

ИСПОЛЬЗОВАНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПОИСКА ПРЕДИКТОРОВ ДЕПРЕССИВНЫХ РАССТРОЙСТВ У ПАЦИЕНТОВ С РАССЕЯННЫМ СКЛЕРОЗОМ В БЕЛОРУССКОЙ ПОПУЛЯЦИИ

Евстафьева В.А., Благочинная К.В., Федулов А.С.

*Белорусский государственный медицинский университет,
г. Минск, Беларусь*

Актуальность. Рассеянный склероз (РС) – тяжелое аутоиммунное нейродегенеративное и демиелинизирующее заболевание центральной нервной системы, приводящее к потере трудоспособности и прогрессирующей инвалидизации у лиц молодого возраста. Наряду с моторным и чувствительным дефицитом в клинической картине РС часто присутствуют нейропсихиатрические и когнитивные нарушения, которые значительно влияют на повседневную и профессиональную активность пациентов. Среди нарушений психической сферы чаще встречаются депрессивные эпизоды, гипомания, паническое расстройство, обсессивно-компульсивное расстройство. Распространенность депрессии у пациентов с РС в 3-5 раз выше, чем у здоровой популяции. Учитывая вышеизложенное, раннее прогнозирование депрессивных расстройств у пациентов с РС может значительно повлиять на диагностический и лечебный процесс. На данный момент наиболее современным и эффективным инструментом для построения систем прогнозирования и поиска предикторов является машинное обучение.

Цель исследования. Сравнить эффективность регрессионных и классификационных моделей машинного обучения для прогнозирования наличия у пациентов с РС депрессивных симптомов.



Материалы и методы. На базе ГУ «Минский научно-практический центр хирургии, трансплантологии и гематологии» был проведен ретроспективный анализ данных 57 пациентов с РС. Средний возраст составил 36 [32; 43] лет, структура по клиническим формам: ремитирующая форма – у 46 (80,7%) пациентов, вторично-прогрессирующая форма – у 11 (19,3%) пациентов. Факторами, которые использовались в качестве независимых переменных (предполагаемых предикторов), были клиническая форма РС, пол, возраст, образование, наличие группы инвалидности, продолжительность заболевания, применение плазмафереза, лечение препаратами, изменяющими течение РС (ПИТРС), связь со стрессом, результаты девятиколышкового теста (9-HPT, 9-Hole Peg Test) для правой и левой руки, результат теста 25-футовой ходьбы, результат по шкале расширенной оценки инвалидизации (EDSS), результаты когнитивных тестов по Монреальской шкале оценки когнитивных функций (MoCA, Montreal Cognitive Assessment), по таблицам Шульте (эффективность работы, степень вработываемости, психическая устойчивость), процент выполнения по символно-цифровому тесту (SDMT, Symbol Digit Modalities Test), результаты по психическому и физическому компоненту опросника по оценке качества жизни (SF-36, The Short Form-36). Для построения модели машинного обучения использовался язык программирования Python и различные классификационные и регрессионные алгоритмы. Для классификационных моделей в качестве вероятных исходов выступало наличие (балл по шкале депрессии Бека (Beck Depression Inventory, BDI) не меньше 10) или отсутствие депрессивных симптомов (балл по шкале BDI меньше 10), для регрессионных моделей в качестве зависимой переменной использовался количественный результат по шкале BDI, который после предсказания также бинарно классифицировался по наличию и отсутствию депрессивных симптомов. Для оценки эффективности работы моделей использовались такие метрики, как точность (precision), полнота (recall), f1 score и ROC AUC score.

Результаты и обсуждение. Наиболее эффективной классификационной моделью машинного обучения оказалась ансамблевая модель с помощью классификатора голосований, включающая в себя как модель «случайных лесов» (random forest), так и модель на основе метода опорных векторов (Support Vector Machine, SVM). Данная модель позволяет прогнозировать наличие или отсутствие депрессивных симптомов у пациентов с РС, основываясь на вышеперечисленных предикторах, с практически 100% точностью (precision=1, recall=1, f1 score=1, ROC AUC score=1). Наиболее эффективной регрессионной моделью машинного обучения оказалась модель на основе алгоритма «случайных лесов» (precision=0,71, recall=1, f1 score=0,83, ROC AUC score=0,86).

Наиболее значимыми предикторами депрессивного расстройства у пациентов с РС для самой эффективной классификационной модели оказались физический компонент SF-36 (15,1%), процент выполнения по SDMT (11,09%), продолжительность заболевания (8,99%), психический компонент SF-36 (8,01%), эффективность работы по таблицам Шульте (6,7%), результат теста МОСА (6,32%), возраст пациента (5,99%), тест 25-футовой ходьбы (5,43%), результат девятиколышкового теста для правой руки (5,15%).

Выводы. Построение систем прогнозирования и поиск предикторов могут быть осуществлены с использованием различных методов машинного обучения. Классификационные ансамблевые методы показывают лучшие результаты по сравнению с работой регрессионных моделей. Наиболее эффективная модель машинного обучения имеет высокую точность предсказаний, а наиболее значимыми предикторами



XXV ЮБИЛЕЙНЫЙ КОНГРЕСС С МЕЖДУНАРОДНЫМ УЧАСТИЕМ **ДАВИДЕНКОВСКИЕ ЧТЕНИЯ**

для ее работы являются физический компонент SF-36, процент выполнения SDMT, продолжительность заболевания, психический компонент SF-36, эффективность работы по таблицам Шульте, результат теста МОСА, возраст, тест 25-футовой ходьбы, результат 9-НРТ для правой руки.

Министерство здравоохранения РФ
Комитет по здравоохранению Санкт-Петербурга
Отделение медицинских наук РАН
Всероссийское общество неврологов
Ассоциация неврологов Санкт-Петербурга
Северо-Западный государственный медицинский университет
им. И.И. Мечникова
Военно-медицинская академия им. С.М. Кирова
Первый Санкт-Петербургский государственный медицинский
университет им. акад. И.П. Павлова
Санкт-Петербургская общественная организация
«Человек и его здоровье»



XXV ЮБИЛЕЙНЫЙ КОНГРЕСС С МЕЖДУНАРОДНЫМ УЧАСТИЕМ ДАВИДЕНКОВСКИЕ ЧТЕНИЯ

под редакцией
проф. Ключевой Е.Г., проф. Голдобина В.В.

МАТЕРИАЛЫ КОНГРЕССА

Санкт-Петербург
2023