

DOI: 10.25205/978-5-4437-1691-6-32

**ПРЕДСКАЗАТЕЛЬНАЯ СПОСОБНОСТЬ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ
НА ОСНОВАНИИ КЛИНИЧЕСКИХ ПОКАЗАТЕЛЕЙ АТЕРОСКЛЕРОЗА***

**PREDICTIVE CAPABILITY OF MACHINE LEARNING MODELS
BASED ON CLINICAL INDICATORS OF ATHEROSCLEROSIS**

Н. А. Орехов, Д. Д. Бородко, Е. М. Плешко, А. В. Омельченко

*Лаборатория молекулярно-генетического моделирования инфламейджинга,
НИИ общей патологии и патофизиологии, Москва*

N. A. Orekhov, D. D. Borodko, E. M. Pleshko, A. V. Omelchenko

*Laboratory of Molecular Genetic Modeling of Inflammaging,
Institute of General Pathology and Pathophysiology, Moscow*

✉ www.fuper@gmail.com

Аннотация

Атеросклероз — хроническое заболевание, входящее в тройку наиболее частых причин смертности в РФ. Цель работы — создать модели машинного обучения для прогнозирования атерогенеза на период 3–12 месяцев по клиническим показателям, в том числе по ЛНП-ЦИК. Данные 101 пациента показали, что метод дерева решений обеспечивает лучшую оценку прогноза (специфичность 98,9 %, точность 94,1 %).

Abstract

Atherosclerosis is a chronic disease and one of the top three causes of mortality in Russia. The aim of this study is to create machine learning models to predict atherogenesis over a period of 3 to 12 months based on clinical indicators, including LDL-CIC. Data from 101 patients demonstrated that the decision tree method provides the best predictive assessment (specificity 98.9 %, accuracy 94.1 %).

Атеросклероз — хроническое заболевание артерий, приводящее к серьезным сердечно-сосудистым заболеваниям, включая инфаркт миокарда и инсульт. Циркулирующие иммунные комплексы, содержащие липопротеины низкой плотности (ЛНП-ЦИК), играют ключевую роль в инициации ранних стадий атерогенеза. Ранее была показана связь между уровнем ЛНП-ЦИК и толщиной интима-медиального слоя аорты (ТИМС). Установлено, что даже низкие уровни ЛНП-ЦИК способны предсказывать отсутствие утолщения ТИМС. Разработка и совершенствование моделей для прогнозирования тяжести атеросклероза является важной задачей для ранней профилактики этого заболевания.

Целью данной работы является создание математических моделей на основе машинного обучения, способных прогнозировать атерогенетические изменения у человека на временном промежутке от трех месяцев до года, с использованием простых диагностических показателей. Для исследования мы использовали выборку из 101 пациента-мужчины в возрасте от 37 до 74 лет с бессимптомным течением атеросклероза на момент первичного обследования. У субъектов измерялось более 30 морфо-анатомических, физиологических и биохимических показателей, которые затем были использованы в качестве предикторов для предсказательной модели.

Толщина интима-медиального слоя сонной артерии измерялась у каждого пациента на начальном этапе исследования, а также через 3, 6, 9 и 12 месяцев. На каждом этапе определялся статус пациента («норма», «предрасположенность», «атеросклероз») по критериям Американского общества эхокардиографов [1, 2]. С помощью языка программирования Python и модуля sklearn строились прогностические регрессионные модели зависимости статуса здоровья от возраста пациента и уровня концентрации циркулирующих иммунных комплексов (ЦИК), а также только от концентрации ЦИК в крови на начальном этапе.

1. Модель логистической регрессии с алгоритмом кросс-валидации.
2. Модель логистической регрессии с разделением выборок в соотношении 3 : 1.
3. Модель деревьев решений.

* Исследование выполнено при поддержке РФФ (проект № 24-65-00027).

© Н. А. Орехов, Д. Д. Бородко, Е. М. Плешко, А. В. Омельченко, 2024

Аналогичные модели строились для прогноза ухудшения статуса здоровья (возникновение предрасположенности к атеросклерозу у здоровых пациентов и развитие атеросклероза у предрасположенных). Для каждой модели проводился ROC-анализ, а также оценивались показатели правильности, неправильности, точности, чувствительности, F-Score и специфичности для выбора наилучшей модели.

В результате работы были созданы математические модели, способные с высокой точностью предсказывать развитие атеросклероза у пациентов на основании их возраста и уровня ЦИК в крови в период от 3 месяцев до года. Интересно, что ни возраст, ни уровень ЦИК в крови по отдельности не являются надежными прогностическими факторами, однако их совместное использование в моделях обеспечивает необходимую точность и специфичность. Наиболее эффективным методом оказался алгоритм на основе дерева решений, продемонстрировавший специфичность 98,9 %, точность 94,1 % и AUC 81,6 % для прогноза атеросклероза на полугодовой период.

Литература

1. Groot de E., Leuven van S. I., Duivenvoorden R. et al. Measurement of carotid intima-media thickness to assess progression and regression of atherosclerosis // *Nat. Clin. Pract. Cardiovasc. Med.* 2008. Vol. 5. P. 280–288.
2. Stein J. H., Korcarz C. E., Hurst R. T. et al. American Society of Echocardiography Carotid Intima-Media Thickness Task Force. Use of carotid ultrasound to identify subclinical vascular disease and evaluate cardiovascular disease risk: a consensus statement from the American Society of Echocardiography Carotid Intima-Media Thickness Task Force. Endorsed by the Society for Vascular Medicine // *J. Am. Soc. Echocardiogr.* 2008. Vol. 21. P. 93–111.