

УДК 681

РАЗРАБОТКА СТАТИСТИЧЕСКИХ МОДЕЛЕЙ КЛАССИФИКАЦИИ СТЕПЕНЕЙ ТЯЖЕСТИ ТЕЧЕНИЯ ХРОНИЧЕСКОЙ ОБСТРУКТИВНОЙ БОЛЕЗНИ ЛЕГКИХ

Студенты гр. 4910 Бенчак Г. Ю., Жихарева В. Е., Леготкина Л. А., Шумилова М. Д.,
студент гр. 4911 Зайцева Е. А.

Кандидат мед. наук Бразовская Н. Г., кандидат мед. наук Калачева Т. П.
Сибирский государственный медицинский университет, Томск, Россия

Введение. В настоящее время хроническая обструктивная болезнь легких (ХОБЛ) является социально значимым заболеванием и занимает третье место по смертности в мире.

Большое внимание фтизиатров и пульмонологов привлекает определение факторов, влияющих на степень тяжести течения ХОБЛ.

Цель. Выявить признаки, ассоциированные с тяжестью течения ХОБЛ, и построить оптимальную модель классификации.

Материалы и методы. В исследование было включено 64 пациента мужского пола от 40 до 70 лет (56 ± 9 лет) с установленным диагнозом ХОБЛ. Набор данных включал в себя анамнестические и антропометрические признаки, показатели кардиогемодинамики (трансторакальная эхокардиография (ЭхоКГ)), показатели, характеризующие функцию дыхательной системы (спирометрия), включая данные инструментальных методов исследования (компьютерная томография высокого разрешения).

Статистический анализ данных проводился с помощью пакета программ RStudio версии 4.3.1 и пакета прикладных программ Statistika 13.3.

Результаты. В зависимости от степени тяжести течения ХОБЛ по классификации GOLD 2014 обследуемые были разделены на две группы: в первую вошли 43 пациента с умеренной степенью тяжести ХОБЛ, во вторую – 21 пациент с тяжелой степенью тяжести ХОБЛ. Путем оценки межгрупповых сравнений были выделены ассоциированные с течением заболевания признаки, которые использовались в качестве предикторов для построения модели: результаты теста шестиминутной ходьбы (ТШХ), эхокардиографические параметры, включая давление в легочной артерии, а также данные компьютерной томографии высокого разрешения. Для построения моделей были использованы алгоритмы машинного обучения: бинарная логистическая регрессия и метод дерева решений.

В уравнении логистической регрессии были использованы предикторы: ТШХ, степень регургитации митрального клапана, наличие «воздушных ловушек», повышения давления в легочной артерии, признаков панлобулярной эмфиземы.

При помощи ROC-анализа выбрана оптимальная точка отсечения – 0,7. Чувствительность модели составила 77 %, специфичность 75 %, общий процент правильных решений – 77 %.

Для построения модели дерева решений были использованы предикторы: ТШХ, наличие «воздушных ловушек» и признаков панлобулярной эмфиземы.

Построенное дерево решений представляет собой в структурированном виде следующие правила классификации вида «если – то»: если $\text{ТШХ} < 335$, то группа 2; если $\text{ТШХ} \geq 335$, нет панлобулярная эмфиземы и «воздушных ловушек», то группа 1; если $\text{ТШХ} \geq 335$, есть панлобулярная эмфиземы и «воздушные ловушки», то группа 2.

Чувствительность модели составила 76 %, специфичность 88 %, общий процент правильных решений – 84 %.

Выводы. В результате проведенного исследования была выбрана наиболее оптимальная модель классификации – метод дерева решений, который обладает более высокими качествами классификации, чем модель логистической регрессии.

При построении моделей с помощью разных методов получен схожий набор признаков, ассоциированных со степенью тяжести течения ХОБЛ. Устойчивость результата подтверждает, что сформированный набор характеристик имеет прочные ассоциации с тяжестью течения заболевания.