

УДК 004.89

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И МАШИННОЕ ОБУЧЕНИЕ ДЛЯ ГЕНЕРАЦИИ ДАТЧИКОВ МАЛЫХ МОЛЕКУЛ НА ОСНОВЕ АПТАМЕРНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Красовская З.С. (Университет ИТМО), Серов Н.С. (Университет ИТМО), Патра Х.К. (Университет ИТМО)

Научный руководитель – доктор химических наук,
профессор Виноградов В.В. (Университет ИТМО)

Введение. В современном мире пищевой промышленности требуется постоянное совершенствование методов контроля качества и безопасности продукции. Одним из ключевых аспектов этого процесса является обнаружение и анализ малых молекул, которые могут влиять на характеристики пищевых продуктов и, в определенных случаях, представлять потенциальную угрозу для здоровья потребителей. Аптамеры, как молекулярные зонды с высокой специфичностью к целевым молекулам, представляют собой перспективное решение для создания чувствительных датчиков, способных быстро и надежно обнаруживать нежелательные соединения в пищевых продуктах.

Разработка аптамеров, коротких последовательностей ДНК или РНК, с высокой аффинностью и селективностью к молекулярным мишеням, является актуальным направлением в современной биотехнологии и медицине. Однако существующие методы их создания, включая традиционный экспериментальный метод SELEX, часто требуют значительных временных и ресурсных затрат, ограничивая их широкое применение, особенно в разработке диагностических систем point-of-care.

Основная часть. Для решения этой проблемы предлагается новый подход, основанный на использовании искусственного интеллекта и методов машинного обучения для прогнозирования аптамерных последовательностей на основе физико-химических и структурных свойств молекулярных мишеней. Это позволит значительно сократить время и ресурсы, затрачиваемые на разработку аптамеров, и создать инструменты, способствующие более быстрой и эффективной генерации молекулярных зондов с желаемыми характеристиками.

В рамках проекта совместно с экспериментаторами мы собрали уникальную базу данных "аптамер-мишень". Она содержит информацию о структуре аптамеров, молекулярных мишенях, аффинности аптамеров и условиях экспериментов. Датасет охватывает 229 уникальных таргетных молекул и 664 единичных нуклеотидных последовательностей, из которых активными являются 500 аптамеров. Половина данных представляет собой количественные значения, с более чем 50% значений для K_d (константы диссоциации). Большинство аптамеров являются каноническими, с менее чем 10% модифицированных последовательностей, преимущественно длиной до 40 нуклеотидов, при этом 80% всех последовательностей относятся к ДНК.

Датасет разделен на три четкие группы с низким, средним и высоким значениями K_d , что делает его подходящим для машинного обучения. Большинство целевых молекул являются амфифильными и водорастворимыми, охватывая широкий диапазон молекул от малых до крупных. Также были получены SMILES модифицированных нуклеотидов из структуры с использованием PubChem Sketcher.

В дальнейшем планируется разработать как прямую модель, способную определять аффинность аптамеров к заданным молекулярным структурам с учетом условий экспериментов, так и генеративную нейросеть, способную генерировать высокоаффинные и селективные аптамеры под заданные молекулы. Кроме того, мы нацелены на анализ частоты встречаемости нуклеотидов, что позволит нам формулировать общие гипотезы о механизмах связывания аптамеров с конкретными молекулярными структурами.

Для удобной работы с аптамерами будет создан общедоступный веб-ресурс, объединяющий модели и инструменты, что ускорит процесс создания аптамеров для их применения в медицине и биотехнологии.

Выводы. В исследовании был предложен новый подход к разработке аптамеров в пищевой промышленности, использующий методы машинного обучения и искусственного интеллекта. В результате была создана и проанализирована уникальная база данных, содержащая 664 уникальные последовательности аптамеров. Произведена оценка качества и валидация собранных данных.

Работа выполнена при финансовой поддержке НИРМА ИТМО