

ПРОГНОЗИРОВАНИЕ СТЕПЕНИ ПОРАЖЕНИЯ ЛЕГКИХ ПРИ COVID-19 НА ОСНОВЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Васильев Юлий Алексеевич

Студент, 1 курс магистратуры

Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия

E-mail: yuliivasiliev@gmail.com

Научный руководитель — Петровский Михаил Игоревич

Всемирная организация здравоохранения 11 марта 2020 года объявила пандемию по заболеванию COVID-19, вызываемому вирусом SARS-CoV-2. В условиях пандемии коронавирусной инфекции компьютерная томография органов грудной клетки (далее - КТ) занимает важное место в диагностике заболевания.

Результаты КТ могут служить предикторами необходимости госпитализации в стационар и вероятности неблагоприятного исхода в отделении интенсивной терапии. Однако, существуют и недостатки КТ: низкая ресурсоемкость, высокие экономические затраты, проблемы радиационной безопасности пациентов и медицинского персонала.

В качестве альтернативного диагностического инструмента разработан способ оценки степени изменения легочной ткани при COVID-19 в экспресс режиме на основе физикальных и клинических признаков пациента.

Из полученных из средств автоматизации города данных была подготовлена выборка, содержащая данные пациентов, которым была проведена КТ. Выборка содержит осмотровые признаки (частота дыхания, температура тела и т.д.), анамнез (хронические заболевания, возраст, пол и т.д.) и результаты лабораторных анализов (включается общий анализ крови, общий анализ мочи и биохимический анализ).

Общий объем выборки составил 200 тысяч записей. Выборка была разделена на обучающую и тестовую часть. Обучающая выборка использована для построения прогнозных моделей. Тестовая – для оценки точности/достоверности работы модели.

В качестве прогнозных моделей рассматривались как классические алгоритмы машинного обучения: ансамбль деревьев решений (метод случайного леса), бустинг ансамбля деревьев решений (метод градиентного бустинга), так и алгоритмы, основанные на нейронных

сетях (рассматривался многослойный перцептрон, регуляризованная нейронная сеть, ансамбль нейронных сетей).

С использованием вышеописанных алгоритмов были построены модели классификации, обеспечивающие решение следующих прикладных задач:

1. Определение вероятности легкой степени тяжести (КТ 0-1) – пациент не требует госпитализации и может проходить лечение на дому
2. Определение вероятности тяжелой пневмонии (КТ 3-4)– пациент должен быть незамедлительно госпитализирован в стационар для проведения интенсивной терапии

При построении моделей был проведен анализ всех используемых факторов на степень значимости. Значимость фактов определялась алгоритмически, без участия человека.

Точность построенных моделей проверялась на тестовой выборке. В качестве метрик качества использовалась площадь под роковой кривой (roc auc). На задаче классификации КТ 0-1 качество модели достигает 0.9307, на задаче классификации КТ 3-4 качество достигает 0.9542.

В рамках взаимодействия с ГКБ №67 г. Москвы был разработан и внедрен в московскую систему ЕМИАС КТ-калькулятор (<https://ct.emias.mos.ru/>), доступ к которому открыт врачам не только Москвы, но и других регионов (<https://www.sobyenin.ru/kalkulyator-pnevmonii-ii-dlya-diagnostiki-covid-19>).

Благодаря разработанному ПО минимизируется необходимость использования КТ для оценки категории изменений легочной ткани при COVID-19. Определение степени изменений в легочной ткани в экспресс режиме позволяет осуществить своевременную госпитализацию в стационар пациентов с высокой вероятностью тяжелого течения пневмонии за счет сокращения диагностического этапа амбулаторного КТ-центра.

Литература

1. Biau G. Analysis of a random forests model. The Journal of Machine Learning Research. 2012.
2. Dmitrienko A., Koch G. G. Analysis of clinical trials using SAS: a practical guide. SAS Institute, 2017.
3. Lenz S. T. Glenn A. Walker, Jack Shostak: Common statistical methods for clinical research with SAS examples 3rd edition, 2012.