

**ВЫЧИСЛИТЕЛЬНЫЕ ПОДХОДЫ К ОЦЕНКЕ И МОДЕЛИРОВАНИЮ  
ВЗАИМОДЕЙСТВИЙ АПТАМЕРОВ С МАЛЫМИ МОЛЕКУЛАМИ**

**Еремеева М. А. (ИТМО)**

**Научный руководитель – кандидат химических наук, Серов Н. С.  
(ИТМО)**

**Введение.** Аптамеры представляют собой короткие одноцепочечные нуклеиновые кислоты, способные избирательно связывать молекулярные мишени. В отличие от антител, они могут быть получены и модифицированы *in vitro*, что делает их перспективными элементами диагностических и сенсорных систем [1]. Однако вычислительный дизайн аптамеров к малым молекулам остается недостаточно разработанным направлением. Существующие модели преимущественно ориентированы на белковые мишени, а отсутствие унифицированного корпуса данных затрудняет сопоставление результатов и объективную оценку методов.

Формирование воспроизводимой инфраструктуры данных и разработка обоснованных протоколов оценки являются необходимыми условиями для систематического развития методов машинного обучения в данной области.

**Основная часть.** В рамках работы сформирован унифицированный корпус взаимодействий «аптамер – малая молекула», включающий 6 413 аннотированных пар, 1 686 уникальных аптамеров (ДНК и РНК) и 1 041 лиганд. Более 30 % записей содержат количественные значения аффинности, что позволяет решать задачи классификации и регрессии.

С использованием данного корпуса решены следующие задачи:

- 1) Проведено сравнительное тестирование моделей машинного обучения при различных схемах разбиения данных (стратифицированное, отдельное по аптамерам, отдельное по молекулам). В задаче классификации наилучший результат достигнут при стратифицированном разбиении (ROC-AUC 0,952), тогда как при разбиении по молекулам показатель снижается до 0,886. В регрессионной постановке максимальное качество продемонстрировала модель LGBM, обеспечившая значение  $R^2$  до 0,668.
- 2) Выполнен анализ архитектур моделей. Установлено, что двухэпистолярные модели на основе сверточных сетей демонстрируют наиболее устойчивые результаты при переносе на новые данные. Эксперименты по абляции признаков показали существенное снижение качества при искажении как аптамерных, так и молекулярных представлений, что подтверждает необходимость совместного учета обеих модальностей.

Дополнительно предложен фрагментно-ориентированный метод генерации аптамеров, основанный на поэтапной сборке последовательностей из коротких мотивов с использованием обученной функции оценки в качестве направляющего сигнала.

**Выводы.** Создан унифицированный бенчмарк взаимодействий аптамеров с малыми молекулами, получены воспроизводимые базовые результаты для задач классификации и регрессии и предложен фрагментно-ориентированный метод генерации, формирующий основу для направленного машинного дизайна аптамеров.

**Список использованных источников:**

1. Ellington A.D., Szostak J.W. In vitro selection of RNA molecules that bind specific ligands // Nature. – 1990. – Т. 346. – С. 818–822.

Автор \_\_\_\_\_ Еремеева М. А.

Научный руководитель \_\_\_\_\_ Серов Н. С.