



<https://doi.org/10.29001/2073-8552-2023-38-2-89-97>
УДК УДК 616.13-004.6-039.36-037

Два подхода к построению моделей прогноза риска прогрессирующего атеросклероза

Н.Г. Ложкина^{1, 2, 5}, Ю.Е. Воскобойников³, В.Н. Копылов⁴, О.М. Пархоменко⁵,
М.И. Воевода¹

¹ Федеральный исследовательский центр фундаментальной и трансляционной медицины,
630117, Российская Федерация, Новосибирск, ул. Тимакова, 2

² Новосибирский национальный исследовательский государственный университет,
630090, Российская Федерация, Новосибирск, ул. Пирогова, 1

³ Новосибирский государственный архитектурно-строительный университет (Сибстрин),
630008, Российская Федерация, Новосибирск, ул. Ленинградская, 113

⁴ Сибирский региональный научно-исследовательский гидрометеорологический институт (СибНИГМИ),
630099, Российская Федерация, Новосибирск, ул. Советская, 30

⁵ Городская клиническая больница № 1,
630047, Российская Федерация, Новосибирск, ул. Залесского, 6

Аннотация

Прогрессирующий или ускоренный атеросклероз сопровождается неблагоприятными клиническими исходами. Изучение и понимание этого процесса, создание персонализированного метода оценки риска и прогноза данного заболевания необходимы для оптимизации подходов к лечению и профилактике.

Цель исследования: сравнить два подхода к созданию прогнозной модели риска прогрессирующего атеросклероза: метод нелинейной регрессионной модели логистического типа и метод свободной кроссплатформенной системы визуального программирования Orange.

Материал и методы. В ретроспективное когортное исследование были включены 202 пациента с подтвержденной ишемической болезнью сердца (ИБС): 147 мужчин и 55 женщин. Средний возраст пациентов составил $53,3 \pm 7,16$ лет. В первую группу вошли пациенты, перенесшие инфаркт миокарда или нестабильную стенокардию, экстренное артериальное стентирование, инсульт, тромбоз периферических артерий, критическую ишемию и ампутацию нижней конечности в течение 2 лет до включения в исследование. Пациенты в группе сравнения не имели этих событий. Прогнозные модели влияния различных исследуемых параметров на вероятность быстрого прогрессирования атеросклероза строились с помощью факторного и корреляционного анализа и свободной кроссплатформенной системы визуального программирования Orange.

Результаты. Предложенные авторские подходы оценки риска прогрессирующего атеросклероза обладают хорошей прогностической точностью: ($K_{\text{чув.}} = 94,1$; $K_{\text{спец.}} = 97,0$; $K_{\text{точ.}} = 95,5$) для регрессионной модели и 0,950 (95,0%) для модели машинного обучения. Однако построение регрессионной модели – более сложная процедура по сравнению со вторым подходом, где выбор информативных показателей для модели прогноза делается системой Orange. Тем не менее, изложенные два подхода могут успешно дополнять друг друга, позволяя строить более точные прогнозные модели риска.

Заключение. Предложенные авторские подходы оценки риска прогрессирующего атеросклероза обладают хорошей прогностической точностью.

Ключевые слова:	нелинейная регрессионная модель, модель машинного обучения, Orange, прогрессирующий атеросклероз.
Конфликт интересов:	авторы заявляют об отсутствии конфликта интересов.
Прозрачность финансовой деятельности:	исследование выполнено с использованием оборудования ЦКП «Протеомный анализ», поддержанного финансированием Минобрнауки России (соглашение № 075-15-2021-691).
Соответствие принципам этики:	исследование было выполнено в соответствии со стандартами надлежащей клинической практики и принципами Хельсинкской декларации и одобрено этическим комитетом НИИ терапии и профилактической медицины – филиала Института цитологии и генетики СО РАН.

Для цитирования:

Ложкина Н.Г., Воскобойников Ю.Е., Копылов В.Н., Пархоменко О.М., Воевода М.И. Два подхода к построению моделей прогноза риска прогрессирующего атеросклероза. *Сибирский журнал клинической и экспериментальной медицины*. 2023;38(2):89–97. <https://doi.org/10.29001/2073-8552-2023-38-2-89-97>.

Two approaches to modeling the risk of progressive atherosclerosis

Natalia G. Lozhkina^{1, 2, 5}, Yurii E. Voskoboynikov³, Vasily N. Kopylov⁴,
Olga M. Parkhomenko⁵, Mikhail I. Voevoda¹

¹ Federal Research Center for Fundamental and Translational Medicine,
2, Timakova str., Novosibirsk, 630117, Russian Federation

² Novosibirsk National Research State University,
1, Pirogova str., Novosibirsk, 630090, Russian Federation

³ Novosibirsk state university of architecture and civil engineering (Sibstrin),
113, Leningradskaya str., Novosibirsk, 630008, Russian Federation

⁴ Siberian Regional Research Hydrometeorological Institute,
30, Sovetskaya str., Novosibirsk, 630099, Russian Federation

⁵ City Clinical Hospital No 1,
6, Zalesskogo str., Novosibirsk, 630047, Russian Federation

Abstract

Progressive or accelerated atherosclerosis is accompanied by unfavorable clinical outcomes. Studying and understanding this process and creating a personalized method for assessing the risk and prognosis of this disease are necessary to optimize approaches to treatment and prevention.

Aim: To compare two approaches to the creation of prognostic risk model of progressive atherosclerosis: non-linear regression model of logistic type and free cross-platform visual programming system Orange method.

Material and Methods. The retrospective cohort study included 202 patients with confirmed coronary heart disease: 147 men and 55 women. The mean age of the patients was 53.3 ± 7.16 years. Group 1 included patients with myocardial infarction or unstable stenocardia, emergency arterial stenting, stroke, peripheral arterial thrombosis, critical ischemia and lower extremity amputation within 2 years before inclusion in the study. Patients in the comparison group did not have these events. Predictive models of the influence of different studied parameters on the probability of rapid progression of atherosclerosis were built using factor and correlation analysis and free cross-platform Orange visual programming system.

Results. The authors' suggested approaches to the evaluation of the risk of progressive atherosclerosis have a good prognostic accuracy (sensitivity 94.1, specificity 97.0 and accuracy 95.5 coefficients, respectively) for the regression model and 0,950 (95,0%) for the machine learning model. However, the construction of the regression model is a more complex procedure compared to the second approach, where the choice of informative indicators for the prediction model is made by Orange. Nevertheless, the above two approaches can successfully complement each other, allowing to build more accurate predictive risk models.

Conclusion. The proposed authors' approaches to assessing the risk of progressive atherosclerosis have a good prognostic accuracy.

Keywords:

nonlinear regression model, machine learning model, Orange, progressive atherosclerosis.

Conflict of interest:

the authors declare no conflict of interest.

Financial disclosure:

the study was performed using the equipment of the Proteomic Analysis Center supported by funding from the Russian Ministry of Education and Science (Agreement No. 075-15-2021-691).

Adherence to ethical standards:

the study was performed in accordance with the standards of good clinical practice and the principles of the Declaration of Helsinki and was approved by the Ethics Committee of Research Institute of Therapy and Preventive Medicine – a branch of the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences.

For citation:

Lozhkina N.G., Voskoboynikov Y.E., Kopylov V.N., Parkhomenko O.M., Voevoda M.I. Two approaches to modeling the risk of progressive atherosclerosis. *The Siberian Journal of Clinical and Experimental Medicine*. 2023;38(2):89–97. <https://doi.org/10.29001/2073-8552-2023-38-2-89-97>.

Введение

Синдром быстро прогрессирующего или ускоренного атеросклероза, в отличие от спонтанного течения процесса, привлекает внимание исследователей в последние 2–3 десятилетия [1]. В его развитии помимо классических факторов риска играют роль и ряд других медиаторов: воспаления, кальцификации, остеопонтина [2]. Пациенты с быстро прогрессирующим атеросклерозом подвержены наибольшему риску развития различных ишемических событий, требующих постоянного медицинского внимания и специфической, более агрессивной липидснижающей и антитромботической терапии [3].

Цель исследования: сравнить два подхода к созданию прогнозной модели риска прогрессирующего атеросклероза: метод нелинейной регрессионной модели логистического типа и метод свободной кроссплатформенной системы визуального программирования Orange.

Материал и методы

Исследование включало 202 пациента с ишемической болезнью сердца (ИБС): 147 мужчин и 55 женщин (ретроспективное когортное методом параллельных групп). Средний возраст пациентов составил $53,3 \pm 7,16$ лет. Диагноз ИБС был установлен в соответствии с набором критериев, разработанных Европейским обществом кардиологов и Американским колледжем кардиологов (2019, 2021) [4, 5]. Ретроспектива охватывала 2 года с момента включения пациентов в исследование (период с января 2019 по январь 2020 гг.) с определением 100 клинических, инструментальных, лабораторных параметров, которые потенциально влияют на возникновение и прогрессирование атеросклероза (по данным анализа литературы).

Данные обследования проводились и оценивались в 2 точках: во время 1-го предшествующего ишемического события в течение 2-летнего ретроспективного периода и в день включения в настоящее исследование.

202 пациента были разделены на 2 группы. В первую – активную (100 человек) включали пациентов, у которых был инфаркта миокарда (ИМ) 1-го типа [6] (54 человека) в течение 2 лет до начала исследования или диагноз ИБС был установлен по данным селективной коронарной ангиографии: наличие стеноза просвета как минимум двух коронарных артерий на 50% и более (46 человек) и которые имели дополнительно два (или более) сердечно-сосудистых события из следующих: ИМ или нестабильная стенокардия, экстренное артериальное стентирование, инсульт, тромбоз периферических артерий, критическая ишемия и ампутация нижних конечностей. Сочетание двух или более из этих сердечно-сосудистых событий, произошедших в течение 2 лет, указывает на быстрое прогрессирование атеросклероза у этих пациентов.

Данные клинические состояния свидетельствуют о нестабильном, или прогрессирующем, течении заболевания; это критерии так называемого «истинного прогрессирования», т. е. сахарный диабет и фибрилляция предсердий также могут привести к вышеперечисленным состояниям, но они имеют под собой несколько иную причину, не связанную с простым повреждением бляшки. Основным документом, пожалуй, следует считать рекомендацию ESC/EAS по применению ингибиторов PCSK9 [7].

В других источниках прогрессирующий атеросклероз описан в категории экстремального сердечно-сосудистого риска. Наиболее подробный обзор по этой проблеме

сделан Р. Shah и соавт. (2015) [1]. Российские ученые М.В. Биленко и соавт. (2010), Т.И. Арефьева и соавт. (2015, 2016) обосновали 3 способа предсказания прогрессирующего атеросклероза, основанные на маркерах воспаления, концентрации Интерлейкин-10-продуцирующих Т-лимфоцитов, способности макрофагов а) продуцировать активные формы кислорода, оцененные по величине люминол-зависимой хемилюминесценции, б) окислять липопротеины низкой плотности (ЛПНП), оцененные по накоплению в них продуктов, реагирующих с тиобарбитуровой кислотой, в) поглощать ЛПНП, оцененные по накоплению в макрофагах общего холестерина и г) по сохранению 50–70% жизнеспособных макрофагов в процессе 1 ч инкубации в безбелковой субстрат-дефицитной среде [8–10]. Ю.И. Рагино в 2019 г. описала лабораторные и биохимические маркеры нестабильной атеросклеротической бляшки, играющей важную роль в возникновении повторных сердечно-сосудистых событий [11].

Вторая группа (группа сравнения) включала 102 пациента с подтвержденной ИБС аналогичным образом (55 человек имели в прошлом только один ИМ 1-го типа, а 45 – ИБС, подтвержденную селективной коронарной ангиографией соответственно), у которых за 2 года до включения в исследование не было сердечно-сосудистых событий, что свидетельствовало о так называемом спонтанном течении атеросклероза. Все пациенты прошли клиническое и инструментальное обследование по следующей программе: электрокардиография, эхокардиография, холтеровское мониторирование, исследование воспалительных цитокинов.

Существенной особенностью прогнозных моделей риска ускоренного развития атеросклероза является вычисление вероятности случайного события, которое заключается в ускоренном развитии атеросклероза. Следовательно, прогнозная модель вычисляет величину (вероятность), которая может меняться в интервале от 0 до 1. При этом если вероятность меньше некоторой пороговой величины, то делается вывод о низком риске, если равно или больше, то о высоком риске. Ниже будут изложены два подхода к построению прогнозных моделей ускоренного развития атеросклероза.

Первый подход содержит построение прогнозной модели на основе нелинейной регрессионной модели логистического типа. При построении регрессионных моделей, устанавливающих связь между переменными (факторами) исследуемого процесса используют понятия зависимой и независимых переменных. Под независимой переменной подразумевают некоторый фактор, воздействующий на другой наблюдаемый фактор, который будем называть зависимой переменной. В данном исследовании зависимой переменной (обозначим Y) является факт ускоренного развития атеросклероза. Примерами независимых переменных являются: привычка курения, наличие ожирения и т. д. Так как зависимая переменная Y принимает только два значения (1 – высокий риск атеросклероза, т. е. неблагоприятный исход (НИ), 0 – низкий риск, т. е. благоприятный исход (БИ)), поэтому для описания исследуемого процесса в качестве математической модели была принята регрессионная логистическая модель, вычисляющая вероятность неблагоприятного исхода.

В общем виде эта регрессионная модель описывается следующим выражением:

$$P(X) = \frac{1}{1 + e^{-Z(X)}}, \quad (1),$$

где $Z(X) = \beta_0 + \sum_{j=1}^M \beta_j \cdot X_j$, β_j – неизвестные коэффициенты регрессионной модели, X_j – независимые переменные (факторы) модели, M – количество переменных моделей. Видно, что значение $P(X)$ может меняться от 0 (если $Z(X) = -\infty$) до 1 (при $Z(X) = +\infty$), и поэтому $P(X)$ можно интерпретировать как вероятность ускоренного развития атеросклероза.

При построении модели (1) будем решать следующие наиболее важные задачи:

- отбор наиболее информативных переменных X_j , которые значимо влияют на значение величины $Z(X)$ и слабо коррелированы между собой;
- вычисление оценок b_j для неизвестных коэффициентов β_j , $j = 0, 1, \dots, M$, и построение уравнения регрессии для оценивания вероятности ускоренного развития атеросклероза в виде:

$$\hat{P}(X) = \frac{1}{1 + e^{-\left(b_0 + \sum_{j=1}^M b_j \cdot X_j\right)}}; \quad (2)$$

- проверка адекватности построенной модели (2) заданной выборке.

Из 100 изучаемых факторов были отобраны следующие 15, которые имели значимую корреляцию с зависимой переменной (исход сердечно-сосудистое событие) и слабую корреляцию этих независимых переменных между собой. Такими переменными оказались: возраст пациента – X_1 , пол пациента – X_2 , стеноз сонных арте-

рий на 50% и более хотя бы на одной стороне – X_3 , артериальная гипертензия – X_4 , курение – X_5 , ожирение более 1 степени – X_6 , стабильная стенокардия в анамнезе – X_7 , уровень лейкоцитов в крови ЕД/л – X_8 , уровень креатинина крови – X_9 , уровень С-реактивного протеина в мг/л – X_{10} , уровень общего холестерина крови в мм/л – X_{11} , уровень холестерина липопротеидов высокой плотности (ХС ЛПВП) в мм/л – X_{12} , уровень холестерина липопротеидов низкой плотности (ХС ЛПНП) в мм/л – X_{13} , фракция выброса левого желудочка в % – X_{14} , многососудистое поражение коронарного русла – X_{15} . Выборка из этих факторов объемом 202 использовалась дальше для построения логистической модели.

К сожалению, табличный процессор Excel не содержит встроенной функции, позволяющей построить логистическую регрессию. Поэтому нахождение оценок b_j для неизвестных коэффициентов β_j , $j = 0, 1, \dots, M$, и вычисление точностных характеристик этих оценок осуществлялось с использованием модуля РЕГРЕССИЯ статистического пакета SPSS (версия 22), где такая возможность существует.

Для определения наиболее информативных независимых переменных в логистической регрессионной модели и вычисления коэффициентов при этих переменных использовался метод Вальда вперед, в котором построение начиналось с простейшей модели с одной независимой переменной, а затем вводилась на каждом шаге новая независимая переменная – наиболее информативная (т. е. максимально влияющая на значение величины $Z(X)$ для данного шага. На рисунке 1 (скрин окна вывода пакета SPSS) приведены значения $K_{чув}$, $K_{спец}$, $K_{точ}$, характеризующие точность построенной модели на каждом шаге введения в модель новых переменных.

Таблица классификации ^а					
Наблюдаемые		Предсказанные			
		VAR00001		Процент правильных	
		,00	1,00		
Шаг 1	VAR00001	,00	75	27	73,5
		1,00	0	100	100,0
	Общая процентная доля				86,6
Шаг 2	VAR00001	,00	89	13	87,3
		1,00	10	90	90,0
	Общая процентная доля				88,6
Шаг 3	VAR00001	,00	92	10	90,2
		1,00	6	94	94,0
	Общая процентная доля				92,1
Шаг 4	VAR00001	,00	94	8	92,2
		1,00	4	96	96,0
	Общая процентная доля				94,1
Шаг 5	VAR00001	,00	95	7	93,1
		1,00	4	96	96,0
	Общая процентная доля				94,6
Шаг 6	VAR00001	,00	96	6	94,1
		1,00	3	97	97,0
	Общая процентная доля				95,5

$K_{чув}$

$K_{спец}$

$K_{точ}$

а. Значение отсечения - ,400

Рис. 1. Значения характеристик точности на каждом шаге построения уравнения регрессии
Fig. 1. The values of the accuracy characteristics at each step of the regression equation

Расшифруем смысл этих коэффициентов:

- коэффициент чувствительности $K_{\text{чув}}$ – оценка вероятности правильного прогноза НИ;
- коэффициент специфичности $K_{\text{спец}}$ – оценка вероятности правильного прогноза БИ;
- коэффициент точности $K_{\text{точ}}$ – оценка вероятности правильного прогноза как НИ, так и БИ.

Коэффициент $K_{\text{точ}}$ является «суммарным» показателем точности прогноза построенной модели, и поэтому этот показатель дальше является определяющим при отборе переменных логистической модели.

Из рисунка 1 видно, что процедура добавления новых переменных в регрессионную модель заканчивается на шаге 6, когда значение коэффициента $K_{\text{точ}}$ достигает максимального значения. На последнем шаге 6 мы имеем следующие значения коэффициентов:

$$K_{\text{чув}} = 94,1\% (0,941), K_{\text{спец}} = 97,0\% (0,97), \\ K_{\text{точ}} = 95,5\% (0,955),$$

которые говорят о высоких точностных характеристиках модели.

На практике для вычисления этих коэффициентов используют следующую четырехпольную таблицу, учитывающую БИ и НИ (табл. 1).

Таблица 1. Четырехпольная таблица вычисления для коэффициентов чувствительности, специфичности и точности

Table 1. Four-field table of calculations for sensitivity, specificity, and accuracy coefficient

Исходы Outcomes	Результат прогнозирования Prediction result		Итого Total
	Благоприятный исход – БИ (0) Favorable outcome – FO (0)	Неблагоприятный исход – НИ (1) Adverse outcome – AO (1)	
Благоприятный исход (0) Favorable outcome	a – число больных с БИ и результат прогнозирования БИ a – the number of patients with FO and the result of FO prediction	b – число больных с НИ и результат прогнозирования НИ b – the number of patients with FO and the result of AO prediction	a + b – общее число больных с БИ a + b – the total number of patients with FO
Неблагоприятный исход (1) Adverse outcome	c – число больных с НИ и результат прогнозирования БИ c – the number of patients with AO and the result of FO prediction	d – число больных с НИ и результат прогнозирования НИ d – the number of patients with AO and the result of AO prediction	c + d – общее число больных с НИ c + d – the total number of patients with AO

Тогда коэффициенты вычисляются следующим образом:

$$K_{\text{чув}} = \frac{d}{c + d}, K_{\text{спец}} = \frac{a}{a + b}, K_{\text{точ}} = \frac{a + d}{a + b + c + d}, \quad (3)$$

В таблице 2 приведены названия независимых переменных, вошедших в регрессионную модель на шаге 6, обозначения этих переменных в регрессионной модели и значения вычисленных коэффициентов b_j . В последнем столбце таблицы 2 приведены так называемые Р-величины. Если эта величина меньше или равна 0,05, то соответствующий вычисленный коэффициент b_j является значимым и переменная X_j может входить в построенное уравнение регрессии. Анализ таблицы 2 показывает, что все коэффициенты $b_j, j = 0, 1, \dots, 6$ являются значимыми и

в построенное уравнение регрессии могут входить все 6 независимых переменных.

Таблица 2. Вычисленные коэффициенты b_j

Table 2. Calculated coefficients b_j

Названия независимых переменных (факторов) Names of independent variables (factors)	Обозначения в логистической модели Designation in the logistical model	Коэффициенты b_j Coefficients b_j	Р-величина P-level
Возраст Age	X_1	–0,766	0,000
Стеноз сонных артерий Stenosis of carotid arteries	X_2	–3,119	0,008
Ожирение Obesity	X_3	–1,678	0,012
Стенокардия в анамнезе Angina in medical history	X_4	4,016	0,000
СРБ CRP	X_5	0,082	0,025
ЛПВП HDL	X_6	–1,989	0,007
Константа Constant	b_0	47,581	0,000

Примечание: СРБ – С-реактивный белок, ЛПВП – липопротеины высокой плотности.

Note: CRP – C Reactive Protein, HDL – high density lipoproteins.

Подставляя вычисленные коэффициенты в уравнение (2), получаем конкретное уравнение логистической регрессии:

$$\hat{P}(X) = \frac{1}{1 + e^{-[47,581 - 0,766X_1 - 3,119X_2 - 1,678X_3 + 4,016X_4 + 0,082X_5 - 1,989X_6]}}, \quad (4)$$

Очевидно, что для вычисления прогноза вероятности $\hat{P}(X)$ ускоренного развития атеросклероза у какого-либо пациента достаточно подставить в регрессию (4) значения его факторов (значения переменных, используемых в уравнении (4)). Прогнозное значение переменной Y (0 или 1) вычисляется по правилу:

$$Y = \begin{cases} 0, & \text{если } \hat{P}(x) \leq 0,4; \\ 1, & \text{если } \hat{P}(x) > 0,4, \end{cases} \quad (5)$$

где величина порога 0,4 найдена из условия максимума коэффициента точности $K_{\text{точ}}$.

Для упрощения использования построенной регрессионной модели формулы (4), (5) были реализованы в табличном процессоре Excel в виде специального калькулятора. Интерфейс калькулятора показан на рисунке 2. Данные пациента (значения переменных, введенных в табл. 2) вводятся в ячейки, выделенные цветом.

Видно предельно простое использование этого калькулятора для прогноза риска ускоренного развития атеросклероза у конкретного пациента.

Второй подход к вычислению риска ускоренного развития атеросклероза существенно использует машинное обучение, представляющее альтернативный подход к моделированию прогнозов исходов, который может учесть нелинейные взаимосвязи [12]. Задача прогнозирования сводится к классификации пациентов на основе наиболее вероятных исходов. Машинное обучение – это один из методов искусственного интеллекта, решающий зада-

чу не прямым способом, а путем поиска закономерностей в данных в результате обучения алгоритма на множестве примеров с известными исходами (БИ или НИ). При этом алгоритм обучается на наборе примеров, корректируется

с помощью тестового набора, а уже потом применяется на реальных объектах (в данном случае – пациентах), нуждающихся в прогнозе. Другими словами, первый этап – это обучение, второй этап – прогноз.

	B	C	D	E	F	G
1	Обозначения переменных регрессионной модели	Обозначения коэффициентов регрессионной модели	Значения коэффициентов регрессионной модели			
2		b0	47,581			
3	X1	b1	-0,766	порог=	0,4	
4	X2	b2	-3,119			
5	X3	b3	-1,678	Z(X)	Вероятность P(X)	Y предск
6	X4	b4	4,016	-4,808	0,01	0
7	X5	b5	0,082			
8	X6	b6	-1,989			
9						
10				Риск ускоренного развития атеросклероза НЕ ВЫСОКИЙ		
11						
12	Введите значения переменных регрессионной модели (6 величин)					
13						
14	X1	X2	X3	X4	X5	X6
15	Возраст	Стеноз сонных артерий	Ожирение	Стенокардия в анамнезе	СРБ	ЛПВП
16	62	1	0	0	2,1	0,98

Рис. 2. Интерфейс калькулятора

Fig. 2. Calculator interface

Для решения поставленных задач использовалась свободная кроссплатформенная система визуального программирования Orange, разработанная Лабораторией биоинформатики Люблянского университета (Словения). Orange – инструмент с открытым исходным кодом для интеллектуального анализа данных, статистических исследований и визуализации данных, который обеспечивает визуальное программирование для построения

рабочих процессов анализа. Пользовательский интерфейс системы легко осваивается специалистами из разных сфер, связанных с большими наборами разнообразных данных. На рисунке 3 показан рабочий процесс классификации объектов с помощью логистической регрессии, полученный визуальным программированием из модулей, называемых виджетами, которые реализуют отдельные алгоритмы и функции.

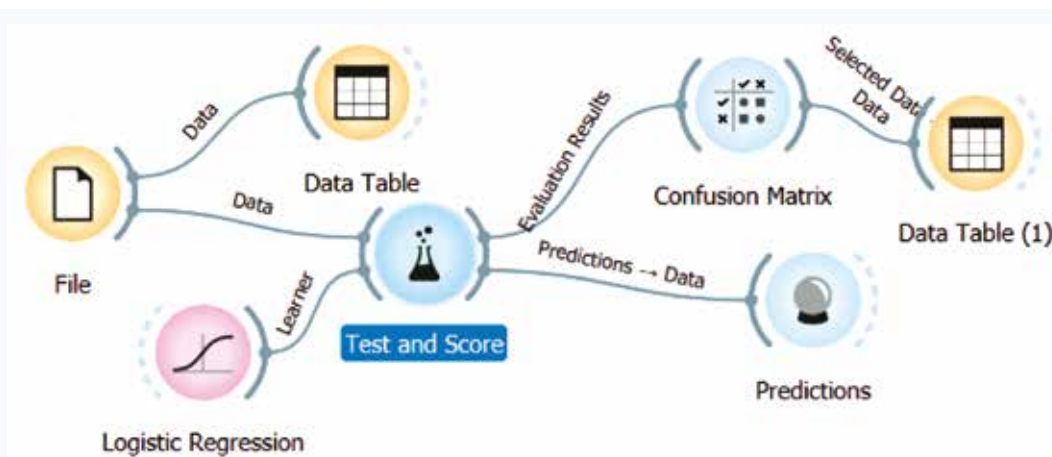


Рис. 3. Рабочий процесс классификации объектов с помощью логистической регрессии
Fig. 3. Workflow of object classification using logistic regression

Была проведена перекрестная проверка (cross-validation) рассматриваемого алгоритма. В основном она используется в ситуациях, когда целью является прогнозирование, и требуется оценить, насколько точно прогностическая модель будет работать на практике. Исходный

набор данных случайным образом делится на N равных частей. Сначала происходит обучение по $N-1$ частям, а тестирование – по N -й части. Затем тестовая часть меняется, и обучение проходит по оставшимся $N-1$ частям. Цикл состоит из N итераций, в ходе которых каждая часть

будет использована в качестве тестовой. Результаты цикла обобщаются (усредняются). В качестве иллюстрации на рисунке 4 представлено окно перекрестной проверки алгоритма логистической регрессии при $N = 3$ и числе независимых переменных, равном 15. Коэффициент точности $K_{точ}$ представлен как CA (classification accuracy) со значением 0,950 (95,0%).

Видно, что используя систему визуального программирования Orange, удается построить прогнозную модель такой же точности, что и в первом подходе (см. рис. 1). На рисунке 5 представлена четырехпольная таблица, называемая в системе Orange «матрицей разногласий» (аналог табл. 1) со значениями $K_{чув}$ и $K_{спец}$, при 15 независимых переменных.



Рис. 4. Результат перекрестной проверки алгоритма логистической регрессии
 Fig. 4. The result of cross-validation of the logistic regression algorithm

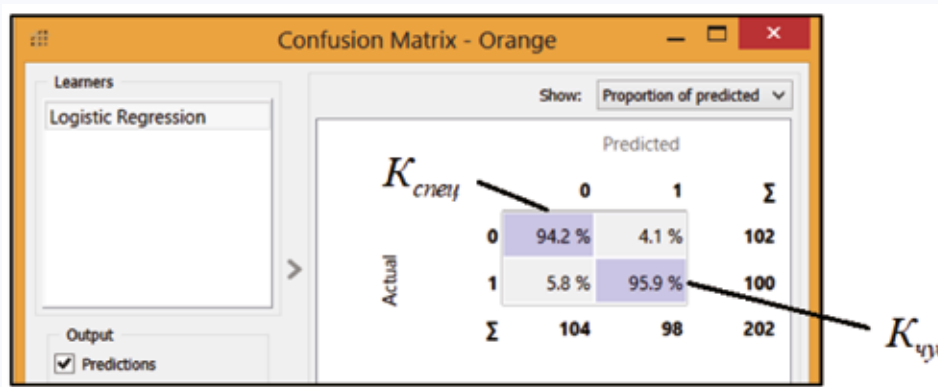


Рис. 5. Окно матрицы разногласий с коэффициентами $K_{чув}$, $K_{спец}$
 Fig. 5. Disagreement matrix window with coefficients $K_{чув}$, $K_{спец}$

Заметим, что в построенной регрессионной модели (4) использовались только 6 наиболее информативных переменных из имеющихся 15. В качестве следующего эксперимента в исходном наборе данных были оставлены эти 6 переменных, и на них проведена перекрестная проверка рассматриваемых алгоритмов, аналогичная продемонстрированной выше проверке для 15 перемен-

ных. В таблице 3 представлены коэффициенты $K_{чув}$ и $K_{спец}$, $K_{точ}$, для 6 и 15 независимых переменных. Максимальные значения коэффициентов выделены жирным шрифтом.

Из таблицы 3 видно, что при числе независимых переменных 15 коэффициенты незначительно увеличиваются.

Таблица 3. Коэффициенты моделей с 6 и 15 независимыми переменными
 Table 3. Coefficients of models with 6 and 15 independent

Алгоритм Algorithm	$K_{чув}$		$K_{спец}$		$K_{точ}$	
	6 перемен-ных 6 variables	15 перемен-ных 15 variables	6 перемен-ных 6 variables	15 перемен-ных 15 variables	6 перемен-ных 6 variables	15 перемен-ных 15 variables
Логистическая регрессия Logistic regression	95,1	95,9	94,1	94,2	94,6	95,0

Эксперименты с рассмотренными алгоритмами, в которых из исходного набора 202 пациентов случайным образом выбирались несколько тестовых пациентов для построения их прогнозов, а остальные использовались

для обучения алгоритма, подтвердили высокую точность прогнозирования исходов. Заметим, что точность прогнозирования можно существенно повысить за счет увеличения обучающей выборки.

Возникает вопрос: какой из двух подходов предпочтительнее? Ответ представляется неоднозначным. Так как оба подхода дают прогноз с практически одинаковой точностью (см. рис. 1, табл. 3), то для ответа необходимо учитывать другие аспекты построения и использования прогнозных моделей. Так, реализация построенной регрессионной модели (первый подход) в виде калькулятора позволяет сравнительно просто (Excel есть у всех) вычислить риск ускоренного развития атеросклероза. Однако построение самой модели (с отбором наиболее информативных независимых переменных) представляется более сложным по сравнению со вторым подходом, где этот выбор делается системой Orange. Кроме того, машинное обучение при построении прогнозной модели является хорошей основой (в силу простоты и компактности программирования алгоритмов обработки) при решении научно-исследовательских задач, связанных с анализом влияния данных разного объема на точность построенной прогнозной модели.

Клинический случай

Пациент К., 46 л., предприниматель, масса тела – 93 кг, рост – 176 см, перенес острый ИМ с подъемом сегмента ST нижней стенки левого желудочка в 2018 г., чрескожная транслюминальная коронарная ангиопластика (ЧТКА) со стентированием правой коронарной артерии (ПКА) в 2018 г. ЧТКА со стентированием передней нисходящей коронарной артерии (ПНА) в 2019 г., 2020 г. Тяжесть хронической сердечной недостаточности – 1-й функциональный класс по NYHA. Факторы риска: артериальная гипертензия, дислипидемия, курение (стаж – 35 лет).

На электрокардиограмме регистрировался синусовый ритм, элевация сегмента ST в III, AVF. На эхокардиограмме, проведенной на 10-е сут индексного события (ИМ в 2018 г.), выявлялся гипокинез нижней области левого желудочка, фракция выброса левого желудочка – 45%. Согласно ультразвуковому исследованию брахиоцефальных артерий, выявлены стенозы справа (45%) и слева (55%). Уровень гликемии в момент включения в исследование – 5,4 ммоль/л, креатинина – 75,6 мкмоль/л, билирубина общего – 7,6 мкмоль/л, билирубин прямой отрицателен, АЛТ – 22 усл. ед/л, АСТ – 14 усл. ед/л, калий – 3,8 ммоль/л, натрий – 138 ммоль/л, общий холестерин – 5,08 ммоль/л, триглицериды – 3,93 ммоль/л, альфа-холестерин – 1,24 ммоль/л, ЛПНП – 2,05 ммоль/л, С-реактивный протеин – 5 мг/л.

Пациент получал стандартную двойную антитромбоцитарную терапию (тикагрелор и аспирин), а также бисопролол, статин, эзетимиб, лозап.

Согласно результатам рискметрии по калькулятору в Excel (регрессионной модели), определялся высокий риск прогрессирующего атеросклероза: вероятность составила 0,9959, а по Orange – 0,9977.

Предсказанный прогноз подтвердился клинически у пациента в виде повторных госпитализаций по поводу нестабильной стенокардии со стентированием ПНА дважды: в 2019 и 2020 гг.

Таким образом, изложенные два подхода могут успешно дополнять друг друга, служа одной цели: построению более точных прогнозных моделей риска ускоренного развития атеросклероза.

Литература / References

- Shah P., Bajaj S., Virk H., Bikkina M., Shamoof F. Rapid progression of coronary atherosclerosis: a review. *Thrombosis*. 2015;2015:634983. DOI: 10.1155/2015/634983.
- Li M., Ren C., Wu C., Li X., Mao W. Bioinformatics analysis reveals diagnostic markers and vital pathways involved in acute coronary syndrome. *Cardiol. Res. Pract.* 2020;2020:3162581. DOI: 10.1155/2020/3162581.
- Кухарчук В.В., Езов М.В., Сергиенко И.В., Арабидзе Г.Г., Бубнова М.Г., Балахонova Т.В. и др. Диагностика и коррекция нарушений липидного обмена с целью профилактики и лечения атеросклероза. Российские рекомендации, VII пересмотр. *Атеросклероз и дислипидемии*. 2020;38(1):7–42. [Kukharchuk V.V., Ezhov M.V., Sergienko I.V., Arabidze G.G., Bubnova M.G., Balakhonova T.V. et al. Diagnostics and correction of lipid metabolism disorders in order to prevent and treat of atherosclerosis Russian recommendations, VII revision. Moscow, 2020. *Atherosclerosis and Dyslipidemia*. 2020;38(1):7–42. (In Russ.)]. DOI: 10.34687/2219-8202.JAD.2020.01.0002.
- Knuuti J., Wijns W., Saraste A., Capodanno D., Barbato E., Funck-Brentano C. et al. 2019 ESC Guidelines for the diagnosis and management of chronic coronary syndromes: The Task Force for the diagnosis and management of chronic coronary syndromes of the European Society of Cardiology (ESC). *Eur. Heart J.* 2020;41(3):407–477. DOI: 10.1093/eurheartj/ehz425.
- Gulati M., Levy P.D., Mukherjee D., Amsterdam E., Bhatt D.L., Birtcher K.K. et al. 2021 AHA/ACC/ASE/CHEST/SAEM/SCCT/SCMR Guideline for the evaluation and diagnosis of chest pain: Executive summary: A Report of the American College of Cardiology. American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation*. 2021;144(22):e368–e454. DOI: 10.1161/CIR.0000000000001029.
- Thygesen K., Alpert J.S., Jaffe A.S., Chaitman B.R., Bax J.J., Morrow D.A. et al. ESC Scientific Document Group. Fourth universal definition of myocardial infarction (2018). *Eur. Heart J.* 2019;40(3):237–269. DOI: 10.1093/eurheartj/ehy462.
- Landmesser U., Chapman M.J., Stock J.K., Amarencu P., Belch J.J.F., Borén J. et al. 2017 Update of ESC/EAS Task Force on practical clinical guidance for proprotein convertase subtilisin/kexin type 9 inhibition in patients with atherosclerotic cardiovascular disease or in familial hypercholesterolaemia. *Eur. Heart J.* 2017;39(14):1131–1143. DOI: 10.1093/eurheartj/ehx549.
- Биленко М.В., Владимиров Ю.А., Хильченко А.В., Павлова С.А. Патент 2408019 Российская Федерация, МПК G01N 33/50. Способ экспресс-диагностики тяжести ишемических повреждений сердца у больного с ишемической болезнью сердца и предрасположенности больного к прогрессированию атеросклероза: № 2009121616/15: заявл. 08.06.2009; опубл. 27.12.2010. [Bilenko M.V., Vladimirov Ju.A., Khil'chenko A.V., Pavlova S.A. Patent Russian Federation, 2408019, Int. Cl. G01N 33/50. Instant severity diagnosis of ischemic heart damage in patient with chronic heart disease and propensity for progressing atherosclerosis. 2009121616/15, date of filing: 08.06.2009, date of publication: 27.12.2010]. URL: http://www.chemilum.ru/files/patent_ru-2408019.pdf (10.05.2023).
- Арефьева Т.И., Балахонova Т.В., Красникова Т.Л., Ноева Е.А., Потехина А.В., Проваторов С.И. и др. Патент RU 2566288 C1 Российская Федерация, МПК G01N 33/53. Способ диагностики предрасположенности к прогрессированию атеросклероза у больных с хронической ишемической болезнью сердца по отношению концентраций Интерлейкина-10 и Интерлейкина-17 в периферической крови: № 2014141099/15, заявл. 13.10.2014, опубл. 20.10.2015. [Aref'eva T.I., Balakhonova T.V., Krasnikova T.L., Noeva E.A., Potekhina A.V., Provatorov S.I. et al. Patent RU 2566288 C1, Int. Cl. G01N 33/53. Diagnostic technique for predisposition to atherosclerosis progression in patients with chronic ischemic heart disease as shown by peripheral blood interleukin-10 and interleukin-17 concentrations. Application: 2014141099/15, date of filing: 13.10.2014, date of publication: 20.10.2015]. URL: <https://patentimages.storage.googleapis.com/99/db/72/ea/591c80020882/RU2566256C1.pdf> (10.05.2023).
- Арефьева Т.И., Балахонova Т.В., Красникова Т.Л., Ноева Е.А., Потехина А.В., Проваторов С.И. и др. Патент 2575791 C1 Российская Федерация, МПК G01N 33/50. Способ диагностики предрасположенности к прогрессированию атеросклероза у больных с хронической ишемической болезнью сердца по содержанию Интерлейкина-10-продуцирующих Т-лимфоцитов в периферической крови: № 2014141098/15: заявл. 13.10.2014; опубл. 20.02.2016. [Aref'eva T.I., Balakhonova T.V., Krasnikova T.L., Noeva E.A., Potekhina A.V., Provatorov S.I. et al. Patent 2575791 C1 Russian Federation, MПК G01N 33/50. Method of diagnosis of predisposition to atherosclerosis progression in patients with chronic ischemic heart disease by the content of Interleukin-10-producing T-lymphocytes in peripheral blood: № 2014141098/15: filed. 13.10.2014; published. 20.02.2016.]

na A.V., Provatorov S.I. et al. Patent RU 2575791 C1, Int. Cl. G01N 33/50. Diagnostic technique for disposition to atherosclerosis progression in patients with chronic ischemic heart diseases by peripheral blood interleukin-10-producing lymphocyte count; Application: 2014141098/15, date of filing: 13.10.2014, date of publ.: 20.02.2016]. URL: <https://patentimages.storage.googleapis.com/d9/b2/ae/bd04c55b44d480/RU2575257C1.pdf> (10.05.2023).

11. Рагино Ю.И. Нестабильная атеросклеротическая бляшка и её лабораторные биохимические маркеры. Новосибирск: Наука; 2019:120. [Ragino Yu.I. Unstable atherosclerotic plaque and its laboratory biochemical markers. Novosibirsk: Nauka; 2019:120. (In Russ.)].
12. Obermeyer Z., Emanuel E.J. Predicting the future – big data, machine learning, and clinical medicine. *New Engl. J. Med.* 2016;375(13):1216–1219. DOI: 10.1056/NEJMp1606181.

Информация о вкладе авторов

Ложкина Н.Г. участвовала в разработке концепции статьи, собрала первичный литературный материал, участвовала в написании текста, подготовила окончательную версию текста.

Воскобойников Ю.Е., Копылов В.Н. участвовали в разработке концепции статьи, написании и редактировании текста, предложили иллюстративный материал, участвовали в подготовке окончательной версии текста.

Пархоменко О.М. собрала первичный литературный материал, участвовала в написании текста.

Воевода М.И. участвовал в разработке структуры статьи, написании и подготовке окончательной версии текста.

Information on author contributions

Lozhkina N.G. article concept, collection of primary literary material, writing the text, preparation of the final version of the text.

Voskoboynikov Yu.E. and Kopylov V.N. article concept, writing and editing the text, suggestion illustrative material and preparation of the final text.

Parkhomenko O.M. collection of primary literary material and writing the text.

Voevoda M.I. working out the structure of the article and writing the final version of the text.

Сведения об авторах

Ложкина Наталья Геннадьевна, д-р мед. наук, профессор, главный научный сотрудник, руководитель группы Клиническая и экспериментальная кардиология, Федеральный исследовательский центр фундаментальной и трансляционной медицины Сибирского отделения Российской академии наук. ORCID 0000-0002-4832-3197.

E-mail: lozhkina.n@mail.ru.

Воскобойников Юрий Евгеньевич, д-р физ.-мат. наук, профессор, факультет инженерных и информационных технологий, заведующий кафедрой прикладной математики, Новосибирский государственный архитектурно-строительный университет. ORCID 0000-0002-5282-6002.

E-mail: voscob@mail.ru.

Копылов Василий Николаевич, д-р техн. наук, профессор, кафедра информационных систем и технологий, Сибирский региональный научно-исследовательский гидрометеорологический институт. ORCID 0000-0000-0000-0000.

E-mail: kopylov@sibnigmi.ru.

Пархоменко Ольга Михайловна, канд. мед. наук, заместитель главного врача по терапии, Городская клиническая больница № 1. ORCID 0000-0003-4736-6491.

E-mail: parh.om@mail.ru.

Воевода Михаил Иванович, д-р мед. наук, академик РАН, профессор, директор Федерального исследовательского центра фундаментальной и трансляционной медицины. ORCID 0000-0001-9425-413X.

E-mail: mvoevoda@ya.ru.

Information about the authors

Natalia G. Lozhkina, Dr. Sci. (Med.), Professor, Chief Research Scientist, Head of Clinical and Experimental Cardiology Group, Federal Research Center for Fundamental and Translational Medicine. ORCID 0000-0002-4832-3197.

E-mail: lozhkina.n@mail.ru.

Yurii E. Voskoboynikov, Dr. Sci. (Phys. and Math.), Professor, Department of Engineering and Information Technologies, Head of Applied Mathematics Department, Novosibirsk State University of Architecture and Civil Engineering. ORCID 0000-0002-5282-6002.

E-mail: voscob@mail.ru.

Vasily N. Kopylov, Dr. Sci. (Phys. and Techn.), Professor, Siberian Regional Hydrometeorological Research Institute. ORCID 0000-0000-0000-0000.

E-mail: kopylov@sibnigmi.ru.

Olga M. Parkhomenko, Cand. Sci. (Med.), Deputy Chief Physician for Internal Medicine, City Clinical Hospital No. 1, Novosibirsk. ORCID 0000-0003-4736-6491.

E-mail: parh.om@mail.ru.

Mikhail I. Voevoda, Dr. Sci. (Med.), Academician of the Russian Academy of Sciences, Professor, Director of the Federal Research Center for Fundamental and Translational Medicine. ORCID 0000-0001-9425-413X.

E-mail: mvoevoda@ya.ru.

 **Natalia G. Lozhkina**, e-mail: lozhkina.n@mail.ru.

 **Ложкина Наталья Геннадьевна**, e-mail: lozhkina.n@mail.ru.

Received April 12, 2023

Поступила 12.04.2023