
РАЗДЕЛ 4 | КЛАССИФИКАЦИЯ, РАСПОЗНАВАНИЕ И
ДИАГНОСТИКА

УДК 004.852

**АЛГОРИТМ ОТБОРА ПРЕДИКТОРОВ
ВНУТРИГОСПИТАЛЬНОЙ ЛЕТАЛЬНОСТИ
ПОСЛЕ КОРОНАРНОГО ШУНТИРОВАНИЯ¹**

Б.И. Гельцер (*boris.geltser@vvsu.ru*)

В.Ю. Рублев (*carinash@vvsu.ru*)

Дальневосточный федеральный университет,

Школа биомедицины, Владивосток

А.Б. Кригер (*alexandra.kriger@vvsu.ru*)

К.И. Шахгельдян (*carinash@vvsu.ru*)

В.Г. Широбоков (*wpn@inbox.ru*)

Владивостокский государственный университет
экономики и сервиса, Владивосток

Институт информационных технологий, Владивосток

Целью исследования является разработка алгоритма отбора и оценки информативности предикторов прогнозирования внутригоспитальной летальности (ВГЛ) после коронарного шунтирования (КШ). Проанализировано 866 историй болезни и 99 факторов, включающих результаты клинических, лабораторных и инструментальных исследований, полученных до КШ. Предикторы ВГЛ выделены путем многоступенчатой процедуры отбора с анализом статистических гипотез и расчетом весовых коэффициентов однофакторных моделей, разработкой многофакторных моделей на основе логистической регрессии (ЛР), случайного леса (СЛ) и искусственных нейронных сетей (ИНС). Мо-

¹ Работа выполнена при финансовой поддержке РФФИ (проекты №№ 18-29-03131 и 19-29-01077).

дели на основе СЛ и ИНС позволили не только построить более точные инструменты прогноза, но и верифицировать 5 дополнительных предикторов ВГЛ. Для выбора наилучшей модели предложен комбинированный показатель качества.

Ключевые слова: отбор предикторов внутригоспитальной летальности, методы машинного обучения, логистическая регрессия, случайный лес, искусственная нейронная сеть

Введение

В последние годы в клинических исследованиях все шире применяются методы машинного обучения (МО), задачей которых является разработка моделей расчета выходных прогнозируемых переменных на основе нескольких входных факторов, характеризующих клинико-функциональный статус пациентов при различных заболеваниях [Karim et al., 2018], [Kurt et al., 2008]. При этом доказательствам предиктивного потенциала факторов прогностических моделей придается особое значение. Для отбора предикторов существуют «классические» статистический методы, к наиболее известным из которых, можно отнести анализ межгрупповых различий и оценку корреляции, которые позволяют верифицировать интенсивность линейных взаимосвязей отдельных факторов и прогнозируемых переменных. Вместе с тем, указанные методы не позволяют выявлять нелинейные взаимосвязи потенциальных предикторов с прогнозируемой переменной. Кроме того, взаимозависимость между факторами порождает проблему мультиколлинеарности и приводит к снижению качества прогностической модели.

Ишемическая болезнь сердца (ИБС) – одна из основных причин смертности населения в большинстве стран мира, в том числе и в Российской Федерации [Бойцов и др., 2015]. Операции коронарного шунтирования (КШ) относятся к одним из наиболее распространенных видов восстановления коронарного кровотока у больных ИБС. В связи с этим постоянно возрастает интерес к изучению роли факторов, влияющих на риск развития неблагоприятных исходов КШ [Yamaoka et al., 2015]. В последние годы в клинической кардиологии все чаще используются унифицированные инструменты прогнозирования неблагоприятных событий, основанные на результатах крупных проспективных исследований. Так, в кардиологии известны шкалы EuroSCORE и EuroSCORE II, оценивающие вероятность внутригоспитальной летальности (ВГЛ) в течении 30-ти дней после КШ [Sedaghat et al., 2013]. К «классическим» моделям прогноза ВГЛ в настоящее время относят шкалу EuroSCORE II, которая представляет собой логистическую регрессию (ЛР) и включает 18 предикторов, характеризую-

щих клиничко-функциональный статус больных до операции [Nashef et al., 2012]. Модель EuroSCORE II содержит предикторы, не все из которых имеют статистическая значимость ($p\text{-value} > 0,05$), что приводит к снижению ее прогнозной ценности. Кроме того, в шкалу включены факторы, которые коррелируют друг с другом, что создает проблему мультиколлинеарности [Faraway et al., 2016]. ЛР-модель является одной из наиболее популярных методов в клинической медицине для решения задачи классификации, к которой можно отнести прогнозируемую вероятность ВГЛ. В качестве предикторов в «идеальной» многофакторной ЛР-модели должны использоваться факторы, имеющие статистическую значимость, не коррелирующие между собой и линейно связанные с прогнозируемой переменной. Реализация этих условий при разработке моделей приводит к отказу от использования части информативных предикторов, которые нелинейно связаны с выходными переменными и разобщает участие в ее построении взаимозависимых факторов. Кроме того, для решения задач клинической медицины чаще используются ЛР-модели, которые не проходят процедуры кросс-валидации, что снижает их прогностическую точность при применении в других популяциях больных. Отсутствие «идеальных» прогностических технологий инициирует все большее количество исследований, в которых на основе современных методов МО реализуются подходы, направленные на совершенствование существующих и разработку новых моделей. Для их построения используются такие методы, как: дерево принятия решений вместе с градиентным бустингом, случайный лес (СЛ), машина опорных векторов, наивный байесовский классификатор, регрессия Кокса, искусственные нейронные сети (ИНС) и др. [Breiman et al., 2001], [Flach, 2012], [Kurt et al., 2008].

1 Материалы и методы исследования

Проведен ретроспективный анализ 866 историй болезни пациентов, оперированных методом КШ в отделении сердечно-сосудистой хирургии КБУЗ «Приморская краевая клиническая больница №1» г. Владивостока. Среди обследованной когорты пациентов было выделено 2 группы лиц. В первую из них вошли 35 (4%) больных, умерших в течении первых 30 суток после КШ, во вторую – 831 (96%) больной с благоприятным исходом операции. Поиск предикторов осуществлялся среди 99 факторов, характеризующих клиничко-anamnestический и функциональный статус каждого больного до КШ.

Отбор предикторов из множества факторов необходим в связи с тем, что те из них, которые не имеют прогностическую ценность для исследуемого события, добавляют шум к оценке взаимосвязей между ними и прогнозируемыми переменными [Faraway et al., 2016]. Кроме того, в клинической ме-

дицине важность обоснования выбора предикторов, степени и характера их влияния на результирующую переменную очень высока и требует убедительных доказательств. В нашем исследовании выбор предикторов включал несколько этапов. На первом из них использовали методы статистического анализа для оценки влияния отдельных факторов на прогнозируемую переменную. В связи с тем, что прогнозируемая переменная была дискретной, нами использовались тесты для верификации межгрупповых различий: Стьюдента, Манна-Уитни, Фишера и χ^2 . На втором этапе исследования для оценки статистической значимости и сравнительного анализа степени влияния отдельных факторов на прогнозируемую переменную были использованы однофакторные ЛР-модели. На третьем этапе исследования мы разработали базовую модель ЛР, используя показатели шкалы EuroSCORE II, которые на предыдущих этапах показали свой предиктивный потенциал. Наряду с этим мы также оценили качество разработанной базовой модели и шкалы EuroSCORE II на анализируемой когорте. Критериальным порогом для оценки вероятности ВГЛ, согласно шкале EuroSCORE II, является величина 5%. Прогностическую ценность моделей классификации оценивали тремя метриками: чувствительности (sen), специфичности (spec), площадь под ROC-кривой (AUC). На четвертом этапе исследования мы расширяли множество предикторов с помощью метода пошагового включения [Faraway et al., 2016]. Все факторы, которые имели линейные взаимосвязи с прогнозируемой переменной и были выделены на предыдущих этапах, прошли процедуру пошагового включения в многофакторную ЛР-модель. Факторы, которые могли иметь нелинейные взаимосвязи, были отобраны с помощью других методов МО: СЛ и ИНС. Улучшение метрик качества модели при включении отдельного фактора свидетельствовало о предикативном потенциале последнего. Если фактор приводил к улучшению качества в разных моделях, это являлось дополнительным доказательством корректности его отбора в качестве предиктора.

Основными инструментами анализа данных и разработки моделей являлись языки R и Python.

2. Результаты анализа

2.1. Предварительный статистический анализ

Результаты первого этапа исследования показали, что только отдельные факторы имеют достоверных межгрупповые различия (табл. 1). Так, из 18 переменных шкалы EuroSCORE II статистически значимые межгрупповые различия фиксировались только по 7 параметрам, к которым относились возраст, фракция выброса (ФВ) левого желудочка (ЛЖ) меньше 30%, ФВ от 30% до 50%, ранее перенесенный инфаркт миокарда (ИМ), атеросклеротическое поражение периферических артериальных бассейнов (ППАБ), экс-

тренность операции (ЭО), хроническая сердечная недостаточность (ХСН) III-IV функциональных классов (ФК). Такие факторы EuroSCORE II, как пол, клиренс креатинина (КК), среднее давление в легочной артерии (СрД-ЛА), нестабильная (НС) и стабильная (СС) стенокардия, наличие хронической обструктивной болезни легких (ХОБЛ), сахарного диабета 2-го типа (СД), не имели значимых межгрупповых различий. Были выделены дополнительные факторы, имеющие межгрупповые различия: систолическое артериальное давление (САД), признаки наличия артериальной гипертензии (АГ) и аортального стеноза (АС). Из остальных факторов, которые не имели статистически значимых различий, мы привели частоту сердечных сокращения (ЧСС), индекс относительной толщины (ИОТ) задней стенки ЛЖ и относительный индекс массы миокарда ЛЖ (ОИММЛЖ).

Табл 1.

Показатели	1-я группа, n=35	2-я группа, n=831	ОШ, (95% ДИ)	p-value
Женщины	10 (28,6%)	171 (20,6%)	1,56 [0,7; 3,2]	0,35
Возраст	66,9±5,5	62,9±7,5	–	0,001
КК ≤50 мл/мин	7 (20%)	65 (10%)	2,3 [0,88; 5,2]	0,109
ППАБ	18 (51,4%)	238 (29,6%)	2,6 [1,3; 5,4]	0,0066
ИМ	19 (55,9%)	140 (17,4%)	5,6 [2,8; 11,4]	<0,0001
ФВ крови, %	50,2±11,8	57,7 ±9,6	–	0,0007
ФВ от 30-50%	16 (45,7%)	168 (22%)	3 [1,5; 6,9]	0,0023
ФВ < 30%	2 (5,7%)	4 (0,52%)	11,8 [1,4; 66,5]	0,013
ХСН III-IV ФК	17 (48,6%)	130 (15,6%)	5,1 [2,5; 10,3]	<0,0001
НС	16 (45,7%)	460 (55,6%)	0,67 [0,3; 1,3]	0,33
СС IV ФК	3 (8,6%)	30 (3,6%)	2,6 [0,56; 7,9]	0,3
ЭО	5 (14,3%)	25 (3%)	5,4 [1,7; 14,3]	0,002
СД	7 (20%)	28 (23,6%)	0,82 [0,3; 1,8]	0,77
ХОБЛ	6 (17,1%)	90 (10,8%)	1,7 [0,63; 4]	0,37
Креатинин, мкмоль/л	111,13±30	103,5±21	–	0,15
ФВ крови*, %	50,2±11,8	57,7 ±9,6	–	0,0007
ЧСС* уд./мин	74,5±16,6	70±10	–	0,12
САД*, мм р.ст.	125,7±17,6	133,2 ±18,2	–	0,018
АГ*	28 (80%)	768 (92,4%)	0,3 [0,13; 0,8]	0,011
ИОТ *	0,39 ±0,1	0,42 ±0,09	–	0,123
ОИММЛЖ*	1,12±0,346	1,1±0,9	–	0,8
АС*	5 (14,3%)	30 (3,6%)	4,5 [1,4; 11,7]	0,0069

Примечание: * – дополнительные факторы; ДИ – доверительный интервал; непрерывные факторы представлены как Me±SD.

На втором этапе исследования для оценки степени влияния отдельных факторов риска на ВГЛ нами были построены однофакторные модели ЛР с расчетом весовых коэффициентов, характеризующих предиктивную ценность анализируемых показателей (табл. 2). Мы рассматривали непрерывные факторы, предварительно нормализованные, наряду с их категориальными аналогами, так как в нашем исследовании непрерывные факторы имели большую степень влияния на ВГЛ.

Табл. 2.

Показатели	Коэффициент	p-value
Возраст (непрерывный)	3,6254	0,0021
Возраст (непрерывный при > 60 лет)	1,8647	0,0033
Креатинин, ммоль/мл	2,3598	0,0413
КК ≤50 мл/мин	0,8143	0,0656
КК 50-85 мл/мин	0,1151	0,7451
ППАБ	0,9253	0,0077
ИМ	1,7285	0,0000009
ФВ крови, %	-4,5188	0,000026
ФВ от 30-50%	1,0851	0,0020
ФВ < 30%	2,4423	0,0057
ХСН III-IV ФК	1,6105	0,000005
ЭО	1,6752	0,0014
ЧСС *	2,8163	0,0172
САД*	-4,1152	0,0160
АГ*	-1,1971	0,0070
ИОТ*	-3,261	0,3058
ОИММЛЖ*	0,0194	0,9109
АС*	1,4929	0,0039

В результате анализа было установлено, что максимальное влияние на ВГЛ имеет ФВ в непрерывной форме (-4,52), а минимальное – поражение периферических артерий (0,93).

2.2. Многофакторные логистические регрессионные модели

На третьем этапе исследования были разработаны 2 варианта многофакторных моделей ЛР. В первой из них были использованы 7 статистически значимых предикторов шкалы EuroSCORE II, а во второй – они были дополнены 3 новыми факторами (ЧСС, признаки АГ и АС). Было установлено, что предикторы в форме непрерывных переменных (возраст, креатинин в сыворотке крови и ФВ) по отношению к их категориальным аналогам, которые используются в шкале EuroSCORE II, обладали лучшими прогностическими свойствами, что подтверждалось весовыми коэффициентами однофакторной модели и метриками качества многофак-

торных моделей. В то же время, такие предикторы EuroSCORE II, как: гендерная принадлежность, ХСН II ФК, стабильная стенокардия IV ФК, ХОБЛ, СД 2-го типа, СрДЛА в авторских многофакторных моделях были также статистически незначимы, как они были и в однофакторных. Напротив, фактор недавно перенесенного ИМ, предсказательная ценность которого в шкале EuroSCORE II была недостаточной ($p > 0,05$), в авторской многофакторной модели приобрела существенное прогностическое значение ($p < 0,0001$).

2.3. Разработка моделей случайного леса и нейронных сетей

Рассчитали базовые вероятности ВГЛ, используя шкалу и предикторы EuroSCORE II. Было установлено, что качество прогноза по метрикам AUC, чувствительности и специфичности для «классической» шкалы EuroSCORE II с 18 предикторами составляло 0,73, 0,25 и 0,92 соответственно. Это свидетельствует о недостаточной точности при апробации на анализируемой когорте и необходимости совершенствования.

На следующем этапе анализа разрабатывали модели ВГЛ на основе методов ЛР, СЛ и ИНС, используя процедуру кросс-валидации, которая обеспечивалась многократным (не менее 500 раз) случайным разделением когорты обследованных на обучающую и тестовую выборки в соотношении 75% и 25%, соответственно. Параметры моделей на обучающих выборках подбирали таким образом, чтобы усредненные значения трех метрик качества на соответствующих тестовых выборках достигали максимальных величин. Для ЛР параметрами модели были отобранные предикторы, которые, имели статистическую значимость в многофакторной модели. В результатах анализа представлены 2 группы метрик качества. Первая из них - это усредненные значения 3-х метрик на тестовых выборках, случайным образом сгенерированных из обследованной когорты. Общими для построенных моделей являются отобранные предикторы, часть из которых ранее не входила ни в одну из систем оценки ВГЛ после КШ. Еще одной задачей является разработка такой модели, которая обеспечит наиболее точный прогноз ВГЛ. С одной стороны, принято считать, что усреднение результатов расчетов моделей позволяет получать более точные прогнозы. С другой стороны, некоторые модели, разработанные при кросс-валидации на несбалансированных выборках, могут снижать прогностическую точность «лучших» моделей. При приемлемой усредненной точности моделей выбрать лучшую модель, которую в дальнейшем можно использовать на неизвестной выборке, может быть более эффективным решением, чем использовать множество моделей с процедурой усреднения результата. В виду несбалансированности выборки для определения «лучшей» модели предлагается выполнить следующую процедуру. Во-первых, необходимо выбрать единственный критерий отбора. Ана-

лиз показал, что ни одна из вышеприведенных метрик не могла характеризовать в целом качество модели. Мы предлагаем определить комбинированный показатель качества как среднее между тремя метриками: AUC, чувствительность и специфичность. Наибольшее среднее соответствует лучшей модели. Во-вторых, мы оценили качество модели как на тестовой выборке, так и на общей когорте пациентов. Худший из этих двух результатов считали оценкой качества лучшей модели, и он приведен во второй группе метрик (табл.3).

Табл. 3

Метод	Усредненные метрики на тестовых выборках			Лучшие суммарные метрики на всей когорте		
	AUC	sen	spec	AUC	sen	spec
ЛР-0	–	–	–	0,75	0,25	0,92
ЛР-I	0,83	0,74	0,78	0,86	0,83	0,76
ЛР-II	0,85	0,7	0,8	0,89	0,83	0,78
СЛ-I	0,71	0,69	0,71	0,89	1	0,79
СЛ-II	0,78	0,82	0,74	0,9	1	0,81
ИНС-I	0,85	0,798	0,895	0,96	0,93	0,91
ИНС-II	0,9	0,86	0,935	0,999	1	0,998
ИНС-III	0,949	0,923	0,975	1	1	1

Архитектура моделей ИНС в процессе исследования менялась от 1 до 200 слоев и от 3 до 100 нейронов в каждом слое. Окончательный вариант включал 3 скрытых слоя с 7 или 10 нейронами в каждом. В качестве функции активации на всех трех слоях наилучшим образом подошел сигмоид. Оптимизатор был представлен алгоритмом Adam с ускоренным градиентом Нестерова. Основными параметрами разработанных моделей СЛ были 1000 деревьев для «голосования» и от 6 до 8 признаков для расщепления.

Сравнительный анализ показал существенные различия моделей EuroSCORE II (ЛР-0 на 18 предикторах) и авторских многофакторных моделей ЛР. Авторская модель ЛР-I, построенная на 7 статистически значимых предикторах EuroSCORE II, обеспечила на тестовой выборке рост показателей чувствительности (0,83) и AUC (0,86) на фоне более низкой специфичности (0,76). Расширенная модель ЛР-II с дополнительным параметром ЧСС, признаком AC и AG позволила увеличить показатели специфичности до 0,78 и AUC до 0,89. Лучшая модель СЛ-I с предикторами EuroSCORE II показала 100% чувствительность (против 0,83 для ЛР-I) при сопоставимом уровне специфичности (0,76 vs 0,79) и AUC (0,86 vs 0,89). Модель СЛ-II с тремя дополнительными предикторами имела лучшее качество и в среднем и при выборе лучшей модели, чем СЛ-I и все модели ЛР. Три модели ИНС показали лучшие результаты, чем модели ЛР и СЛ как в

среднем, так и для наилучшей модели. Добавление 3-х предикторов (ЧСС, САД и признак АС) значительно увеличивало качество модели ИНС-II по сравнению с ИНС-I (7 основных предикторов). Два дополнительных фактора, характеризующих гипертрофию миокарда ИОТ и ОИММЛЖ позволили получить наилучшую модель ИНС-III со 100% точностью.

3. Обсуждение

Алгоритм отбора предикторов для прогнозирования ВГЛ предполагает выделение группы факторов, которые имеют межгрупповые различия или отношение шансов для них статистически значимо. Затем необходимо с помощью однофакторных ЛР-моделей выделить предикторы, которые линейно взаимосвязаны с ВГЛ. Объединение этих выделенных групп факторов (они не совпадают) является базовым набором предикторов, которые в первую очередь включаются в ЛР-модель. Для верификации предикторов используем процедуру кросс-валидации с усреднением трех метрик качества на тестовых выборках. Все валидированные таким образом предикторы должны пройти дополнительную верификацию с помощью моделей СЛ и ИНС. Кроме того, эти модели позволяют выделить еще одну группу факторов, для которых не была установлена линейная взаимосвязь. Наряду с усреднением метрик предлагаем рассматривать лучшие модели, которые выбираются за счет усреднение трех метрик качества и проверки на полной выборке. При однофакторном анализе важна не только величина весовых коэффициентов, но интерпретация их знака. Все значимые факторы могли быть объяснены с точки зрения клинической медицины и физиологии. Исключение составляют показатель САД и АГ, повышение которых снижает риски ВГЛ. Объяснение такого влияния на ВГЛ требует проведение дополнительных клинических исследований, но предположительно, больные АГ находятся на антигипертензивной терапии и это позволяет им удерживать давление в ограниченном диапазоне. При отсутствии АГ и соответствующей терапии этот признак может играть роль фактора риска.

Заключение

Проведенное исследование позволило разработать алгоритм отбора предикторов для прогнозирования ВГЛ после КШ и построить модели, имеющие высокую прогностическую точность. Модели на основе методов МО СЛ и ИНС показали высокую точность и позволили выделить несколько дополнительных предикторов ВГЛ после КШ.

Список литературы

- [**Бойцов и др., 2015**] Бойцов С.А., Калинина А.М., Ипатов П.В. Диспансеризация взрослого населения, как механизм выявления сердечно-сосудистых заболеваний и формирования диспансерного наблюдения // Вестник Росздравнадзора. 2015. Т. 5.
- [**Breiman, 2001**] Breiman L. Random Forests // Machine Learning. 2001. No. 45(1).
- [**Faraway et al., 2016**] Faraway J. Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models, Second Edition // Chapman & Hall/CRC Texts in Statistical Science Book 124), 2nd edition, Kindle Edition, 2016.
- [**Flach, 2012**] Flach P. Machine Learning. The Art and Science of Algorithms that Make Sense of Data. Cambridge. Cambridge University Press. 2012.
- [**Karim et al., 2018**] Karim, M.N., Reid, C.M., Huq, M. Predicting long-term survival after coronary artery bypass graft surgery // Interactive Cardiovascular Thoracic Surgery. 2018. Vol. 26(2).
- [**Kurt et al., 2008**], Kurt I., Ture M, Kurum AT. Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease // Expert Systems with Applications. 2008. Vol. 34(1).
- [**Nashef et al., 2012**] Nashef S.A.M., Roques F., Sharples L.D., Nilsson J., Smith C., Goldstone A.R., Lockowandt U. EuroSCORE II // European Journal of Cardio-Thoracic Surgery, 2012. Vol. 41(4).
- [**Sedaghat et al., 2013**] Sedaghat A., Sinning J.-M., Vasa-Nicotera M., Ghanem A., Hammerstingl C., Grube E. The revised EuroSCORE II for the prediction of mortality in patients undergoing transcatheter aortic valve implantation // Clin. Res. Cardiol. 2013. Vol. 102(11).
- [**Yamaoka et al., 2015**] Yamaoka H., Kuwaki K., Inaba H., Yamamoto T., Kato T.S., Dohi S. Comparison of modern risk scores in predicting operative mortality for patients undergoing aortic valve replacement for aortic stenosis // J. Cardiol. 2015. PII: S0914-5087(15).