

ИСПОЛЬЗОВАНИЕ РЕГРЕССИИ LASSO ДЛЯ МИНИМИЗАЦИИ ТЕСТИРУЕМЫХ ПАРАМЕТРОВ КРОВИ В ОПРЕДЕЛЕНИИ СЕРДЕЧНО-СОСУДИСТЫХ ЗАБОЛЕВАНИЙ

А.Ю. Данилин, Н.В. Силивончик, магистранты
Научный руководитель – **В.Т. Чешевик**, к.б.н., доцент
Полесский государственный университет

Сердечно-сосудистые заболевания (ССЗ) остаются основной причиной заболеваемости и смертности во всем мире. Традиционные методы диагностики сердечно-сосудистых заболеваний часто являются дорогостоящими и требуют много времени. Биохимические изменения, вызванные болезнью, могут существенно повлиять на различные аспекты биоанализа. Однако врачи часто сосредотачиваются на существенно аномальных параметрах, потенциально упуская из виду значительное количество других данных тестов и взаимосвязи между лабораторными параметрами, что может привести к недооценке диагностического потенциала этих тестов [1]. Поэтому крайне важно изучать референтный диапазон и характеристики вариации гематологических и биохимических показателей для раннего выявления предотвращаемых факторов риска и ранней диагностики сердечно-сосудистых заболеваний, особенно показателей, связанных с метаболическим здоровьем, чтобы помочь врачам в выявлении сердечно-сосудистых заболеваний на ранней стадии. Исходя из изученной литературы можно предположить, что для выявления наиболее критичных параметров ССЗ может подойти метод регрессии LASSO [2].

Алгоритмы на основе машинного обучения широко используются в принятии клинических решений. Из них оператор наименьшего абсолютного сжатия и выбора LASSO является одним из наиболее часто используемых алгоритмов, и его клиническая эффективность была продемонстрирована ранее [3].

Регрессия LASSO – это метод линейной регрессии, который включает регуляризацию для предотвращения переобучения модели. Этот подход особенно полезен в ситуациях, когда количество предикторов (исследуемых параметров) значительно превышает количество наблюдений, или когда предикторы сильно взаимоскоррелированы [4].

Регрессия LASSO добавляет штраф к функции потерь, который пропорционален абсолютной величине коэффициентов. Это приводит к "сжатию" коэффициентов, и некоторые из них могут стать равными нулю, что позволяет автоматически отбирать наиболее значимые переменные. Следовательно, регрессия LASSO помогает выявить наиболее значимые предикторы [5, 6].

В связи с этим целью настоящей работы является изучение возможности применения метода регрессии LASSO для сокращения тестируемых параметров крови, необходимых для диагностики и раннего определения ССЗ.

Материалом для исследования явилась сыворотка крови 65 человек: 30 условно-здоровых людей (УЗЛ) в качестве контрольной группы и 35 пациентов с ССЗ. Предметом исследования являются 42 параметра крови, которые были разделены на 3 группы (биохимические показатели – 24, гематологические показатели – 10, метаболические показатели – 8). Статистическую обработку данных осуществляли с помощью пакета программ R статистики (v4.3.3). С использованием пакета stats (v4.3.3) проводили первичную обработку данных. С использованием пакета caret (v4.3.3) проводили построение регрессионной модели машинного обучения с регуляризацией по методу LASSO. Модель принималась достоверной при уровне значимости $p < 0,05$.

В результате исследования было построено 4 модели с использованием регрессии LASSO. Три модели с использованием каждой из 3 групп показателей (биохимических, гематологических и метаболических) и одна модель с совместным использованием показателей всех групп. По результатам исследования было выявлено, что модели, построенные с использованием метаболических показателей, а также модель с совместным использованием показателей всех групп не имели статистически значимых отличий от случайных предсказаний ССЗ, при этом модели построенные с использованием биохимических и гематологических показателей крови показали хорошие показатели согласия (0,722 и 0,998, соответственно) и уровнем статистической значимости ($p=0,027$ и $p=0,0005$, соответственно).

Результаты исследований показывают, что для диагностики и раннего определения ССЗ целесообразнее использовать биохимические и гематологические показатели крови. Однако данное исследование имело ряд ограничений, связанных с количеством выборки и методов построения моделей, что может повлиять на обобщаемость выводов.

Исследования проведены при финансовой поддержке Министерства образования Республики Беларусь (договор № 65 от 05.05.2021) в рамках ГПНИ «Биотехнологии-2» (Рег. № НИР 20241017).

Список использованных источников

1. Cheng, X. Application of Artificial Intelligence in Cardiovascular Medicine / X. Cheng [et al.] // Compr Physiol. – 2021. – Vol. 11, №4. – P. 2455–2466.
2. Rabbani, N. Applications of machine learning in routine laboratory medicine: Current state and future directions / N. Rabbani [et al.] // Clin Biochem. – 2022. – Vol. 103, № 1. – P. 1–7.
3. Ichimasa, K. Artificial intelligence may help in predicting the need for additional surgery after endoscopic resection of T1 colorectal cancer / K. Ichimasa [et al.] // Endoscopy. – 2018. – Vol. 50, № 3. – P. 230–240.
4. Ronzio, L. Has the Flood Entered the Basement? A Systematic Literature Review about Machine Learning in Laboratory Medicine / L. Ronzio [et al.] // Diagnostics. – 2021. – Vol. 11, № 2. – P. 372–384.
5. Panjiyar, B. K. A Systematic Review: Do the Use of Machine Learning, Deep Learning, and Artificial Intelligence Improve Patient Outcomes in Acute Myocardial Ischemia Compared to Clinician-Only Approaches? / B. K. Panjiyar [et al.] // Cureus. – 2023. – Vol.15, № 8. –P. 245–257.
6. Chen, L. The emerging roles of machine learning in cardiovascular diseases: a narrative review / L. Chen [et al.] // Ann Transl Med. – 2022. – Vol. 10, № 10. – P. 610–622.