генотоксической активности. Точность модели составила Acc=83%, чувствительность Sens=85%, специфичность Spec=80%.

Эндокринологичность. Эндокринными разрушителями называют соединения, которые, попадая в организм человека, могут вызвать расстройства эндокринных функций, а также повлиять на его репродуктивное здоровье. Перестройка эндокринной системы может быть вызвана вмешательством в процессы синтеза, секреции, транспорта, связывания, действия и элиминации естественных гормонов в организме¹⁶. Химические соединения могут проявлять себя как эндокринные разрушители, имитируя гормоноподобные эффекты, усиливая или ингибируя при этом действие естественных гормонов. Механизмы действия эндокринных разрушителей делятся на прямые и косвенные. Прямой механизм реализуется путем связывания вещества с ядерными рецепторами, в результате чего нарушаются их функции путем чрезмерной активации или подавления. При непосредственном механизме действия эндокринные разрушители могут изменять экспрессию этих рецепторов и влиять на передачу сигнала, а также на эпигенетические механизмы, лежащие в их основе [56, 57]. На данный момент значительное число лекарственных препаратов не испытано на эндокринологичность, поэтому применение методик *in silico* для оценки этого свойства является весьма актуальным [58].

Метод DNN был использован при прогнозировании эндокринных нарушений, связанных с агонистическим действием на прогестероновые рецепторы, на основе данных программы Tox21 [59]. Путем вращения построенных 3D-моделей химических соединений с заданным угловым шагом для каждой структуры создавали набор молекулярных снимков — изображений в виде файлов *.png размером 256×256 пикселей. Эти изображения затем обрабатывали с помощью ресурса DeepSNAP¹⁷, позволяющего проводить построение сетей глубокого обучения на основе графических образов. Точность

¹⁴ ICH guideline S2 (R1) on genotoxicity testing and data interpretation for pharmaceuticals intended for human use.

¹⁵ National Geospatial-Intelligence Agency: NGAgeInt/SARpy. <https://github.com/ngageoint/sarpy>

¹⁶ CSTE Opinion on human and wildlife health effects of endocrine disrupting chemicals, with emphasis on wildlife and on ecotoxicology test methods. https://ec.europa.eu/health/archive/ph_risk/committees/sct/documents/out37_en.pdf

¹⁷ DeepSNAP. <https://snap.stanford.edu/deepsnap/>

найденной таким образом модели при тестировании на внешней выборке составила $Acc=97,6\%$, коэффициент корреляции Мэтью превысил 0,8.

В работе [60] с использованием структурных отпечатков PubChem [61] и OpenBabel [38] на больших обучающих выборках из проектов CERAPP (Collaborative Estrogen Receptor Activity Prediction Project) и CoMPARA (Collaborative Modeling Project for Androgen Receptor Activity) построены двенадцать бинарных и многоклассовых моделей для прогноза агонизма и антагонизма к рецепторам эстрогена и андрогенов. Сбалансированная точность некоторых из них достигала $BA=93\%$, при покрытии $Coverage=89\%$.

Цитотоксичность. По определению, цитотоксичность есть способность вещества оказывать на клетку токсическое действие. По данным литературы, наблюдается статистически достоверная зависимость показателей острой токсичности *in vivo* от результатов *in vitro* тестов на цитотоксичность, без цитотоксических исследований невозможно оценить острую токсичность соединений [62].

В работе [63] с помощью платформы AdmetSAR выполнен прогноз величины цитотоксичности у фенольных соединений. Комбинированное описание структуры включало молекулярные и квантово-химические дескрипторы, рассчитанные с использованием программы HyperChem 7¹⁸. Регрессионная зависимость вычислена методом MLR. Коэффициент детерминации составил $R^2=0,93$.

Кардиотоксичность. Ген *hERG* кодирует калиевый канал сердца Kv11.1, блокирование которого удлиняет интервал QT и может привести к фатальной кардиотоксичности фармакологически активных соединений.

В работе [64] была создана выборка из 7889 соединений с верифицированными экспериментальными данными по гену *hERG*. Показано, что модели на основе многозадачной DNN превосходят по точности прогноза *hERG*-активности модели, построенные с помощью однозадачной DNN (с единственным выходным нейроном), NBC, SVM, RF и графовой CNN. На тестовом наборе площадь под кривой в ROC-анализе для лучшей модели deepHERG составила $AUC_{ROC}=96,7\%$. Более того, из 1824 одобренных FDA препаратов 29,6% по результатам вычислений deepHERG идентифицированы как соединения с потенциальной ингибирующей *hERG* активностью.

Гепатотоксичность. Печень играет центральную роль в метаболизме ксенобиотиков. Некоторые лекарственные средства при передозировке или даже в терапевтических концентрациях могут приводить к гепатотоксическим эффектам. Лекарственно-индуцированное повреждение печени в настоящее время является серьезным основанием для ограничений в использовании ранее уже одобренных препаратов [65].

Одним из механизмов гепатотоксичности лекарственных препаратов является их трансформация в печени с образованием метаболитов, обладающих высокой реакционной способностью, которые и вызывают повреждение этого органа [66]. Достоверное определение гепатотоксической активности является достаточно сложной задачей, поскольку данные, полученные на животных, плохо экстраполируются на человека и не все экспериментальные методы достаточно валидированы [67]. В связи с этим предварительная *in silico* оценка гепатотоксичности сегодня рассматривается как эффективная стратегия ускорения темпов создания безопасных лекарственных средств.

В исследовании [65] представлена ансамблевая модель для прогнозирования риска гепатотоксичности ксенобиотиков, включающая использование восьми классификаторов и молекулярных Marvin-дескрипторов и построенная на основе выборки из 1254 известных экспериментально изученных соединений. Для проверки точности были использованы три внешних тестовых набора и большой набор данных для веществ, не проявляющих гепатотоксичности. Прогностические показатели модели составили: точность $Acc=78,3\%$, чувствительность $Sens=81,8\%$, специфичность $Spec=74,8\%$, $AUC_{ROC}=85,9\%$.

В статье [68] описаны *in silico* модели гепатотоксичности, построенные методами RF и SVM, с применением битовых молекулярных отпечатков. Точность прогноза составила: в перекрестном контроле $Acc=73,8\%$ и $Acc=72,6\%$, $AUC_{ROC}=79,1\%$ и $AUC_{ROC}=76,8\%$; при независимом тестировании $Acc=60,1\%$ и $Acc=61,1\%$ – для RF и SVM соответственно. Полученные модели были использованы для прогноза гепатотоксической активности 13966 растительных соединений.

Иммунотоксичность. К токсическим эффектам в отношении иммунной системы можно отнести развитие аллергии, аутоиммунизацию, увеличение частоты или тяжести инфекционных

¹⁸ Hypercube, Inc. <http://www.hypercubeusa.com/>

и онкологических заболеваний. В настоящее время доклиническое тестирование на иммуно-токсичность проводится *in vivo* в соответствии с руководством S8 ICH¹⁹. Перенос таких данных с животных на человека затруднителен, поэтому дополнительно используются методы *in vitro*, такие как иммуноферментный анализ, проточная цитометрия, определение экспрессии цитокинов/хемокинов, секвенирование и др. Все перечисленные методы являются весьма дорогостоящими и времязатратными, поэтому сейчас они предваряются методами *in silico*.

Например, в работе [69] выполнен прогноз иммунотоксичности наномедицинских препаратов с вычислительным определением сайтов связывания toll-подобных рецепторов и последующим докингом в них углеродных наночастиц.

Приведенная выше обзорная информация показывает, что в зависимости от вида свойств, качества, разнообразия и степени подготовки исходных данных и метода построения решающих правил точность прогноза *in silico* основных токсикологических ADMET-характеристик может варьироваться от 71% (острая токсичность [43]) до 98% (мутагенность [49]).

Методы компьютерного прогноза фармакокинетических ADMET-характеристик

Всасываемость в желудочно-кишечном тракте.

Для успешной доставки к терапевтической мишени лекарственных препаратов, вводимых перорально, они должны обладать хорошей растворимостью и всасываемостью в желудочно-кишечном тракте человека (Human Intestinal Absorption, HIA) [70]. Поэтому создание компьютерных моделей для прогноза этого параметра представляет собой достаточно актуальную задачу.

Так, в исследовании [71], используя обучающую выборку, содержащую 1516 соединений, действие которых на HIA было экспериментально изучено, методом RF построена классификационная модель, точность которой в пятикратном перекрестном контроле составила 83%. Другая модель [72], основанная на методе RF и на обучающей выборке из 970 соединений, показала точность прогноза на тестовом наборе 86,6%.

В работе [73] исследована применимость для оценки HIA шести классификационных методов: SVM, kNN, ANN в двух модификациях, PLS и LDA. Обучающая выборка включала 745 соединений, тестовая — 497 соединений. Точность

прогноза составила 91,5, 88,3, 84,3, 86,5, 79,1 и 80,1% соответственно. Как видно, самым эффективным оказался метод SVM.

Биодоступность при пероральном приеме (oral bioavailability, OB) — это доля введенного внутрь лекарственного препарата, которая достигает системного кровообращения [74].

В статье [75] описана интегрированная модель прогноза OB, построенная с помощью методов SVM и PLS на трех наборах структурных параметров с использованием двух фармакокинетических моделей и включающая девять регрессионных решающих правил. Прогностическая точность модели составила $Q^2=0,50$.

В работе [71] по базе данных из 1822 соединений методом RF создана классификационная модель, точность которой достигла $Acc=71\%$.

В исследовании [76] по выборке из 410 химических соединений построены модели на основе методов PLS, SVM, ANN, RF и k-NN. На тестовом наборе точность прогноза составила 20,6, 23,3, 15,1, 21,9 и 19,2% соответственно, что нельзя признать удовлетворительным.

Кажущийся объем распределения. После всасывания лекарственные препараты попадают в кровотоки и распределяются в органах и тканях организма. Этот процесс характеризуется кажущимся объемом распределения V_D (volume of distribution), который представляет собой отношение количества введенного лекарственного средства к его концентрации в плазме. Данный параметр отражает предположительный объем жидкости, который потребовался бы в стационарном состоянии для равномерного распределения в нем вещества [77].

В работе [78] проводили обучение методом RF с различным числом классификационных деревьев и использованием четырех наборов дескрипторов, описывающих химическую структуру 1301 соединений. Для лучшей регрессионной модели коэффициент детерминации составил $R^2=0,52$.

В исследовании [79] методом RF на обучающей и тестовой выборках из 1090 и 213 соединений, описанных шестью типами дескрипторов, построена модель для прогноза V_D . Показано, что для большинства соединений значение этого показателя можно предсказать в пределах двухкратного диапазона среднегеометрической ошибки модели.

Клиренс. Одним из ключевых параметров количественной характеристики элиминации

¹⁹ ICH S8 Immunotoxicity studies for human pharmaceuticals — Scientific guidelines. CHMP/167235/2004. ICH; 2006.

веществ является клиренс. Общий клиренс Cl (clearance) отражает объем плазмы крови, освобождающийся от лекарственного средства в единицу времени; он определяется клиренсом всех отдельных органов элиминации и включает главным образом клиренс печени и почек [74].

Почечный клиренс Cl_r (renal clearance) зависит от процессов фильтрации, секреции и реабсорбции. Он определяется отношением скорости экскреции к средней концентрации лекарственного средства в крови в соответствующий период времени или отношением кумулятивной экскреции к площади под кривой «концентрация–время» (AUC) лекарственного средства в крови [74]. Печеночный клиренс Cl_H (hepatic clearance) связан с такими процессами, как биотрансформация лекарственного средства в гепатоцитах и его последующее выделение с желчью.

Общий клиренс. В работе [80] исходная выборка содержала информацию о 1114 химических соединениях. Регрессионные модели, полученные методами RF и SVM, показали наилучшую точность прогнозирования клиренса среди восьми моделей, построенных различными методами машинного обучения.

В статье [81] для построения регрессионных моделей кроме методов SVM и RF использовали два метода повышения градиента GBM (Gradient Boosting Machine). Обучающая выборка содержала 1352 соединения. Точность оценивали с помощью трех видов перекрестной проверки. Лучшую точность показала RF модель, на тестовой выборке коэффициент детерминации составил $R^2=0,875$.

Почечный клиренс. В исследовании [82] на обучающей выборке из 636 химических соединений с использованием методов MLR и RF для прогноза Cl_r построены общая модель и частные модели для различных механизмов выведения. Коэффициент детерминации глобальной модели, построенной по методу RF, составил $R^2=0,20$, а по методу MLR – $R^2=0,17$, такую точность можно охарактеризовать как неудовлетворительную.

В работе [83] методами RF, SVM, PLS и ANN с использованием набора данных о структуре 401 соединения были построены классификационные и регрессионные модели для прогноза Cl_r . Классификационные модели прогнозировали механизм почечной экскреции веществ: чистая реабсорбция, чистая секреция

и промежуточный тип. Регрессионные модели осуществляли количественную оценку значения Cl_r соединений каждого типа экскреции. Лучшие регрессионные модели были получены с помощью метода RF, коэффициент детерминации составил $R^2=0,66$, $0,62$ и $0,92$ для трех типов экскреции соответственно. Точность прогноза на тестовой выборке для лучшей классификационной модели, полученной с помощью метода RF, была равна $Acc=32\%$.

Печеночный клиренс. В статье [84] входной набор включал 1600 соединений обучающей выборки и две независимых тестовых выборки из 99 и 2113 соединений (получены из ToxCast²⁰ и ChEMBL²¹). Для построения модели прогноза уровня Cl_H использовали метод RF применительно к четырем разным дескрипторным описаниям. Точность для лучшей модели составила на обучающей выборке $Acc=58,5\%$, на независимых тестовых выборках из ChEMBL $Acc=48,9\%$ и из ToxCast $Acc=69,4\%$.

В исследовании [85] для прогноза количественных значений Cl_H использовали метод RF в модификации для регрессии. На тестовом наборе коэффициент детерминации составил $R^2=0,265$.

Период полувыведения. Период полувыведения $T_{1/2}$ характеризует время, в течение которого концентрация лекарственного средства снижается вдвое. Важность доэкспериментального исследования данного параметра определяется его клиническим значением – он используется для подбора оптимальных терапевтических дозировок лекарственных препаратов и интервалов их введения для создания стабильной концентрации препарата в условиях многократного применения [74].

В работе [76] для прогноза $T_{1/2}$ были построены многозадачные классификационные модели на основе PLS, SVM, ANN, RF и k-NN. Исходный набор данных содержал 969 соединений. Точность прогноза Acc составила 56, 56, 52, 66 и 66% соответственно.

В исследовании [81] для прогноза $T_{1/2}$ использовалась база данных из 1352 соединений. Применены методы машинного обучения SVM, RF, две модификации GBM. Для лучшей регрессионной модели, построенной методом RF, коэффициент детерминации составил $R^2=0,832$.

Приведенная информация показывает, что в зависимости от вида свойств, качества, разнообразия и степени подготовки исходных

²⁰ Exploring ToxCast Data. <https://www.epa.gov/chemical-research/exploring-toxcast-data>

²¹ ChEMBL. <https://www.ebi.ac.uk/chembl/>

данных и метода построения решающих правил точность прогноза *in silico* основных фармакокинетических ADMET-характеристик для лучших классификационных моделей может варьировать от $Acc=32\%$ (почечный клиренс [83]) до $Acc=91,5\%$ (всасываемость в желудочно-кишечном тракте (ЖКТ) [73]), а коэффициент детерминации для регрессионных моделей — от $R^2=0,265$ (печеночный клиренс [85]) до $R^2=0,92$ (почечный клиренс [83]).

Прогноз ADMET-характеристик консенсусным методом и методом искусственных нейронных сетей

Анализ современного состояния проблемы прогноза *in silico* ADMET-характеристик химических соединений показал, что компьютерные методы широко и успешно применяются для вычислительной оценки указанных свойств. Однако следует подчеркнуть, что все ADMET-характеристики являются неспецифическими, механизмы их формирования связаны с множеством сложных физиологических, фармакокинетических, фармакодинамических и биохимических процессов и почти всегда не полностью изучены. Недостаток для многих ADMET-показателей верифицированной и унифицированной информации по структуре и активности экспериментально изученных соединений создает значительные трудности в формировании подходящих наборов данных для обучения моделей. В связи с этим в ряде случаев пока не достигнута удовлетворительная точность прогноза, что является предпосылкой для создания новых концептуальных методологий компьютерного моделирования ADMET-свойств фармакологически активных соединений.

Ниже представлены результаты наших собственных исследований: проведенного нами консенсусного и нейросетевого прогнозирования двенадцати ADMET-характеристик фармакологически активных соединений: токсикологических — острой токсичности, канцерогенности, мутагенности, генотоксичности, эндокринологической токсичности, проникновение через гематоэнцефалический барьер; фармакокинетических — всасываемости в желудочно-кишечном тракте, объема распределения,

клиренса общего, клиренса почечного, клиренса печеночного, периода полувыведения.

Консенсусный прогноз ADMET-характеристик. Методика консенсусного прогноза детально изложена в монографии [86], на применяемый для этого оригинальный программный комплекс IT Microcosm получено свидетельство о государственной регистрации²². В рамках консенсусного подхода IT Microcosm совместно с другими системами применялся для оценки острой токсичности [87] и канцерогенности [88]. Разработана методика интегральной оценки токсикологических [89] и фармакокинетических [90] ADMET-показателей в виде общего единого числового показателя.

В системе IT Microcosm химические структуры в виде *.sdf файлов транслируются в совокупность фрагментных подструктурных QL-дескрипторов 11 разных типов, детальное описание которых приведено в монографии [86]. Для расчета ансамбля из 44 прогнозных зависимостей используются 4 метода машинного обучения: Байеса, расстояния, k-ближайших соседей, локального распределения (NBC, примененный к соединениям обучающей выборки, наиболее сходным по структуре с прогнозируемым соединением). В рамках консервативной стратегии интегральное решающее правило формируется на основе общего невзвешенного консенсуса, при котором для прогнозируемого соединения методом простого голосования обобщаются 44 прогнозных оценки наличия/отсутствия данной активности.

Для токсикологических показателей исходная информация по структуре и активности была нами получена из следующих источников: острая токсичность LD_{50} для крыс перорально — ChemIDplus [91]; канцерогенность для человека — IARC [92], ChEMBL [93]; мутагенность в тесте Эймса — C. Xu et al. [49]; генотоксичность в тесте ДНК-комет — оригинальная база данных «Генотоксические соединения»²³; эндокринологическая токсичность для человека — DEDuCT [94]; проникновение через гематоэнцефалический барьер — Z. Wang et al. [95].

Точность прогноза Acc оценивалась методом скользящего контроля с исключением по одному (результаты приведены в *таблице 1*).

²² Васильев ПМ, Кочетков АН. ИТ «Микрокосм». Свидетельство о государственной регистрации программы для ЭВМ № 2011618547 (Россия); 2011.

²³ Васильев ПМ, Островский ОВ, Сиреканян АГ, Кочетков АН. Генотоксические соединения. Свидетельство о государственной регистрации базы данных № 2022621777 (Россия); 2022. <https://www1.fips.ru/ofpstorage/Doc/PrEVM/RUNWDB/000/002/022/621/777/2022621777-00001/DOCUMENT.PDF>

В проведенном исследовании при анализе фармакокинетических показателей использовались количественные данные. Система IT Microsoft предназначена для решения только классификационных задач, регрессионные методы в ней не применяются. Поэтому для фармакокинетических показателей консенсусные классификационные модели в IT Microsoft не рассчитывались.

Нейросетевое моделирование ADMET-характеристик. Эвристический характер технологии искусственных нейронных сетей ANN позволяет использовать этот подход для построения самых разнообразных моделей [21]. В соответствии с теоремой Колмогорова [96] с помощью двухслойной ANN может быть аппроксимирована зависимость любой сложности. Поэтому наиболее применяемой классической архитектурой ANN является двухслойный перцептрон. В сравнении с другими методами машинного обучения нейронная сеть с архитектурой многослойного перцептрона имеет следующие достоинства: 1) учитывает взаимное влияние каждого из входных нейронов (переменных) на все другие нейроны; 2) слабо чувствительна к наличию шума; 3) обрабатывает любые типы входных данных (бинарные, дискретные, непрерывные); 4) позволяет аппроксимировать любые сложные нелинейные или дискретные зависимости и строить любые разделяющие функции.

Исходная информация по химической структуре и фармакокинетическим показателям для человека известных соединений была нами получена из следующих источников: всасываемость в ЖКТ – [97]; объем распределения, клиренс общий, клиренс почечный, клиренс печеночный, период полувыведения – ChEMBL [93].

Для всех соединений с помощью программы DruLiTo [29] были рассчитаны 15 физико-химических и структурных параметров.

Оценку токсикологических свойств проводили с помощью классификационных ANN моделей, оценку фармакокинетических – регрессионных. В качестве общей архитектуры ANN был выбран двухслойный перцептрон с узким горлом $MLP\ k-m-l(f_1, f_2)$, где k – число входных нейронов, в данном случае это 15 параметров структуры соединений; m – число скрытых нейронов, устанавливается программой от 3 до 14, поскольку $l < m < k$; l – число выходных нейронов, равно двум для классификационных моделей и единице для регрессионных; f_1, f_2 – активационные

функции для скрытого и выходного слоев соответственно.

Обучение при моделировании каждого ADMET-показателя проводили с помощью программы Statistica 7 [98] в два этапа. Сначала с варьированием числа скрытых нейронов от 3 до 14 обучали по 2000 нейросетей с выбором одной с наибольшей точностью. Затем на основании этой нейросети с фиксированным числом скрытых нейронов обучали еще 200 нейросетей, оптимизируя веса синапсов и выбирая из них наилучшую по точности нейросеть. В случае классификационных моделей активационными функциями были *logistic* для скрытого и *softmax* для выходного слоев нейронов, а в случае регрессионных моделей – *logistic* для обоих слоев.

Таким образом, классификационные модели имели архитектуру нейросети $MLP-15-m-2(logistic, softmax)$, а регрессионные – $MLP-15-m-1(logistic, logistic)$.

Точность моделей оценивали методом случайного скользящего контроля, при котором тестовая выборка формируется с использованием генератора случайных чисел. Для классификационных зависимостей рассчитывали точность прогноза *Acc* (табл. 1), для регрессионных – коэффициент детерминации R^2 (табл. 2).

Всего в процессе построения ANN моделей для двенадцати ADMET-характеристик были обучены более 50 тысяч нейронных сетей.

Анализ данных таблицы 1 показывает, что точность прогноза в IT Microsoft только в двух случаях из шести превышает значения точности моделей, описанных в литературе, в то время как для ANN-моделей точность выше в четырех случаях. Анализ данных таблицы 2 показывает, что в трех случаях из шести точность ANN-моделей превышает значения точности моделей, описанных в литературе. Таким образом, из 12 параметров ADMET в семи случаях полученные нами ANN-модели обеспечивают более высокую точность прогноза в сравнении с литературными данными.

Полученные результаты доказывают, что предлагаемая методология построения моделей прогноза ADMET на основе ANN с классической архитектурой многослойного перцептрона [98] в сочетании с комбинированным описанием химической структуры параметрами DruLiTo [29] может быть эффективно использована для достаточно точной оценки токсикологических и фармакокинетических характеристик химических соединений.

Таблица 1. Оценка точности прогноза токсикологических ADMET-свойств химических соединений консенсусным и нейросетевым методами**Table 1.** Accuracy of prediction of toxicological ADMET properties of chemical compounds using consensus and neural network methods

Свойство <i>Property</i>	Число химических соединений <i>Number of compounds</i>	Точность прогноза Acc, % (по данным авторов) <i>Prediction accuracy, Acc, % (authors' data)</i>		Точность прогноза Acc, % (по данным литературы) <i>Prediction accuracy, Acc, % (literature data)</i>	Метод, источник литературы <i>Method, reference</i>
		IT Microcosm	ANN		
Острая токсичность* <i>Acute toxicity*</i>	3480	60,1	75,1**	74,0	RF [43]
Канцерогенность <i>Carcinogenicity</i>	489	85,1**	93,5**	80,1	SVM [47]
Мутагенность <i>Mutagenicity</i>	8390	79,6	78,7	98,0	kNN [49]
Генотоксичность <i>Genotoxicity</i>	370	81,7	91,1**	89,4	SVM [99]
Эндокринотоксичность <i>Endocrine toxicity</i>	1198	80,1	84,5	97,6	DNN [59]
Проникновение через гематоэнцефалический барьер <i>BBB penetration</i>	1820	93,4**	93,8**	91,9	SVM [95]

Таблица составлена авторами / The table is prepared by the authors

Примечание. ANN – искусственные нейронные сети; RF – метод случайного леса; SVM – метод опорных векторов; kNN – метод k-ближайших соседей; DNN – нейросети глубокого обучения.

* Для модели острой токсичности приведена сбалансированная точность BA.

** Превышает точность моделей, описанных в литературе.

Note. ANN, Artificial Neural Networks; RF, Random Forest; SVM, Support Vector Machine; kNN, k-Nearest Neighbours; DNN, Deep Neural Networks; BBB, blood-brain barrier.

* For the acute toxicity model, the table indicates balanced accuracy (BA) values.

** This value exceeds the accuracy of models described in the literature.

Сравнительный анализ методов прогнозирования ADMET-характеристик

В настоящее время применение компьютерных методов прогнозирования ADMET-характеристик препаратов-кандидатов становятся одним из важных этапов доклинического изучения новых лекарственных средств. В нормативных документах Российской Федерации и Евразийского экономического союза возможность использования методов *in silico* не рассматривается (или не регламентирована), но в соответствии с требованиями зарубежных фармакопей эти методы могут применяться на доклиническом этапе для оптимизации исследований на млекопитающих.

В *таблице 3* приведены совокупные данные по частоте применения десяти методов *in silico*, используемых в проанализированных в настоящем обзоре работах для построения моделей прогноза рассмотренных токсикологических и фармакокинетических ADMET-характеристик фармакологически активных соединений.

Из представленных в *таблице 3* данных видно, что наиболее востребованными являются методы RF и SVM (в сумме 58,6% использования). Однако они не всегда обеспечивают достижение максимальной точности: в случае прогноза наличия/отсутствия токсикологических свойств (классификация) максимальное значение Acc≈98% получено для других методов, kNN и DNN (*табл. 1*); но в случае количественного прогноза фармакокинетических показателей (регрессия) максимальное значение $R^2=0,920$ получено для метода RF (*табл. 2*).

В *таблице 1* приведены максимальные показатели точности, полученные в проанализированных публикациях при классификационном прогнозе токсикологических свойств. Наиболее эффективен, по данным литературы, прогноз мутагенности Acc=98,0% [49] и эндокринотоксичности Acc=97,6% [59].

В *таблице 2* приведены максимальные показатели точности, полученные в проанализированных публикациях при количественном прогнозе фармакокинетических свойств. Наиболее

эффективно, по данным литературы, прогнозируется почечный клиренс $R^2=0,920$ [83], несколько менее точно – общий клиренс $R^2=0,875$ [81].

Несмотря на это имеются ADMET-параметры, точность вычислений которых не очень высока. Так, по данным литературы, при прогнозе острой токсичности $Acc=74,0\%$ [43], а при прогнозе канцерогенности $Acc=80,1\%$ [47] (табл. 1). Также по данным литературы при прогнозе печеночного клиренса значение $R^2=0,265$ [85], а при прогнозе объема распределения $R^2=0,520$ [78] (табл. 2).

В проведенных нами исследованиях с использованием оригинальных нейросетевых моделей для указанных параметров получено наиболее существенное улучшение качества прогноза: до $Acc=93,5\%$ в случае канцерогенности, до $R^2=0,501$ в случае печеночного клиренса и до $R^2=0,618$ в случае объема распределения.

Для семи ADMET-характеристик (острая токсичность, канцерогенность, генотоксичность, проникновение через гематоэнцефалический барьер, всасываемость в ЖКТ, объем распределения, печеночный клиренс) повышение точности прогноза с применением построенных нами нейросетевых моделей изменялось в диапазонах от $\Delta Acc=1,1\%$ (острая токсичность) до $\Delta Acc=13,4\%$

(канцерогенность) и от $\Delta R^2=0,033$ (всасываемость в ЖКТ) до $\Delta R^2=0,236$ (печеночный клиренс).

Разработанный нами подход к прогнозу ADMET на основе технологии искусственных нейронных сетей позволяет эффективно прогнозировать ряд ADMET-параметров, в том числе: канцерогенность, генотоксичность и проникновение через гематоэнцефалический барьер с точностью Acc более 90%, а всасываемость в ЖКТ – с точностью $R^2=0,825$.

Заключение

Выполнен анализ публикаций по методам прогноза *in silico* ряда важных токсикологических и фармакокинетических ADMET-характеристик преимущественно за последние пять лет. Далеко не все проанализированные современные подходы реализованы их авторами в виде свободно доступных программ или web-ресурсов. Тем не менее часть из них используется на семи web-платформах, ссылки на которые приведены в соответствующей части настоящего обзора. Эти системы могут быть использованы как в процессе обучения методам компьютерной оценки ADMET-параметров химических соединений, так и в реальных практических исследованиях

Таблица 2. Точность прогноза фармакокинетических ADMET-свойств химических соединений нейросетевым методом

Table 2. Accuracy of prediction of pharmacokinetic ADMET properties of chemical compounds using the neural network method

Свойство <i>Property</i>	Число химических соединений <i>Number of compounds</i>	Коэффициент детерминации R^2 , метод ANN (по данным авторов) <i>Coefficient of determination, R^2, ANN (authors' data)</i>	Коэффициент детерминации R^2 (по данным литературы) <i>Coefficient of determination, R^2 (literature data)</i>	Метод, источник литературы <i>Method, reference</i>
Всасываемость в желудочно-кишечном тракте <i>Human intestinal absorption</i>	572	0,825*	0,792	SVM [100]
Объем распределения** <i>Volume of distribution**</i>	1119	0,618*	0,520	RF [78]
Клиренс общий <i>Total clearance</i>	663	0,530	0,875	RF [81]
Клиренс почечный <i>Renal clearance</i>	640	0,597	0,920	RF [83]
Клиренс печеночный <i>Hepatic clearance</i>	632	0,501*	0,265	RF [85]
Период полувыведения <i>Half-life period</i>	861	0,509	0,832	RF [81]

Таблица составлена авторами / The table is prepared by the authors

Примечание. ANN – искусственные нейронные сети; SVM – метод опорных векторов; RF – метод случайного леса.

* Для модели объема распределения значение R^2 рассчитано в 10-кратном скользящем контроле.

** Превышает точность моделей, описанных в литературе.

Note. ANN, Artificial Neural Networks; SVM, Support Vector Machine; RF, Random Forest.

* For the model describing the volume of distribution, the R^2 value was obtained in 10-fold cross validation.

** This value exceeds the accuracy of models described in the literature.

Таблица 3. Показатели частоты использования различных методов *in silico* при построении моделей ADMET-свойств химических соединений**Table 3.** Frequency of using various *in silico* methods for constructing ADMET models for chemical compounds

Метод <i>Method</i>	Вид модели <i>Model type</i>	Частота использования <i>Frequency of use</i>		Источник литературы <i>Reference</i>
		Доля, % <i>Percentage, %</i>	Число <i>Number</i>	
Множественная линейная регрессия (MLR) <i>Multiple linear regression (MLR)</i>	Регрессия <i>Regression</i>	4,3	3	[47, 63, 82]
Частичная регрессия наименьших квадратов (PLS) <i>Partial least squares regression (PLS)</i>	Регрессия <i>Regression</i>	7,1	5	[73, 75, 76, 83]
Линейный дискриминантный анализ (LDA) <i>Linear Discriminant Analysis (LDA)</i>	Классификация <i>Classification</i>	1,4	1	[73]
Наивный байесовский классификатор (NBC) <i>Naive Bayes Classifier (NBC)</i>	Классификация <i>Classification</i>	4,3	3	[43, 49, 64]
Метод k-ближайших соседей (kNN) <i>k-Nearest Neighbours (kNN)</i>	Классификация <i>Classification</i>	8,6	6	[45, 49, 73, 76]
Метод опорных векторов (SVM) <i>Support Vector Machine (SVM)</i>	Регрессия, классификация <i>Regression, classification</i>	24,3	17	[43, 49, 64, 68, 73, 75, 76, 80, 81, 83, 95, 99, 100]
Метод случайного леса (RF) <i>Random Forest (RF)</i>	Регрессия, классификация <i>Regression, classification</i>	34,3	24	[43, 45, 49, 64, 68, 71, 72, 76, 78–85]
Искусственная нейронная сеть (ANN) <i>Artificial Neural Network (ANN)</i>	Регрессия, классификация <i>Regression, classification</i>	8,6	6	[49, 73, 76, 83]
Сверточная нейронная сеть (CNN) <i>Convolutional Neural Network (CNN)</i>	Классификация <i>Classification</i>	1,4	1	[64]
Глубинная нейронная сеть (DNN) <i>Deep Neural Network (DNN)</i>	Классификация <i>Classification</i>	5,7	4	[45, 59, 64]

Таблица составлена авторами / The table is prepared by the authors

по прогнозу *in silico* безопасности новых фармакологически активных веществ.

Однако в соответствии с вышеописанными результатами наиболее перспективным и практически значимым направлением в разработке

систем *in silico* прогноза токсикологических и фармакокинетических характеристик новых лекарственных препаратов следует считать использование технологии искусственных нейронных сетей.

Литература / References

1. Tsaion K, Kates SA, eds. *ADMET for Medicinal Chemists: A Practical Guide*. New York: Wiley; 2011.
2. Дурнев АД. Лекарственная токсикология занимает важнейшее место в структуре доклинических исследований. *Ведомости Научного центра экспертизы средств медицинского применения. Регуляторные исследования и экспертиза лекарственных средств*. 2023;13(1):8–13. Durnev AD. Pharmaceutical toxicology is the most important component of preclinical studies. *Bulletin of the Scientific Centre for Expert Evaluation of Medicinal Products. Regulatory Research and Medicine Evaluation*. 2023;13(1):8–13 (In Russ.). <https://doi.org/10.30895/1991-2919-2023-13-1-8-13>
3. Boroujerdi M. *Pharmacokinetics and Toxicokinetics*. New York: CRC Press; 2015.
4. Shaker B, Ahmad S, Lee J, Jung C, Na D. *In silico* methods and tools for drug discovery. *Comput Biol Med*. 2021;137:104851. <https://doi.org/10.1016/j.compbiomed.2021.104851>
5. Chang Y, Hawkins BA, Du JJ, Groundwater PW, Hibbs DE, Lai F. A guide to *in silico* drug design. *Pharmaceutics*. 2023;15(1):49. <https://doi.org/10.3390/pharmaceutics15010049>
6. Sammut C, Webb GI, eds. *Encyclopedia of machine learning*. New York: Springer; 2011.
7. Devillers J, Balaban AT, eds. *Topological indices and related descriptors in QSAR and QSPR*. New York: CRC Press; 2000.
8. Васильев ПМ, Спасов АА. Языки фрагментарного кодирования структуры соединений для компьютерного прогноза биологической активности. *Российский химический журнал (Журнал Российского химического обще-*

- ства им. Д.И. Менделеева). 2006;50(2):108–27.
- Vassiliev PM, Spasov AA Fragmentary encoding languages of compound structure for computer prediction of biological activity. *Žurnal Vsesoúznogo himičeskogo obšestva im. D.I. Mendeleeva*. 2006;50(2):108–27 (In Russ.). EDN: HTUUSP
- Engel T, Gasteiger J, eds. *Cheminformatics: Basic Concepts and Methods*. Weinheim: Wiley-VCH; 2018.
 - Wang L, Ding J, Pan L, Cao D, Jiang H, Ding X. Quantum chemical descriptors in quantitative structure–activity relationship models and their applications. *Chemometr Intell Lab Syst*. 2021;217:104384. <https://doi.org/10.1016/j.chemolab.2021.104384>
 - Batool M, Ahmad B, Choi S. A structure-based drug discovery paradigm. *Int J Mol Sci*. 2019;20(11):2783. <https://doi.org/10.3390/ijms20112783>
 - Kar S, Leszczynski J. Open access *in silico* tools to predict the ADMET profiling of drug candidates. *Expert Opin Drug Discov*. 2020;15(12):1473–87. <https://doi.org/10.1080/17460441.2020.1798926>
 - Fahrmeir L, Kneib T, Lang S, Marx BD. *Regression: models, methods and applications*. New York: Springer; 2021.
 - Gramatica P. Principles of QSAR modeling: comments and suggestions from personal experience. *Int J Quant Struct Prop Relatsh*. 2020;5(3):61–97. <https://doi.org/10.4018/IJQSPR.20200701.oa1>
 - Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12(85):2825–30.
 - Poroikov VV, Filimonov DA, Borodina YV, Lagunin AA, Kos A. Robustness of biological activity spectra predicting by computer program pass for noncongeneric sets of chemical compounds. *J Med Chem*. 2000;40(6):1349–55. <https://doi.org/10.1021/ci000383k>
 - Zhang Z. Introduction to machine learning: K-nearest neighbors. *Ann Trans Med*. 2016;4(11):218–24. <https://doi.org/10.21037/atm.2016.03.37>
 - Cristianini N, Shawe-Taylor J. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge: Cambridge University Press; 2000.
 - Genuer R, Poggi J-M. *Random forests with R (Use R!)*. New York: Springer; 2020.
 - Jang IS, Ghazoui Z, Ahsen ME, Vogel R, Neto EC, Norman T, et al. Community assessment to advance computational prediction of cancer drug combinations in a pharmacogenomic screen. *Nat Commun*. 2019;10(1):2674. <https://doi.org/10.1038/s41467-019-09799-2>
 - Aggarwal CC. *Neural networks and deep learning: a textbook*. New York: Springer; 2018.
 - Han S-H, Kim KW, Kim S, Youn YC. Artificial neural network: understanding the basic concepts without mathematics. *Dement Neurocogn Disord*. 2018;17(3):83–9. <https://doi.org/10.12779/dnd.2018.17.3.83>
 - Che J, Chen L, Guo ZH, Wang S, Aorigele C. Drug target group prediction with multiple drug networks. *Comb Chem High Throughput Screen*. 2020;23(4):274–84. <https://doi.org/10.2174/1386207322666190702103927>
 - Pinzi L, Rastelli G. Molecular docking: shifting paradigms in drug discovery. *Int J Mol Sci*. 2019;20(18):4331. <https://doi.org/10.3390/ijms20184331>
 - Wang Z, Sun H, Yao X, Li D, Xu L, Li Y, et al. Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: the prediction accuracy of sampling power and scoring power. *Phys Chem Chem Phys*. 2016;18(18):12964–75. <https://doi.org/10.1039/c6cp01555g>
 - Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. 2010;31(2):455–61. <https://doi.org/10.1002/jcc.21334>
 - Dickson CJ, Velez-Vega C, Duca JS. Revealing molecular determinants of hERG blocker and activator binding. *J Chem Inf Model*. 2020;60(1):192–203. <https://doi.org/10.1021/acs.jcim.9b00773>
 - Plewczynski D, Łażniewski M, Augustyniak R, Ginalski K. Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. *J Comput Chem*. 2011;32(4):742–55. <https://doi.org/10.1002/jcc.21643>
 - Joshi T, Sharma P, Joshi T, Pundir H, Mathpal S, Chandra S. Structure-based screening of novel lichen compounds against SARS Coronavirus main protease (Mpro) as potential inhibitors of COVID-19. *Mol Divers*. 2021;25(3):1665–77. <https://doi.org/10.1007/s11030-020-10118-x>
 - Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*. 2001;46(1-3):3–26. [https://doi.org/10.1016/s0169-409x\(00\)00129-0](https://doi.org/10.1016/s0169-409x(00)00129-0)
 - Oprea TI. Property distribution of drug-related chemical databases. *J Comput Aided Mol Des*. 2000;14(3):251–64. <https://doi.org/10.1023/a:1008130001697>
 - Veber DF, Johnson SR, Cheng H-Y, Smith BR, Ward KW, Kopple KD. Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem*. 2002;45(12):2615–23. <https://doi.org/10.1021/jm020017n>
 - Ghose AK, Viswanadhan VN, Wendoloski JJ. A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *J Comb Chem*. 1999;1(1):55–68. <https://doi.org/10.1021/cc9800071>
 - Pardridge WM. CNS drug design based on principles of blood-brain barrier transport. *J Neurochem*. 1998;70(5):1781–92. <https://doi.org/10.1046/j.1471-4159.1998.70051781.x>
 - Bickerton GR, Paolini Gaia V, Besnard J, Muresan S, Hopkins AL. Quantifying the chemical beauty of drugs. *Nat Chem*. 2012;4(2):90–8. <https://doi.org/10.1038/nchem.1243>
 - Ahlberg E, Carlsson L, Boyer S. Computational derivation of structural alerts from large toxicology data sets. *J Chem Inf Model*. 2014;54(10):2945–52. <https://doi.org/10.1021/ci500314a>
 - Daina A, Michielin O, Zoeteb V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep*. 2017;7:42717. <https://doi.org/10.1038/srep42717>
 - O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: an open chemical toolbox. *J Cheminform*. 2011;3(1):33. <https://doi.org/10.1186/1758-2946-3-33>
 - Banerjee P, Eckert AO, Schrey AK, Preissner R. ProTox-II: a webserver for the prediction of toxicity of chemicals. *Nucleic Acids Res*. 2018;46(W1):W257–63. <https://doi.org/10.1093/nar/gky318>
 - Pires DEV, Blundell TL, Ascher DB. pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J Med Chem*. 2015;58:4066–72. <https://doi.org/10.1021/acs.jmedchem.5b00104>
 - Dong J, Wang N-N, Yao Z-J, Zhang L, Cheng Y, Ouyang D,

- et al. ADMETlab: a platform for systematic ADMET evaluation based on a comprehensively collected ADMET database. *J Cheminform.* 2018;10(1):29.
<https://doi.org/10.1186/s13321-018-0283-x>
42. Yang H, Lou C, Sun L, Li J, Cai Y, Wang Z, et al. admetSAR 2.0: web-service for prediction and optimization of chemical ADMET properties. *Bioinformatics.* 2019;35(6):1067–69.
<https://doi.org/10.1093/bioinformatics/bty707>
43. Lunghini F, Marcou G, Azam P, Horvath D, Patoux R, Van Miert E, et al. Consensus models to predict oral rat acute toxicity and validation on a dataset coming from the industrial context. *SAR QSAR Environ Res.* 2019;30(12):879–97.
<https://doi.org/10.1080/1062936X.2019.1672089>
44. Ruggiu F, Marcou G, Varnek A, Horvath D. ISIDA property-labelled fragment descriptors. *Mol Inform.* 2010;29(12):855–68.
<https://doi.org/10.1002/minf.201000099>
45. Sosnin S, Karlov D, Tetko IV, Fedorov MV. Comparative study of multitask toxicity modeling on a broad chemical space. *J Chem Inf Model.* 2018;59(3):1062–72.
<https://doi.org/10.1021/acs.jcim.8b00685>
46. Tice RR, Bassan A, Amberg A, Anger LT, Beal MA, Bellion P, et al. *In silico* approaches in carcinogenicity hazard assessment: current status and future needs. *Comput Toxicol.* 2021;20:100191.
<https://doi.org/10.1016/j.comtox.2021.100191>
47. Zhong M, Nie X, Yan A, Yuan Q. Carcinogenicity prediction of noncongeneric chemicals by a support vector machine. *Chem Res Toxicol.* 2013;26(5):741–9.
<https://doi.org/10.1021/tx4000182>
48. Sushko I, Novotarskiy S, Körner R, Pandey AK, Kovalevskiy VV, Prokopenko VV, et al. Applicability domain for *in silico* models to achieve accuracy of experimental measurements. *J Chemom.* 2010;24(3–4):202–8.
<https://doi.org/10.1002/cem.1296>
49. Xu C, Cheng F, Chen L, Du Z, Li W, Liu G, et al. *In silico* prediction of chemical Ames mutagenicity. *J Chem Inf Model.* 2012;52(11):2840–7.
<https://doi.org/10.1021/ci300400a>
50. Patlewicz G, Jeliakova N, Safford RJ, Worth AP, Aleksiev B. An evaluation of the implementation of the Cramer classification scheme in the Toxtree software. *SAR QSAR Environ Res.* 2008;19(5–6):495–524.
<https://doi.org/10.1080/10629360802083871>
51. Стрелкова ЮН. Понятия «генотоксичность» и «мутатогенность». В кн. *Advances in Science and Technology. Сборник статей XXVI международной научно-практической конференции.* М.: Актуальность.РФ; 2020. Ч. I. С. 37–8.
Strelkova YuN. Concepts of “genotoxicity” and “mutagenicity”. In: *Advances in science and technology. Collected papers of XXVI international scientific-practical conference.* Moscow: Aktualnost.RF; 2020. Part 1. P. 37–8 (In Russ.). EDN: [IKZUNJ](https://doi.org/10.1007/978-1-0716-1960-5_8)
52. Baderna D, Van Overmeire I, Lavado GJ, Gadaleta D, Mertens B. *In silico* Methods for chromosome damage. In: *In silico methods for predicting drug toxicity.* New York: Springer US; 2022. P. 185–200.
https://doi.org/10.1007/978-1-0716-1960-5_8
53. Kusko R, Hong H. Machine learning and deep learning promote computational toxicology for risk assessment of chemicals. In: *Machine learning and deep learning in computational toxicology.* Cham: Springer International Publishing; 2023. P. 1–17.
<https://doi.org/10.1007/978-3-031-20730-3>
54. Baderna D, Gadaleta D, Lostaglio E, Selvestrel G, Raitano G, Golbamaki A, et al. New *in silico* models to predict *in vitro* micronucleus induction as marker of genotoxicity. *J Hazard Mater.* 2020;385:121638.
<https://doi.org/10.1016/j.jhazmat.2019.121638>
55. Ferrari T, Cattaneo D, Gini G, Bakhtyari NG, Manganaro A, Benfenati E. Automatic knowledge extraction from chemical structures: the case of mutagenicity prediction. *SAR QSAR Environ Res.* 2013;24(5):365–83.
<https://doi.org/10.1080/1062936X.2013.773376>
56. Combarrous Y, Diep Nguyen TM. Comparative overview of the mechanisms of action of hormones and endocrine disruptor compounds. *Toxics.* 2019;7(1):5–14.
<https://doi.org/10.3390/toxics7010005>
57. La Merrill MA, Vandenberg LN, Smith MT, Goodson W, Browne P, Patisaul HB, et al. Consensus on the key characteristics of endocrine-disrupting chemicals as a basis for hazard identification. *Nat Rev Endocrinol.* 2020;16(1):45–57.
<https://doi.org/10.1038/s41574-019-0273-8>
58. Ruiz P, Sack A, Wampole M, Bobst S, Vracko M. Integration of *in silico* methods and computational systems biology to explore endocrine-disrupting chemical binding with nuclear hormone receptors. *Chemosphere.* 2021;178:99–109.
<https://doi.org/10.1016/j.chemosphere.2017.03.026>
59. Matsuzaka Y, Uesawa Y. Molecular image-based prediction models of nuclear receptor agonists and antagonists using the deepsnap-deep learning approach with the Tox21 10K library. *Molecules.* 2020;25(12):2764.
<https://doi.org/10.3390/MOLECULES25122764>
60. Collins SP, Barton-Maclaren TS. Novel machine learning models to predict endocrine disruption activity for high-throughput chemical screening. *Front Toxicol.* 2022;4:1–13.
<https://doi.org/10.3389/ftox.2022.981928>
61. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.* 2019;47(D1):D1102–D1109.
<https://doi.org/10.1093/nar/gky1033>
62. Halle W, Halder M, Worth A, Genschow E. The registry of cytotoxicity: toxicity testing in cell cultures to predict acute toxicity (LD50) and to reduce testing in animals. *Altern Lab Anim.* 2003;31(2):89–198.
<https://doi.org/10.1177/026119290303100204>
63. Petrescu AM, Paunescu V, Ilia G. The antiviral activity and cytotoxicity of 15 natural phenolic compounds with previously demonstrated antifungal activity. *J Environ Sci Health B.* 2019;54(6):498–504.
<https://doi.org/10.1080/03601234.2019.1574176>
64. Chuipu C, Guo P, Zhou Y, Zhou J, Wang Q, Zhang F, et al. Deep learning-based prediction of drug-induced cardiotoxicity. *J Chem Inf Model.* 2019;59(3):1073–84.
<https://doi.org/10.1021/acs.jcim.8b00769>
65. He S, Ye T, Wang R, Zhang C, Zhang X, Sun G, et al. An *in silico* model for predicting drug-induced hepatotoxicity. *Int J Mol Sci.* 2019;20(8):1897.
<https://doi.org/10.3390/ijms20081897>
66. Ye H, Nelson LJ, Moral MGD, Martínez-Naves E, Cubero FJ. Dissecting the molecular pathophysiology of drug-induced liver injury. *World J Gastroenterol.* 2018;24(13):1373–85.
<https://doi.org/10.3748/wjg.v24.i13.1373>
67. García-Cañaveras JC, Jiménez N, Gómez-Lechón MJ, Castell JV, Donato MT, Lahoz A. LC-MS untargeted metabolomic analysis of drug-induced hepatotoxicity in HepG2 cells. *Electrophoresis.* 2015;36(18):2294–302.
<https://doi.org/10.1002/elps.201500095>
68. Kim E, Nam H. Prediction models for drug-induced hepatotoxicity by using weighted molecular fingerprints. *BMC Bioinformatics.* 2017;18(Suppl 7):25–34.
<https://doi.org/10.1186/s12859-017-1638-4>
69. Turabekova M, Rasulev B, Theodore M, Jackman J, Leszczynska D, Leszczynski J. Immunotoxicity of nanoparticles: a computational study suggests that CNTs and C60

- fullerenes might be recognized as pathogens by Toll-like receptors. *Nanoscale*. 2014;6(7):3488–95. <https://doi.org/10.1039/C3NR05772K>
70. Roos C. *Intestinal absorption of drugs: the impact of regional permeability, nanoparticles, and absorption-modifying excipients*. Uppsala: Acta Universitatis Upsaliensis; 2018.
 71. Venkatraman V. FP-ADMET: a compendium of fingerprint-based ADMET prediction models. *J Cheminform*. 2021;13(1):75–86. <https://doi.org/10.1186/s13321-021-00557-5>
 72. Wang N-N, Huang C, Dong J, Yao Z-J, Zhu M-F, Deng Z-K, et al. Predicting human intestinal absorption with modified random forest approach: a comprehensive evaluation of molecular representation, unbalanced data, and applicability domain issues. *RSC Adv*. 2017;7(31):19007–18. <https://doi.org/10.1039/C6RA28442F>
 73. Kumar R, Sharma A, Siddiqui MH, Tiwari RK. Prediction of human intestinal absorption of compounds using artificial intelligence techniques. *Curr Drug Discov Technol*. 2017;14(4):244–54. <https://doi.org/10.2174/1570163814666170404160911>
 74. Миронов АН, ред. *Руководство по проведению доклинических исследований лекарственных средств*. Ч. 2. М.: Гриф и К; 2012.
Mironov AN, ed. *Guide for conducting preclinical studies of medicines*. Part 2. Moscow: Grif i K; 2012 (In Russ.).
 75. Fagerholm U, Hellberg S, Spjuth O. Advances in predictions of oral bioavailability of candidate drugs in man with new machine learning methodology. *Molecules*. 2021;26(9):2572–82. <https://doi.org/10.3390/molecules26092572>
 76. Ye Z, Yang Y, Li X, Cao D, Ouyang D. An integrated transfer learning and multitask learning approach for pharmacokinetic parameter prediction. *Mol Pharm*. 2019;16(2):533–41. <https://doi.org/10.1021/acs.molpharmaceut.8b00816>
 77. Currie GM. Pharmacology, part 2: Introduction to pharmacokinetics. *J Nucl Med Technol*. 2018;46(3):221–30. <https://doi.org/10.2967/jnmt.117.199638>
 78. Murad N, Pasikanti KK, Madej BD, Minnich A, McComas JM, Crouch S, et al. Predicting volume of distribution in humans: performance of *in silico* methods for a large set of structurally diverse clinical compounds. *Drug Metab Dispos*. 2021;49(2):169–78. <https://doi.org/10.1124/dmd.120.000202>
 79. Lombardo F, Bentzien J, Berellini G, Muegge I. *In silico* models of human PK parameters. Prediction of volume of distribution using an extensive data set and a reduced number of parameters. *J Pharm Sci*. 2021;110(1):500–9. <https://doi.org/10.1016/j.xphs.2020.08.023>
 80. Kosugi Y, Hosea N. Direct comparison of total clearance prediction: computational machine learning model versus bottom-up approach using *in vitro* assay. *Mol Pharm*. 2020;17(7):2299–309. <https://doi.org/10.1021/acs.molpharmaceut.9b01294>
 81. Wang Y, Liu H, Fan Y, Chen X, Yang Y, Zhu L, et al. *In silico* prediction of human intravenous pharmacokinetic parameters with improved accuracy. *J Chem Inf Model*. 2019;59(9):3968–80. <https://doi.org/10.1021/acs.jcim.9b00300>
 82. Chen J, Yang H, Zhu L, Wu Z, Li W, Tang Y, et al. *In silico* prediction of human renal clearance of compounds using quantitative structure-pharmacokinetic relationship models. *Chem Res Toxicol*. 2020;33(2):640–50. <https://doi.org/10.1021/acs.chemrestox.9b00447>
 83. Watanabe R, Ohashi R, Esaki T, Kawashima H, Natsume-Kitani Y, Nagao C, et al. Development of an *in silico* prediction system of human renal excretion and clearance from chemical structure information incorporating fraction unbound in plasma as a descriptor. *Sci Rep*. 2019;9(1):18782–92. <https://doi.org/10.1038/s41598-019-55325-1>
 84. Dawson DE, Ingle BL, Phillips KA, Nichols JW, Wambaugh JF, Tornero-Velez R. Designing QSARs for parameters of high-throughput toxicokinetic models using open-source descriptors. *Environ Sci Technol*. 2021;55(9):6505–17. <https://doi.org/10.1021/acs.est.0c06117>
 85. Feinberg EN, Joshi E, Pande VS, Cheng AC. Improvement in ADMET prediction with multitask deep featurization. *J Med Chem*. 2020;63(16):8835–48. <https://dx.doi.org/10.1021/acs.jmedchem.9b02187>
 86. Vassiliev PM, Spasov AA, Kosolapov VA, Kucheryavenko AF, Gurova NA, Anisimova VA. Consensus drug design using IT Microcosm. In: Gorb L, Kuz'min V, Muratov E, eds. *Application of computational techniques in pharmacy and medicine; challenges and advances in computational chemistry and physics*. Vol. 17. Springer, Dordrecht; 2014. P. 369–431. https://doi.org/10.1007/978-94-017-9257-8_12
 87. Васильев ПМ, Спасов АА, Кочетков АН, Бабков ДА, Литвинов РА. Консенсусная оценка *in silico* острой токсичности мультитаргетных RAGE-ингибиторов. *Вестник Волгоградского государственного медицинского университета*. 2020;17(1):31–3.
Vassiliev PM, Spasov AA, Kochetkov AN, Babkov DA, Litvinov RA. *In silico* consensus assessment of the acute toxicity of multitarget RAGE inhibitors. *Journal of Volgograd State Medical University*. 2020;17(1):31–3 (In Russ.). [https://doi.org/10.19163/1994-9480-2020-1\(73\)-31-33](https://doi.org/10.19163/1994-9480-2020-1(73)-31-33)
 88. Васильев ПМ, Спасов АА, Кочетков АН, Бабков ДА, Литвинов РА. Консенсусный прогноз *in silico* канцерогенной опасности мультитаргетных RAGE-ингибиторов. *Волгоградский научно-медицинский журнал*. 2020;(1):55–7.
Vassiliev PM, Spasov AA, Kochetkov AN, Babkov DA, Litvinov RA. Consensus *in silico* prediction of the carcinogenic hazard of multitarget RAGE inhibitors. *Volgograd Scientific and Medical Journal*. 2020;(1):55–7 (In Russ.). EDN: [TJAFRE](https://doi.org/10.19163/1994-9480-2020-1(73)-31-33)
 89. Васильев ПМ, Спасов АА, Кочетков АН, Перфильев МА, Королева АР, Голубева АВ и др. Консенсусная оценка *in silico* общей безопасности мультитаргетных RAGE-ингибиторов. *Волгоградский научно-медицинский журнал*. 2020;(2):47–51.
Vassiliev PM, Spasov AA, Kochetkov AN, Perfiliev MA, Koroleva AR, Golubeva AV, et al. *In silico* consensus assessment of the overall safety of multitarget RAGE inhibitors. *Volgograd Scientific and Medical Journal*. 2020;(2):47–51 (In Russ.). EDN: [VYKKZP](https://doi.org/10.19163/1994-9480-2020-2(74)-100-104)
 90. Васильев ПМ, Спасов АА, Кочетков АН, Перфильев МА, Королева АР, Голубева АВ и др. Консенсусный прогноз *in silico* фармакокинетической предпочтительности мультитаргетных RAGE-ингибиторов. *Вестник Волгоградского государственного медицинского университета*. 2020;(2):100–4.
Vassiliev PM, Spasov AA, Kochetkov AN, Perfiliev MA, Koroleva AR, Golubeva AV, et al. *In silico* consensus prediction of pharmacokinetic preference for multitarget RAGE inhibitors. *Journal of Volgograd State Medical University*. 2020;(2):100–4 (In Russ.). [https://doi.org/10.19163/1994-9480-2020-2\(74\)-100-104](https://doi.org/10.19163/1994-9480-2020-2(74)-100-104)
 91. Tomasulo P. ChemIDplus — super source for chemical and drug information. *Med Ref Serv*. 2002;21(1):53–9. https://doi.org/10.1300/j115v21n01_04
 92. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*. Vol. 1–131. Lyon: IARC; 1972–2023.
 93. Mendez D, Gaulton A, Bento AP, Chambers J, De Veij M,

- Felix E, et al. ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.* 2019;47(D1):D930–40. <https://doi.org/10.1093/nar/gky1075>
94. Karthikeyan BS, Ravichandran J, Aparna SR, Samal A. DE-DuCT 2.0: An updated knowledgebase and an exploration of the current regulations and guidelines from the perspective of endocrine disrupting chemicals. *Chemosphere.* 2021;267:128898. <https://doi.org/10.1016/j.chemosphere.2020.128898>
95. Wang Z, Yang H, Wu Z, Wang T, Li W, Tang Y, et al. *In silico* prediction of blood–brain barrier permeability of compounds by machine learning and resampling methods. *Chem Med Chem.* 2018;13(20):2189–201. <https://doi.org/10.1002/cmdc.201800533>
96. Колмогоров АН. О представлении непрерывных функций нескольких переменных в виде суперпозиции непрерывных функций одного переменного. *Доклады АН СССР.* 1958;114(5):953–6.
- Kolmogorov AN. On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. *Proc. USSR Acad. Sci.* 1957;114(5):953–6 (In Russ.).
97. Shen J, Cheng F, Xu Y, Li W, Tang Y. Estimation of ADME properties with substructure pattern recognition. *J Chem Inf Model.* 2010;50(6):1034–41. <https://doi.org/10.1021/ci100104j>
98. Hilbe JM. Statistica 7: an overview. *Am Stat.* 2007;61(1):91–4.
99. Li H, Ung CY, Yap CW, Xue Y, Li ZR, Cao ZW, et al. Prediction of genotoxicity of chemical compounds by statistical learning methods. *Chem Res Toxicol.* 2005;18(6):1071–80. <https://doi.org/10.1021/tx049652h>
100. Yan A, Wang Z, Cai Z. Prediction of human intestinal absorption by GA feature selection and support vector machine regression. *Int J Mol Sci.* 2008;9(10):1961–76. <https://doi.org/10.3390/ijms9101961>

Вклад авторов. Все авторы подтверждают соответствие своего авторства критериям ICMJE. Наибольший вклад распределен следующим образом: П.М. Васильев – концептуализация и методология исследования, общее руководство проектом, написание и редактирование текста рукописи, формулировка выводов, утверждение окончательного варианта рукописи для публикации; А.В. Голубева, А.Р. Королева, М.А. Перфильев, А.Н. Кочетков – написание и редактирование текста рукописи.

Authors' contributions. All the authors confirm that they meet the ICMJE criteria for authorship. The most significant contributions were as follows. Pavel M. Vassiliev elaborated the concept and methodology of the study, supervised the research project, drafted and edited the manuscript, formulated the conclusions, and approved the final version of the manuscript for publication. Arina V. Golubeva, Anastasia R. Koroleva, Maksim A. Perfilev, and Andrey N. Kochetkov drafted and edited the manuscript.

ОБ АВТОРАХ / AUTHORS

Васильев Павел Михайлович, д-р биол. наук, ст. науч. сотр., доцент

ORCID: <https://orcid.org/0000-0002-8188-5052>
pvassiliev@mail.ru

Голубева Арина Владимировна

ORCID: <https://orcid.org/0000-0001-8268-8811>
arina_arina_golubeva@mail.ru

Королева Анастасия Руслановна

ScopusID: [57224934568](https://orcid.org/0000-0001-8268-8811)
koroLyova1698@mail.ru

Перфильев Максим Алексеевич

ORCID: <https://orcid.org/0000-0002-5326-3299>
maxim.firu@yandex.ru

Кочетков Андрей Николаевич

ORCID: <https://orcid.org/0000-0003-3077-1837>
akocha@mail.ru

Поступила 06.06.2023

После доработки 03.11.2023

Принята к печати 17.11.2023

Pavel M. Vassiliev, Dr. Sci. (Biol.), Senior Researcher, Associate Professor

ORCID: <https://orcid.org/0000-0002-8188-5052>
pvassiliev@mail.ru

Arina V. Golubeva

ORCID: <https://orcid.org/0000-0001-8268-8811>
arina_arina_golubeva@mail.ru

Anastasia R. Koroleva

ScopusID: [57224934568](https://orcid.org/0000-0001-8268-8811)
koroLyova1698@mail.ru

Maksim A. Perfilev

ORCID: <https://orcid.org/0000-0002-5326-3299>
maxim.firu@yandex.ru

Andrey N. Kochetkov

ORCID: <https://orcid.org/0000-0003-3077-1837>
akocha@mail.ru

Received 6 June 2023

Revised 3 November 2023

Accepted 17 November 2023