

УДК 004.89
МРНТИ 28.23.33

РАЗРАБОТКА МЕТОДИКИ ОЦЕНКИ ЭФФЕКТИВНОСТИ SMART-ТЕХНОЛОГИИ ПРОГНОЗИРОВАНИЯ СВОЙСТВ ЛЕКАРСТВЕННЫХ СОЕДИНЕНИЙ И АНАЛИЗА БАЗ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ СОВРЕМЕННЫХ ПРОГРАММНЫХ СРЕДСТВ

САМИГУЛИНА Г.А.¹, САМИГУЛИНА З.И.²

¹Институт информационных и вычислительных технологий

²Казахстанско-Британский технический университет

Аннотация: В настоящее время перспективным является разработка инновационных методик для создания новых лекарственных препаратов с заданными свойствами с целью снижения временных и финансовых затрат. Поиск эффективных лекарственных соединений является сложным, много-стадийным процессом, при котором необходимо обрабатывать огромный объем химических данных. Актуально применение современных методов искусственного интеллекта для прогнозирования зависимости «структура-свойство» лекарственных соединений. В статье представлена разработанная Smart-технология прогнозирования на основе модифицированных алгоритмов искусственных иммунных систем. Оценка эффективности Smart-технологии осуществляется с помощью методологии FMEA (Failure Mode and Effects Analysis) для анализа причин и последствий возникновения дефектов. Разработана модель FMEA для оценки рисков функционирования этапов Smart-технологии. В качестве примера рассматриваются лекарственные соединения сульфаниламидной группы.

Ключевые слова: лекарственные соединения, прогнозирование зависимости «структура-свойство», Smart-технология, выделение информативных дескрипторов, модифицированные алгоритмы искусственных иммунных систем, сульфаниламиды, модель FMEA для оценки рисков

ЗАМАНАУИ БАҒДАРЛАМАЛЫҚ ҚҰРАЛДАРЫН ПАЙДАЛАНА ОТЫРЫП МӘЛІМЕТТЕР ҚОРЫН ТАЛДАУ ЖӘНЕ ДӘРІЛІК ҚОСЫЛЫСТАРДЫҢ ҚАСИЕТТЕРІН БОЛЖАУДЫҢ SMART-ТЕХНОЛОГИЯСЫНЫҢ ТИІМДІЛІГІН БАҒАЛАУ ӘДІСІН ӨЗІРЛЕУ

Андатпа: Қазіргі таңда уақытты және қаржылық шығынды азайту мақсатында белгіленген қасиеттер бойынша жаңа дәрілік препараттарды жасауға арналған инновациялық әдістерді әзірлеудің келешегі зор. Тиімді дәрілік қосылыстарды іздеу өте үлкен көлемдегі химиялық мәліметтерді өңдеуге тура келетін күрделі, көп сатылы үдеріс болып табылады. Дәрілік қосылыстардың «құрылым-қасиет» тәуелділігін болжау үшін жасанды интеллекттің заманауи әдістерін қолдану өзекті болып келеді. Мақалада жасанды иммунды жүйелердің модификацияланған алгоритмдерінің негізінде жасалған болжаудың Smart-технологиясы ұсынылған. Smart-технологиясының тиімділігін бағалау ақаудың туындау салдары мен себептерін талдауға арналған FMEA (Failure Mode and Effects Analysis) әдіс-төмесінің көмегімен жүзеге асырылады. Зерттеу жасалған FMEA моделі Smart-технологиясының этаптарының жұмыс істеуінің тәуекелін бағалауға арналған.

Түйінді сөздер: дәрілік қосылыстар, «құрылым-қасиет» тәуелділігін болжау, Smart-технология, ақпараттық сипаттауыштарды бөліп алу, жасанды иммунды жүйелердің модификацияланған алгоритмдері, сульфаниламидтер, тәуекелді бағалауға арналған FMEA моделі

DEVELOPMENT OF A METHOD OF SMART-TECHNOLOGY EFFICIENCY ASSESSMENT FOR PREDICTING MEDICINAL COMPOUNDS PROPERTIES AND ANALYSIS OF DATABASES USING MODERN SOFTWARE

Abstract: Nowadays, it is promising to develop innovative methods for creating new drugs with desired properties in order to reduce time and financial costs. The search for effective drug compounds is a complex, multi-stage process, in which it is necessary to process a huge amount of chemical data. The application of modern artificial intelligence methods to predict the structure-property dependence of drug compounds is relevant. The article presents the developed Smart-technology for prediction based on modified algorithms of artificial immune systems. Smart-technology effectiveness assessment is carried out using the FMEA (Failure Mode and Effects Analysis) methodology in order to analyze the causes and consequences of defects. An FMEA model has been developed for assessing the risks of the functioning of Smart technology stages. As an example, there are considered medicinal compounds of the sulfanilamide group.

Key words: medicinal compounds, prediction of the “structure-property” dependence, Smart-technology, selection of informative descriptors, modified algorithms of artificial immune systems, sulfonamides, FMEA risk assessment model

Введение

В связи с распространением опасных вирусов и ростом заболеваемости среди населения актуальна разработка отечественных технологий по созданию новых лекарственных соединений. Создание лекарственных препаратов является сложным процессом, сопряжённым с анализом большого объёма химической информации. В связи с чем ведущие фармакологические компании сотрудничают с организациями, занимающимися исследованиями в области искусственного интеллекта. Крупнейшие научные издательства, такие как Springer Nature, BMC, Elsevier, Scientific American, Apress и др. также занимаются поддержкой исследований в области искусственного интеллекта в медицине и здравоохранении.

Методы искусственного интеллекта широко применяются для решения задач выделения информативных дескрипторов, описывающих структуру химического соединения, распознавания образов и прогнозирования зависимости «структура-свойство» лекарственных соединений (Quantitative Structure Activity Relationship, QSAR). Перспективными являются алгоритмы роевого интеллекта (ПИ) [1], метод оптимизации серых волков (Grey wolf optimization, GWO) [2], метод опыления

цветов (Flower pollination algorithm, FPA) [3], генетические алгоритмы (Genetic algorithms, GA) [4] и т.д.

Особый интерес вызывает подход искусственных иммунных систем (ИИС) и модифицированные алгоритмы на его основе. Алгоритмы ИИС имитируют механизмы иммунной системы живого организма и обладают свойствами самоорганизации, памяти, возможности распознавания образов на границе классов. В настоящее время существует множество разновидностей ИИС и модифицированных алгоритмов на их основе. Например, в работе [5] представлен улучшенный алгоритм распознавания образов искусственной иммунной системой (Artificial Immune Recognition Systems, AIRS). Модификация названа AIRS3. В статье [6] рассматривается искусственная иммунная система для ассоциативной классификации при диагностике рака молочной железы. Работа [7] посвящена разработке гибридного алгоритма клональной селекции с модифицированной комбинаторной рекомбинацией и адаптивной мутацией для решения задач численной оптимизации.

По мере возникновения модифицированных алгоритмов искусственного интеллекта и

новых технологий анализа данных на их основе актуально применение современных методов оценки рисков для обеспечения точности прогнозов. В настоящее время существует множество эффективных методологий оценки рисков, таких как: FMEA (Failure Mode and Effects Analysis), PFMEA (Process Failure Mode and Effects Analysis), FTA (Fault Tree Analysis) и т.д. Данные методологии получили широкое распространение не только при оценке производственных рисков, но и в области здравоохранения.

Активно применяется в области фармакологии, медицины и производства лекарственных препаратов подход FMEA. Например, в работе [8] рассматривается методология FMEA при оценке рисков программного обеспечения для медицинских устройств. В работе предложена модификация классического подхода FMEA на основе подхода нечёткой логики и проведён сравнительный анализ полученных результатов. Исследования [9] посвящены использованию модели FMEA для снижения медицинских рисков. В работе [10] представлено приложение методологии FMEA при разработке новых лекарственных препаратов. Данный подход позволяет визуализировать сложные процессы в фармацевтических исследованиях, выявлять технологические дефекты и осуществлять количественную оценку нежелательных эффектов, напрямую влияющих на качество и эффективность.

Таким образом, анализ литературы доказывает актуальность разработки новых инновационных технологий для разработки лекарственных препаратов на основе подхода искусственных иммунных систем с применением современных методов оценки рисков.

1. Постановка задачи

В настоящее время разработка эффективных технологий для создания новых лекарственных соединений с заранее заданными свойствами является сложной задачей, находящейся на стыке нескольких дисциплин: фармакологии, биоинформатики, хемотрики и IT-технологий. При разработке подоб-

ных систем необходимо учитывать следующие требования:

- 1) возможность обработки баз данных дескрипторов лекарственных соединений высокой размерности;
- 2) объединение последних достижений в области фармакологии и искусственного интеллекта;
- 3) наличие модульной структуры и возможности расширения системы;
- 4) интеграция с современными программными продуктами по интеллектуальному анализу данных;
- 5) простота реализации;
- 6) оценка рисков возникновения дефектов при функционировании разработанной технологии поиска новых лекарственных соединений.

Постановка задачи исследований формулируется следующим образом: необходимо разработать методику оценки эффективности Smart-технологий прогнозирования зависимости «структура-свойство» лекарственных соединений на основе модифицированных алгоритмов ИИС для анализа баз данных дескрипторов лекарственных соединений с помощью подхода FMEA с использованием современных программных средств.

2. Методология FMEA

Рассмотрим подход FMEA для анализа видов и последствий дефектов. Данная методология впервые была разработана в США (NASA) и применялась в следующих областях: аэрокосмической, автомобильной промышленности, в области разработки стандартов (Германия), в медицине и фармакологической отрасли [9-10]. На рисунке 1 представлена таблица оценки рисков с помощью FMEA (Failure Mode and Effects Analysis).

В таблице описываются потенциальные сбои (рис. 1, столбцы 1-3) или дефекты системы и причины их возникновения для предупреждения ошибок (например, оценка работы информационной системы, оценка этапов технологического процесса или лабораторных исследований и т.д.) вводятся следующие определения (рис. 1, столбцы 4-6).

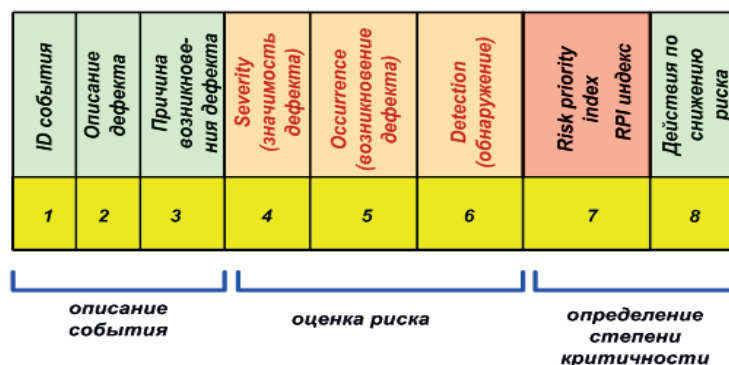


Рис. 1 – Методология FMEA

Определение 1. Значимостью дефекта (Severity) считается степень последствий отказа (степень значимости ранжируется в диапазоне).

Определение 2. Возникновение дефекта (Occurrence) это оценка, представляющая собой вероятность возникновения причины дефекта (критерий устанавливается в диапазоне).

Определение 3. Обнаружение дефекта (Detection) – это вероятность обнаружения дефектов, причин их возникновения и их последствий на всю систему (критерий ранжируется в диапазоне).

Определение 4. Индекс приоритета рисков (Risk priority index) представляет собой количественную оценку комплексного риска. В случае, если значение риска для нескольких отказов будет иметь одинаковое значение, наиболее серьёзным считается риск, у которого критерий Severity имеет наибольшее значение.

Данный коэффициент рассчитывается по формуле.

Определение 5. Матрица рисков (Risk Matrix) представляет собой механизм оценки степени критичности отказа. Данная оценка эффективна в случае неверной расстановки экспертами классов приоритетов рисков (рис.1, графы 4-6) и позволяет оценить критичность дефектов с разных точек зрения.

На рисунке 2 представлена визуализация матрицы рисков с учётом ранжирования критериев в диапазоне от 1 до 10. Матрица поделена на три зоны:

- «ЗОНА А» – низкая степень риска;
- «ЗОНА Б» – средняя степень риска;
- «ЗОНА В» – высокая степень риска.

Методология FMEA может успешно применяться для оценки разработанной Smart-технологии прогнозирования зависимости «структура-свойство» лекарственных соединений на основе подхода искусственных иммунных систем.

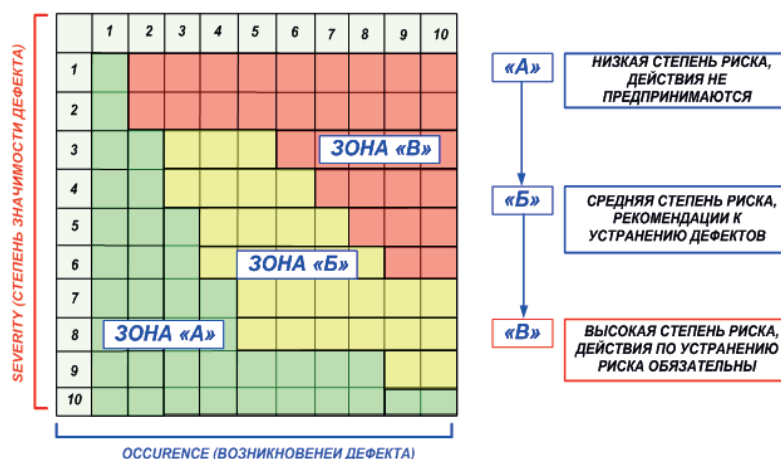


Рис. 2 – Матрица рисков методологии FMEA

3. Этапы Smart-технологии прогнозирования лекарственных соединений на основе

модифицированных алгоритмов ИИС

Разработана Smart-технология прогнозирования зависимости «структура-свойство» новых лекарственных соединений [11] с заданными свойствами. Так как в настоящее время эффективность результатов моделирования во многом зависит от качества исходной выборки данных, то необходима обработка баз данных дескрипторов лекарственных соединений для построения оптимального набора. Рассмотрим основные этапы функционирования Smart-технологии на примере лекарственных соединений сульфаниламидной группы.

1 этап. Формирование базы данных дескрипторов сульфаниламидов (Mol-Instincts, PubChem, ChemSynthesis, ABC chemistry и т.д.).

2 этап. Классификация сульфаниламидов по продолжительности действия:

1 класс – сульфаниламиды короткой продолжительности действия (менее 10 часов);

2 класс – сульфаниламиды средней продолжительности действия (10-24 часов);

3 класс – сульфаниламиды длительного действия (24-48 часов);

4 класс – сульфаниламиды сверхдлительного действия (более 48 часов).

Выбор целевых классов для решения задачи распознавания образов.

3 этап. Формирование оптимального набора дескрипторов сульфаниламидов.

3.1 Предварительная обработка данных (заполнение пропущенных данных, нормирование и т.д.).

3.2 Выделение информативных дескрипторов на основе современных оптимизационных алгоритмов: метода оптимизации серых волков (GWO), метода опыления цветов (FPA), генетического алгоритма (GA).

3.2 Оценка эффективности алгоритмов выделения информативных дескрипторов.

3.3 Построение оптимального набора данных по результатам анализа оптимизационных алгоритмов.

4 этап. Прогнозирование свойств лекарственных соединений сульфаниламидов на основе модифицированных алгоритмов ИИС.

5 этап. Оценка эффективности прогноза.

5.1 Расчёт оценки: Accuracy, %.

5.2 Расчёт оценки: Classification error, %.

5.3 Получение оценок по каждому классу: Precision, Recall, Sensitivity, Specify, f-measure.

5.4 Расчёт AUC характеристики.

5.5 ROC-анализ.

5.6 Получение оценки: confusion matrix.

6 этап. Отбор лекарственных соединений кандидатов с заданными свойствами.

Моделирование обработки данных на основе алгоритмов искусственного интеллекта осуществляется на базе современного пакета прикладных программ WEKA.

4. Оценка рисков функционирования Smart-технологии прогнозирования зависимости «структура-свойство» лекарственных соединений на основе методологии FMEA

Рассмотрим пример оценки рисков функционирования этапов Smart-технологии с помощью методологии FMEA. На рисунке 3 представлен фрагмент таблицы FMEA. Таблица содержит сведения об этапе функционирования Smart-технологии и потенциальном дефекте. Этапы кодируются по следующему шаблону (где – «step» этап, на котором возник дефект; порядковый номер этапа; порядковый номер дефекта, например – этап первый, дефект второй). Далее описывается наименование сбоя и возможная причина его возникновения.

Согласно критериям оценки, рассмотренным в разделе 3, экспертами осуществляется оценка степени критичности дефекта. Заполняются значения граф «Severity», «Occurrence» и «Detection». Рассчитывается значение индекса приоритета рисков RPI. Максимальное значение коэффициента:

$$RPI = 10 \times 10 \times 10 = 100$$

В случае, если несколько этапов имеют одинаковое значение RPI выбирается риск, у которого показатель «Severity» наибольший. Чем меньше значение коэффициента RPI, тем

менее опасен риск для функционирования Smart-технологии.

Последовательная структура оценки рисков на основе FMEA представлена на рисунке 4.

Таким образом, методология FMEA позволяет выявить потенциальные дефекты Smart-технологии, выявить их причину и осуществить оценку степени их критичности и влияния на весь процесс в целом.

№ Этапа	Описание дефекта	Причина возникновения дефекта	Severity (значимость дефекта)	Occurrence (возникновение дефекта)	Detection (обнаружение)	RPI индекс	ДЕЙСТВИЯ ПО СНИЖЕНИЮ РИСКА
S1.1	Отсутствие соединения по сети	Ошибка сети	5 высокая степень значимости	10 критическая степень	10 высокая степень	500	1. Проверка подключения к сети интернет. 2. Настройка и конфигурация сети. 3. Проверка провайдера.
S1.2	Падение сервера	Большая нагрузка	6 высокая степень значимости	10 критическая степень	10 высокая степень	600	1. Проведение профилактических работ на сервере. 2. Оптимальное распределение нагрузки.
S1.3	Сбой компьютера	Ошибка конфигурации	5 высокая степень значимости	3 низкая степень	3 низкая степень	45	1. Перезагрузка ПК. 2. Переустановка системных конфигураций. 3. Техническое обслуживание.
S1.4	Вирусная атака	Низкий уровень защиты ПК	5 высокая степень значимости	3 низкая степень	1 низкая степень	15	1. Установка антивирусной программы. 2. Настройка брандмауэра ПК.
S2.1	Ошибка ввода данных	Человеческий фактор	6 высокая степень значимости	5 средняя степень	3 низкая степень	90	1. Техническая проверка данных. 2. Уведомления об отсутствии записей.
S2.2	Отказ программы	Окончание лицензии	7 высокая степень	5 средняя степень	10 высокая степень	350	1. Мониторинг лицензионного соглашения.
...	...	...	...	...	...	...	...

Рис. 3 – Таблица оценки рисков FMEA для Smart-технологии прогнозирования зависимости «структура-свойство» сульфаниламидов

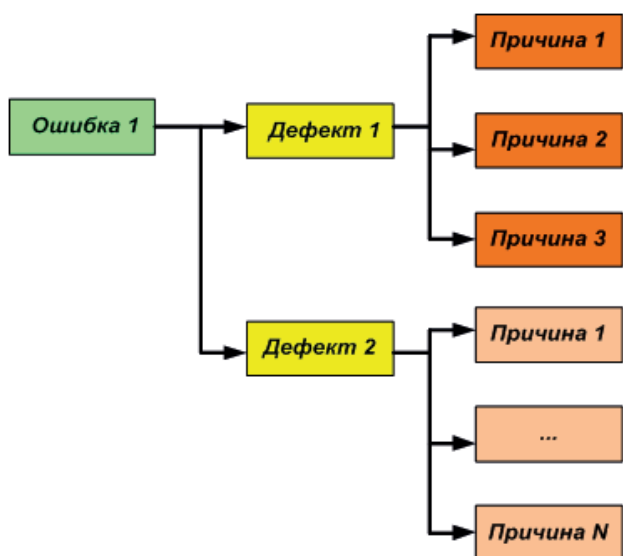


Рис. 4 – Структура оценки рисков на основе FMEA

6. Заключение

Разработанная Smart-технология для прогнозирования зависимости «структура-свойство» лекарственных соединений сочетает в себе последние достижения в области искусственного интеллекта и включает в себя современные оптимизационные алгоритмы, а также перспективное направление искусственных иммунных систем для поиска соединений кандидатов с заданными свойствами. Моделирование интеллектуальных алгоритмов на основе современных программных средств, таких как WEKA позволяет обрабатывать большие массивы химической информации, осуществлять визуализацию и архивирование данных. Оценка эффективности

разработанной Smart-технологии на основе методологии FMEA позволяет исключить потенциальные риски возникновения ошибок и повысить качество прогноза.

Работа выполнена по гранту КН МОН РК на тему: «Разработка и анализ баз данных для информационной системы прогнозирования зависимости «структура-свойство» лекарственных соединений на основе алгоритмов искусственного интеллекта» (2018-2020 гг.).

ЛИТЕРАТУРА

1. Chenye Qiu. A novel multi-swarm particle swarm optimization for feature selection // Genetic programming and evolvable machines. – Springer, 2019. - №20. – P. 503-529.
2. Zheng-Ming Gao, Juan Zhao. An Improved Grey Wolf Optimization Algorithm with Variable Weights // Computational Intelligence and Neuroscience. – 2019. – Vol.1. - №3. – P.1-13.
3. Mohamed Abdel-Basset, Laila A. Shawky. Flower pollination algorithm: a comprehensive review // Artificial Intelligence Review. – 2019. – Vol. 52. – P. 2533-2557.
4. Xi Wang. The Application of Genetic Algorithms in the Biological Medical Diagnostic Research // International journal of BIO automation. – 2016. – Vol.20. - №4. – P. 493 – 504.
5. Ilyes Jenhani, Zied Elouedi. Re-visiting the artificial immune recognition system: a survey and an improved version // Artificial Intelligence Review. – 2014. – Vol. 42. – P. 821–833.
6. David González-Patiño, Yenny Villuendas-Rey, Amadeo José Argüelles-Cruz, Oscar Camacho-Nieto, Cornelio Yáñez-Márquez. AISAC: An Artificial Immune System for Associative Classification Applied to Breast Cancer Detection // Applied science. – 2020. – Vol.10.- №2. – P. 515.
7. Weiwei Zhang, Kui Gao, Weizheng Zhang, Xiao Wang, Qiuwen Zhang & Hua Wang. A hybrid clonal selection algorithm with modified combinatorial recombination and success-history based adaptive mutation for numerical optimization // A hybrid clonal selection algorithm with modified combinatorial recombination and success-history based adaptive mutation for numerical optimization. – Applied Intelligence, 2018. – Vol.49. – P.819-836.
8. Khulan Batbayar, Márta Takács, Miklos Kozlovsky. Medical device software risk assessment using FMEA and fuzzy linguistic approach: case study // 11th IEEE International Symposium on Applied Computational Intelligence and Informatics. – 2016. – P.197-202.
9. Chiozza M.L., Ponzetti C. FMEA: a model for reducing medical errors // Clinica Chimica Acta; International Journal of Clinical Chemistry. – 2009. – Vol. 404 (1). – P. 75-78.
10. Hirotaka Inoue, Shu Yamada. Failure mode and effects analysis in pharmaceutical research // International Journal of Quality and Service Sciences. – 2010.- Vol. 2(3). – P. 369-382.
11. Samigulina G.A., Samigulina Z.I. Development of multi-agent technology for prediction of the «structure-property» dependence of drugs on the basis of modified algorithms of artificial immune systems // Proceedings of International Work Conference on Bioinformatics and Biomedical Engineering, IWBBIO 2018, April 25-27. –Spain, Granada. – 2018. – P. 1-2.
12. Самигулина Г.А., Самигулина З.И. Информационная система ведения научных исследований для прогнозирования зависимости «структура-свойство» лекарственных соединений на основе модифицированных алгоритмов искусственных иммунных систем // Проблемы информатики. – 2019. – № 3(44). – С. 31-46.