



**СЕЧЕНОВСКИЙ
УНИВЕРСИТЕТ**
НАУК О ЖИЗНИ

ФГАОУ ВО Первый МГМУ
имени И. М. Сеченова Минздрава России

+7 (495) 609-14-00 доб. 20-63, 21-67
pr@sechenov.ru
www.sechenov.ru

Большая Пироговская ул., дом 2, стр. 4
119991, Москва, Россия

ПРЕСС-РЕЛИЗ

9 декабря 2019 года

Ученые рассказали о новейших технологиях анализа данных в медицине

*Исследователи Института персонализированной медицины Сеченовского университета проследили, как применяется машинное обучение и анализ больших данных в персонализированной медицине, в особенности в онкологии. Они указали на способ повышения гибкости алгоритмов и предложили собственный метод обработки данных. Работа [опубликована](#) в журнале *Frontiers in Oncology*.*

Хотя принципы и основы алгоритмов машинного обучения были сформулированы более полувека назад, более или менее широкое распространение в медицине они получили только в последние 20 лет. До этого алгоритмы изменили способ принятия решений в некоторых областях проектирования и разработки, банковской сфере, сельском хозяйстве, работе служб безопасности. Переломный для сферы здравоохранения момент произошел, когда появились способы получения больших объемов медицинских данных, сведений о ДНК (геном), белках (протеом), РНК (транскриптом) и соединениях, обеспечивающих обмен веществ в клетке (метаболом).

Работа алгоритмов машинного обучения основана на создании математической модели и ее корректировке на основе «тренировочного» массива данных (например, сведений о состоянии пациента, проведенном лечении и, главное, результатах терапии). Получившуюся модель используют для предсказания исхода в новых случаях («тестовый» массив). В медицине такие алгоритмы внедряются медленно из-за сложности и часто недостаточности данных – к примеру, современные методы секвенирования ДНК и РНК позволяют выделить намного больше признаков (мутаций отдельных генов), чем было обследовано пациентов, и для построения прогнозов требуется значительная обработка данных и сведение массивов из разных источников.

В области персонализированной медицины для предсказания результата той или иной стратегии лечения используются два типа данных: во-первых, сведения о поле, возрасте пациента, истории болезни, факторах риска, результатах обследований и лабораторных тестов, во-вторых, омиксные данные (состав и строение ДНК, РНК, белков, продуктов обмена веществ). Вторую группу данных можно сравнивать с выводами исследований, проведенных в клеточной культуре – с тем, как меняется активность (экспрессия) генов в ответ на применение



лекарства.

Работа с алгоритмами машинного обучения состоит из трех основных этапов. Первый из которых - подготовка данных и составление «тренировочного» датасета. Здесь исследователь сталкивается с той же недостаточностью данных (признаков больше, чем обследованных пациентов), и должен отбирать наиболее важные для анализа характеристики. Это могут быть гены, кодирующие конкретный фермент, или мутации, с наибольшей точностью выделяющие разные группы пациентов. Следующий этап это - применение алгоритма. Ученые могут либо выбрать подходящий метод из набора алгоритмов классификации и кластеризации, либо использовать сочетание нескольких из них. И в финале - обработка «тестового» массива данных и оценка результата.

Исследователи лаборатории клинической и геномной биоинформатики Сеченовского университета [предложили](#) способ отбора наиболее важных признаков для построения классификатора на основе методов машинного обучения. Новый алгоритм обеспечивает гибкий набор характеристик, «подстраивающийся» под каждый конкретный набор данных. Он позволяет значительно повысить точность уже существующих методов машинного обучения и позволяет начать применять их даже на небольших наборах медико-биологических данных. Это внушает надежду на скорую разработку десятков молекулярно-диагностических тестов, способных предсказывать эффективность разных видов онкотерапии для индивидуальных больных прежде всего исходя из молекулярных характеристик опухоли.

Исследование проводилось совместно со специалистами ИБХ РАН и компании OmicsWay Corporation (США).

Работа исследователей была представлена в рамках визита в Сеченовский университет членов Совета по повышению конкурентоспособности ведущих университетов Российской Федерации среди ведущих мировых научно-образовательных центров (Совет Проекта 5-100) 24 октября 2019 года, где состоялась презентация передовых научных разработок и проектов, созданных в университете в рамках «дорожной карты» Проекта 5-100. Проведение данного исследования осуществлялось за счет субсидии программы.

