

# **ИСПОЛЬЗОВАНИЕ РАЗВЕТВЛЕННЫХ ГЛУБОКИХ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ПРЕДСКАЗАНИЯ ВТОРИЧНОЙ СТРУКТУРЫ БЕЛКА**

**Богданова Е.А., Новоселецкий В.Н., Шайтан К.В.**

Кафедра биоинженерии, биологический факультет, Московский Государственный Университет имени М.В. Ломоносова; Россия, 119234, г. Москва, Ленинские горы, д. 1, стр. 12, E-mail: elizawea@yandex.ru

Определение структуры белков играет важнейшую роль в изучении их функциональной активности и для применения в фармацевтике и биотехнологии. Третичная структура белка во многом определяется его аминокислотной последовательностью, поэтому изучение сложной взаимосвязи структура-последовательность является одной из основных проблем современной биоинформатики. Информация о вторичной структуре белка позволяет существенно упростить решение данной проблемы. Существующие методы моделирования предсказывают вторичную структуру с точностью 72% (при классификации на 8 типов). Так как это значение относительно невысоко, продолжается разработка методов высокоточного предсказания вторичной структуры по аминокислотной последовательности.

В настоящей работе мы представляем метод машинного обучения, основанный на разветвленной глубокой свёрточной нейронной сети, для предсказания вторичной структуры в рамках классификации на 8 типов (Q8). Машинное обучение уже на протяжении нескольких лет применяется для решения подобных задач, однако, наш метод позволяет принимать во внимание при предсказании не только ближние, но и дальние взаимодействия аминокислотных остатков в первичной структуре. Это достигается путём введения во входные данные дополнительного окна сдвига, характеризующего аминокислотное окружение для каждого остатка. В качестве обучающих данных был использован набор CullPDB, содержащий информацию о консервативности и вторичной структуре аминокислотных последовательностей 6133 различных белков. Таким образом, параллельно происходит предсказание для данных с окнами разного размера, затем результаты интегрируются в один слой и осуществляется финальная классификация каждого аминокислотного остатка по типу вторичной структуры. Предложенный нами метод не уступает в точности другим существующим методам предсказания вторичной структуры. В дальнейшем планируется создание веб-сервиса, реализующего разработанный нами алгоритм.