

КОЛИЧЕСТВЕННОЕ ПРЕДСКАЗАНИЕ ЦИТОТОКСИЧНОСТИ НАНОМАТЕРИАЛОВ

Дин Е. (ИТМО), Широкий Н.И. (ИТМО), Серов Н.С. (ИТМО), Разливина Ю.С. (ИТМО).
Научный руководитель – д.х.н., проф. Виноградов В.В. (ИТМО)

Введение. Машинное обучение изменило область органической химии, сделав возможным не только прогнозирование физико-химических свойств молекул, но и генерацию новых активных молекул на основе желаемых свойств [1]. Эти результаты позволяют сводить к минимуму количество чрезмерно трудоемких экспериментов. Однако, когда дело доходит до наночастиц, классические дескрипторы для работы с молекулами становятся недостаточными, поскольку они не несут никакой информации об их структуре и поверхностных свойствах. Токсичность наноматериалов, в свою очередь, является одним из ключевых параметров в исследованиях, касающихся медицины, который описывает их взаимодействие с живыми организмами. На данный момент существует множество моделей, предсказывающих цитотоксичность органических молекул [2–4], но в случае наночастиц представлены лишь несколько работ, которые частично восполняют этот пробел, предсказывая, токсичен наноматериал или нет. Таким образом, становится актуальным переход к количественному предсказанию цитотоксичности с использованием важных для ученых-экспериментаторов характеристик: условий протекания реакций, характеристик наночастиц и концентраций. Это позволит быстро разрабатывать наноматериалы для медицинских целей, проводя первичный отбор кандидатов *in silico*.

Основная часть. В этом исследовании были использованы сведения о 8076 образцах наночастиц. Обработка данных проводилась с использованием языка программирования Python 3. После обработки данных и устранения ошибок окончательный набор для обучения модели содержал информацию о 3087 образцах. Далее был проведен анализ данных, который включал в себя построение корреляционной матрицы, визуальных представлений распределения численных параметров и гистограммам, изображающих распределение категориальных переменных.

С целью определить оптимальный алгоритм, было проведено тестирование на 40 различных моделях с их исходными гиперпараметрами, в результате которого были выбраны пять лучших моделей, на которых далее было проведено обучение с тонкой настройкой гиперпараметров с помощью GridSearchCV. В результате основным алгоритмом был выбран LGBMRegressor из библиотеки lightgbm, который использует метод повышения градиента. Для интерпретации функциональности алгоритмов была использована информация о значимости признаков в модели и SHAP-анализ.

Выводы. Финальный алгоритм продемонстрировал точность R -квадрат=0,86 и среднеквадратичное отклонение (RMSE) равное 12,2% в численном предсказании клеточной токсичности *in vitro*. Оценка важности параметров моделей выявила, что ключевые факторы, воздействующие на токсичность, включают в себя концентрацию наночастиц в среде, диаметр частиц, тип клеточной линии и заряд частиц, что соответствует теоретическим данным. Дополнительно были построены кривые, демонстрирующие зависимость токсичности наноматериала, исключенного из общего набора данных в процессе обучения, от концентрации и времени. Эти кривые, обычно используемые при оценке цитотоксичности материалов, выявили отклонение от экспериментальных значений токсичности в пределах стандартного отклонения алгоритма, что свидетельствует о способности прогнозировать целевую переменную на новых данных и подтверждает возможность его применения для оценки токсичности наночастиц *in silico* на стадии разработки.

Список использованных источников:

1. J. M. Stokes, K. Yang, K. Swanson, W. Jin, A. Cubillos-Ruiz, N. M. Donghia, C. R. MacNair, S. French, L. A. Carfrae, Z. Bloom-Ackermann, V. M. Tran, A. Chiappino-Pepe, A. H. Badran, I. W. Andrews, E. J. Chory, G. M. Church, E. D. Brown, T. S. Jaakkola, R. Barzilay, J. J. Collins, *Cell*. 2020, 180, 688.
2. A. K. Sharma, G. N. Srivastava, A. Roy, V. K. Sharma, *Front Pharmacol* 2017, 8.
3. S. Zheng, J. Xiong, Y. Wang, G. Liang, Y. Xu, F. Lin, *J. Chem. Inf. Model.* 2020, 60, 3231.
4. A. P. Challa, A. L. Beam, M. Shen, T. Peryea, R. R. Lavieri, E. S. Lippmann, D. M. Aronoff, *Reprod. Toxicol.* 2020, 95, 148.