

## МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ РИСКА РАЗВИТИЯ ФИБРИЛЛЯЦИИ ПРЕДСЕРДИЙ ПОСЛЕ КОРОНАРНОГО ШУНТИРОВАНИЯ<sup>1\*</sup>

*Шахгельдян К.И., д.т.н., профессор,  
директор института Информационных технологий ВГУЭС,  
заведующий лабораторией  
анализа больших данных в биомедицине и  
здравоохранении ДВФУ  
e-mail: [carinash@vvsu.ru](mailto:carinash@vvsu.ru)*

*Рублев В.Ю., врач сердечно-сосудистый хирург,  
аспирант ДВФУ  
e-mail: [dr.rublev.v@gmail.com](mailto:dr.rublev.v@gmail.com)*

*Гельцер Б.И., д.м.н., профессор, чл.-корр.  
РАН, директор департамента клинической медицины школы биомедицины ДВФУ  
e-mail: [boris.geltser@vvsu.ru](mailto:boris.geltser@vvsu.ru)*

*Щеглов Б.О., лаборант школы биомедицины ДВФУ  
e-mail: [b.shcheglov@mail.ru](mailto:b.shcheglov@mail.ru)*

*Щеглова С.Н., к.пед.н., доцент, доцент  
кафедры точных и естественных наук СВГУ  
e-mail: [schegloff@mail.ru](mailto:schegloff@mail.ru)*

### 1. ВВЕДЕНИЕ

Внедрение технологий искусственного интеллекта (ИИ) в различные области клинической медицины позволяет повысить качество медицинской помощи за счет расширения диагностических возможностей и улучшения прогностических оценок. Так, например, в кардиохирургии с помощью алгоритмов машинного обучения и других методов интеллектуального анализа данных разрабатываются прогностические шкалы оценки индивидуального риска пациента, которые позволяют оптимизировать тактику ведения лечения и минимизировать вероятность послеоперационных осложнений.

К одному из наиболее частых осложнений кардиохирургической практики относится послеоперационная фибрилляция предсердий (ПоФП), которая, в зависимости от типа оперативного вмешательства отмечается у 25-40% пациентов [1, 2]. Анализ публикаций по прогнозированию фибрилляции предсердий (ФП) у больных ишемической болезнью сердца (ИБС) с помощью методов машинного обучения показал эффективность применения искусственных нейронных сетей с целью разработки точных прогностических моделей. Так, в исследовании Sidrah Liaqat и соавт. [3] разработали сверточную нейронную сеть с метрикой качества площадь под ROC-кривой (AUC), равной 0.81, LSTM (AUC - 0.83) и сверточную LSTM (AUC - 0.80). Исследователями Hill N.R. и соавт. [4] на основе данных электрокардиографии (ЭКГ) была разработана модель нейронной сети с AUC 0,83. Для прогноза развития ФП были получены и более точные модели. В исследовании Yong Xia и соавт. [5] получена модель сверточной нейронной сети с чувствительностью 98,34%, специфичностью 98,24% и точностью 98,29%. Lown M. и соавт. [6] была создана модель SVM с чувствительностью 99,2% и специфичностью 99,5%.

В кардиохирургии важнейшее значение играет риск развития ПоФП. Ее возникновение после коронарного шунтирования (КШ) в 4 раза повышает риск развития инсульта, послеоперационных кровотечений и острой почечной недостаточности и в 2 раза увеличивает вероятность смертности на 30-дневном и 6-месячном горизонтах наблюдения [1, 7]. Несмотря на большое количество исследований, до настоящего времени не существует единой патофизиологической концепции, подробно описывающей механизмы развития ПоФП и унифицированной шкалы для расчета риска ее возникновения после КШ. Однако, в ряде работ с помощью алгоритмов, применяемых для бинарной классификации (логистической регрессии (ЛР), метода случайного леса (СЛ), деревьев решений и многослойного перцептрона) были

<sup>1\*</sup> Работа выполнена при финансовой поддержке РФФИ, проекты №18-29-03131 и №19-29-01077

выделены ряд предикторов и разработаны прогностические модели [8, 9]. Например, в исследовании Mathew J.P. и соавт. [10] показатель точности модели ЛР по метрике AUC составил 0,7. Выборка включала 1503 пациентов с ИБС после КШ, а в качестве предикторов были определены возраст больных, ФП у пациентов в дооперационном периоде, наличие операций на клапанах сердца в анамнезе, послеоперационная отмена  $\beta$ -адреноблокаторов и ингибиторов ангиотензинпревращающего фермента. В ретроспективном исследовании GÜngör и соавт. [11] на данных 125 пациентов была получена модель ЛР, включающая в качестве предикторов возраст пациентов, тромбоцитарно-лимфоцитарный индекс. Точность модели по метрике AUC составила 0,634, а чувствительность и специфичность - 64% и 56%, соответственно. В исследовании Magee M.J. и соавт. [12] разработанная ЛР-модель включала 14 факторов (AUC 0,72). К наиболее значимым предикторам отнесились дооперационные и интраоперационные факторы: возраст пациентов, потребность в длительной искусственной вентиляции (24 часа и более), использование искусственного кровообращения и наджелудочковые аритмии в анамнезе. Zaman A.G. и соавт. [13] в ходе анализа данных 326 пациентов (среди которых ПоФП наблюдалась у 92) с помощью ЛР выявили следующие предикторы в модели при прогнозировании ФП с чувствительностью и специфичностью в 49% и 84% случаев соответственно: возраст пациентов (>75 лет), увеличение длительности зубца P (>155мс) и мужской пол. В исследовании David Amaret и соавт. [14] с помощью ЛР разработали модель с AUC 0,69, которая включала следующие факторы: сумма отфильтрованных сигналов от 3 ортогональных отведений ЭКГ до операции, пожилой возраст пациентов и мужской пол. В исследовании Todorov и соавт. [15] на выборке из 999 пациентов была разработана многофакторная модель ЛР с AUC 0,68 и выделены следующие до- и послеоперационные предикторы: длительность нахождения на искусственной вентиляции легких и пребывания в отделении интенсивной терапии, пожилой возраст, повышение концентрации С-реактивного белка и креатинина в плазме крови. Для прогнозирования ПоФП после КШ многие исследователи применяли прогностические шкалы EuroSCORE [16], Koles [17], PAFAC (Predictors of AF After CABG) [18], CHA2DS2-VASc и HAS-BLED, рассчитанные с помощью ЛР, показывающие точность по метрике AUC от 0,65 до 0,68 [19, 20].

Таким образом, данные литературы указывают на высокую точность прогноза ФП у больных ИБС, но недостаточную точность разработанных ранее моделей для прогноза впервые выявленной ПоФП после КШ, что побудило авторов к более детальному анализу данной проблемы. Цель исследования состояла в разработке в совершенствовании алгоритма отбора предикторов в прогнозировании ПоФП у больных ИБС после КШ.

## 2. МАТЕРИАЛЫ И МЕТОДЫ

Разработка моделей выполнялась на основе данных ретроспективного исследования, содержащих электронные истории болезни (ЭИБ) 886 больных ИБС (181 женщина и 685 мужчин) в возрасте от 35 до 81 года с медианой (МЕ) 63 года и 95% доверительным интервалом (ДИ) [63; 64], которым выполнялось изолированное КШ в условиях искусственного кровообращения в период с 2008 по 2019 гг. на базе кардиохирургического отделения ГБУЗ «Приморская краевая клиническая больница №1» г. Владивостока. Из исследования были исключены больные с любой формой ФП в анамнезе. Общее количество таких больных составило 85 человек. Таким образом, датасет был представлен результатами обследования 801 больного ИБС. Верификация ПоФП осуществлялась по результатам мониторинга ЭКГ в течение не менее 96 часов после КШ. Среди обследованной когорты было выделено 2 группы лиц. В первую из них вошли 153 (19,1%) больных, у которых в послеоперационном периоде были зарегистрированы пароксизмы ФП, во вторую – 648 (80,9%) больных без нарушений сердечного ритма. В первой группе госпитальная летальность составила 9,8% (15 пациентов), а во второй - 4,6% (30 пациентов).

Конечная точка исследования представляла собой ПоФП в форме категориального бинарного признака (“отсутствие” или “развитие”). Входные признаки – подгруппа 100 потенциальных предикторов выражалась в форме непрерывных и категориальных переменных. Для обработки и анализа данных использовали методы статистического анализа и машинного обучения. Первые из них были представлены тестами Хи-квадрат, Фишера, Манна-Уитни и однофакторной ЛР с расчетом весовых коэффициентов. Вторые - многофакторной ЛР, стохастическим градиентным бустингом (СГБ) и искусственной нейронной сетью (ИНС). Архитектура ИНС, представляющей многослойный перцептрон, подбиралась путем максимизации площади под ROC-кривой (AUC) и в окончательном виде состояла из двух скрытых слоев по 90 и 80 нейронов в каждом. В качестве функции активации ИНС использовались “sigmoid” и “relu”. В окончательные модели с учетом лучшей точности ИНС вошла только функция “sigmoid”. Точность моделей оценивали по 4 метрикам качества: AUC, чувствительность (Sen), специфичность (Spec) и точность (Acc). Разработка моделей включала процедуру кросс-валидации методом k-блоков. Модели были построены на обучающей выборке (9/10) пациентов и верифицированы на тестовой (1/10). Все значения метрик качества, приведенные в работе, получены путем усреднения оценок прогноза на тестовых выборках.

Дизайн исследования включал 4 этапа. На первом из них применяли статистический анализ, с помощью которого проводили межгрупповые сравнения потенциальных предикторов ПоФП. Для непрерывных переменных использовали тест Манна-Уитни, так как предварительная оценка близости данных к нормальному распределению методом Шапиро-Уилки показала отрицательный результат. Для сравнения категориальных переменных использовали критерий хи-квадрат, а для оценки отношения шансов (ОШ) и их ДИ - тест Фишера. На втором этапе с помощью данных методов определяли границы значений анализируемых факторов с наилучшим предиктивным потенциалом. Данная процедура включала проверку гипотез о равенстве распределений признаков в группах сравнения. Выделение прогностически значимых диапазонов осуществляли с шагом тестирования 0,05-0,1 усл. ед. для различных показателей. Критериям отбора соответствовали границы значений факторов, p-value которых имели минимальную, а ОШ - максимальную величину. На третьем этапе исследования по нормированным признакам с помощью однофакторной ЛР определяли весовые коэффициенты, соответствующие значимости влияния отдельных признаков на развитие ПоФП. На четвертом этапе были разработаны многофакторные модели на основе ЛР, СГБ и ИНС, структура которых пошагово дополнялась потенциальными предикторами ПоФП с оценкой метрик качества. При увеличении значений последних считали, что включенный в модель показатель может рассматриваться в качестве предиктора ПоФП. Обработка и анализ данных выполнялись на языке R в среде R-studio и языке Python с помощью пакетов keras, tensorflow и xgboost.

### 3. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Сравнительный межгрупповой (пациенты с и без ПоФП) анализ факторов, характеризующих дооперационный клиничко-функциональное состояние пациентов после КШ, показал, что достоверно значимые межгрупповые различия фиксировались только по 10 показателям: возраст; составляющие части электрокардиограмм - длительность сегмента PQ, комплекса деполяризации желудочков QRS, интервала RR, электрической систолы сердца QT; эхокардиографические параметры сердца - верхне-нижний (LA3) и поперечный (LA2) размеры левого предсердия (ЛП), продольный (RA1) и поперечный (RA2) размеры правого предсердия (ПП), конечный систолический размер (КСР) левого желудочка (ЛЖ); наличие недостаточности трикуспидального клапана (НТК) в анамнезе; фракция выброса (ФВ) крови ЛЖ во время систолы (табл. 1).

Наибольшая статистически значимая разница в группах пациентов с наличием или отсутствием ПоФП наблюдалась у следующих факторов: RA1, RA2 и длительности интервала QT (p-value < 0,0001). Относительно пациентов без аритмии у больных ПоФП показатели ЛП (LA2, LA3) были достоверно значимо выше, а длительность QRS статистически значимо меньше. Дальнейший анализ показал, что попарное произведение показателей LA2 и LA3, но и RA1 и RA2 имели более высокий уровень достоверной значимости межгрупповых различий, чем оценка влияния отдельных геометрических параметров камер сердца. Менее заметные, но статистически значимые отличия пациентов с послеоперационной аритмией ассоциировались с более старшим возрастом, увеличением КСР ЛЖ, продолжительностью интервала QRS и наличием НТК.

**Таблица 1. Клиничко-функциональная характеристика больных**

Показатели	Объем выборки	1 группа, n=153	2 группа, n= 648	p-value
Возраст, лет	801	64 [63; 66]	63[62; 64]	0,00076
ФВ, %	783	59 [57;60]	60[60; 60]	0,039
LA2, мм	734	41 [40; 42]	39 [39; 40]	0,026
LA3, мм	734	38 [37; 39]	37 [36; 37]	0,013
LA2*LA3, мм <sup>2</sup>	734	160 [147; 168]	144 [141; 148]	0,011

RA1, мм	734	39,5 [3,9; 40]	37 [36; 37]	0,00007
RA2, мм	734	43 [41; 43]	39 [38; 40]	0,00003
RA1 * RA2, мм <sup>2</sup>	734	164[160; 176]	144 [140;148]	0,000012
PQ, мс	801	160 [150; 160]	150[140; 150]	0,1
QRS, мс	801	80[80; 100]	100[80; 100]	0,0019
RR, мс	761	936,5[909; 1000]	920[882,4; 950]	0,22
QT, мс	761	400 [400; 410]	400[380; 400]	0,00012
КСР ЛЖ, мм	733	350 [330; 360]	340[ 330; 350]	0,037
НТК, абс (%)	801	34 (22,2%)	79 (12,2%)	0,002

На втором этапе исследования среди показателей, имеющих достоверные различия в группах сравнения, верифицировали диапазоны их значений, обладающие наибольшим предиктивным потенциалом (табл. 2). Были выделены следующие границы непрерывных факторов, имеющие статистически значимое влияние на развитие ПоФП: возраст мужчин старше 55 и младше 75 лет и женщины старше 60 и младше 79 лет (ОШ = 3,4,  $p < 0,0001$ ); длительность сегмента PQ больше 160 и меньше 200 мс (ОШ=2,2,  $p=0,0004$ ), комплекса деполяризации желудочков QRS больше 50 и меньше 100 мс (ОШ=1,5,  $p=0,021$ ), интервал RR в диапазоне от 1000 до 1100 мс, электрической систолы сердца QT больше 420 мс; ЭхоКГ параметры сердца - произведение LA3 на LA2 больше 150 мм<sup>2</sup>, произведение RA1 на RA2 больше 160 мм<sup>2</sup>, КСР ЛЖ больше 4,9 см с повышением риска в 2,9 раза; ФВ более 45 и менее 60 % увеличивала риск возникновения осложнения в виде мерцательной аритмии в 1,7 раза ( $p=0,0058$ ). Процедура поиска границ непрерывных переменных позволила выделить не только значимые диапазоны предикторов и таким образом определить факторы риска, но и выделить дополнительные предикторы, которые ранее не были таковыми верифицированы (PQ и RR).

**Таблица 2. Границы значений факторов риска ПоФП с наилучшим предиктивным потенциалом**

Предиктор и его граница	1 группа, n=153	2 группа, n= 648	ОШ, 95% ДИ	p-value
Возраст, лет М 55- 74 Ж 60-78	142 (92,8 %)	511 (78,9%)	3,4 [1,9; 6,9]	0,0001
ФВ,% 45-60%	81 (52,9%)	268 (41,4%)	1,68[1,17; 2,42]	0,0058
КСР > 490 мм	7 (4,6%)	11 (1,7%)	2,9 [1,05; 7,7]]	0,049
LA2*LA3 >160 мм <sup>2</sup>	67 (43,8%)	191 (29,5%)	2,1 [1,4; 3]	0,0002
RA1 * RA2 >150 мм <sup>2</sup>	92 (60%)	269 (41,5%)	2,5[1,7; 3,8]	<0,0001

PQ 170-210 мс	38 (24,8%)	47 (7,3%)	2,2 [2,6; 6,8]	0,0004
QRS 50-80 мс	88 (57,5%)	303 (46,8%)	1,53 [2,2; 14,4]	0,021
RR 1000-1100 мс	70 (45,8 %)	175 (27%)	2 [1,4; 2,9]	0,00033
QT, мс М>420 мс Ж>440 мс	45 (30,8%)	100 (16,3%)	2,3 [1,5; 3,5]	0,00013
НТК, абс (%)	34 (22,2%)	79 (12,2%)	2 [1,3; 3,2]	0,002

Примечание: М – мужчины; Ж – женщины

Для верификации возможных взаимосвязей факторов риска с вероятностью развития ПоФП на третьем этапе исследования нами были построены однофакторные модели ЛР с расчетом весовых коэффициентов, характеризующих предсказательную ценность анализируемых показателей (табл.3). Данный подход позволяет более детально оценить влияние потенциальных предикторов на результирующую переменную. Наибольшее влияние на ПоФП оказывали возраст, QRS и PQ.

**Таблица 3. Весовые коэффициенты однофакторных моделей ЛР для оценки риска ПоФП**

Показатель	Коэффициент	p-value
Возраст, лет М 55- 74 Ж 60-78	1,24	0,00015
ФВ: 45 - 60%	0,42	0,03
LA2*LA3 >160 мм <sup>2</sup>	0,73	0,00016
КСР ЛЖ > 490 мм	1,07	0,03
RA1 * RA2 >150 мм <sup>2</sup>	0,94	<0,0001
PQ 170 - 210 мс	1,44	<0,0001
QRS 50-80 мс	1,63	0,0005
RR 1000- 1100 мс	0,87	<0,0001
QT, мс М>420 мс Ж>440 мс	0,78	0,00016
НТК, абс (%)	0,72	0,0016

На четвертом этапе исследования, используя пошаговое включение предикторов в непрерывной или в категориальной форме, были разработаны модели ЛР (табл. 3), СГБ (табл. 4) и ИНС (табл. 5). Одной из задач нашего исследования являлась оценка эффективности использования предикторов в категориальной форме по сравнению с их непрерывными аналогами. Повышение точности многофакторных моделей на основе ЛР фиксировалось в моделях 2-7 (табл.4). Максимальная точность по метрике Sen (0,67) была достигнута в моделях 6 и 7 при AUC 0,71 и 0,7 соответственно. Уровень Spes составил 0,66. Дальнейшее расширение возможных предикторов снижало метрику Sen, сохраняя уровни других метрик. Данный факт свидетельствовал о том, что такие параметры как КСР, ФВ и НТК связаны, возможно линейно, с показателями размеров предсердий, данных ЭКГ или возраста. Анализ моделей на основе ЛР, использующих категориальные переменные (модели 11-20 в табл.4), показал, что точность моделей повышалась аналогично непрерывному случаю, но она могла достигать лучших значений.

Например, сравнение моделей 7 и 17, показывает, что последняя имеет лучшую точность по всем 4 метрикам качества (0,69 vs 0,67, 0,71 vs 0,66, 0,74 vs 0,7 и 0,71 vs 0,67). Дальнейшее расширение категориальных предикторов моделей 18-20 существенно не меняло точность прогноза.

**Таблица 4. Оценка точности прогностических моделей ПоФП на тестовых выборках для многофакторной ЛР**

N	Предикторы	Граница предиктора	Метрики точности			
			Sen	Spec	AUC	ACC
1	Возраст	-	0.46	0.63	0.58	0.59
2	Возраст, PQ	-	0.51	0.62	0.59	0.6
3	Возраст, PQ, QRS	-	0.6	0.6	0.62	0.6
4	Возраст, PQ, QRS, RR	-	0.56	0.61	0.63	0.6
5	Возраст, PQ, QRS, RR, QT	-	0.62	0.61	0.67	0.61
6	Возраст, PQ, QRS, RR, QT, RA1*RA2	-	0.67	0.66	0.71	0.66
7	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3	-	0.67	0.66	0.7	0.67
8	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСР	-	0.64	0.68	0.71	0.66
9	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСР, ФВ	-	0.62	0.68	0.7	0.66
10	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСР, ФВ, НТК	-	0.58	0.7	0.7	0.68
11	Возраст	<i>М 55- 74 Ж 60-78</i>	0.93	0.21	0.57	0.35
12	Возраст + PQ	<i>170-210 мс</i>	0.25	0.93	0.64	0.8
13	Возраст, PQ + QRS	<i>50-80 мс</i>	0.72	0.46	0.66	0.5
14	Возраст, PQ, QRS + RR	<i>1000-1100 мс</i>	0.56	0.75	0.7	0.71
15	Возраст, PQ, QRS, RR + QT	<i>М &gt; 420 мс Ж &gt; 440 мс</i>	0.67	0.69	0.72	0.69
16	Возраст, PQ, QRS,	<i>&gt; 160 мм2</i>	0.63	0.74	0.75	0.72

	RR, QT + RA1*RA2					
17	Возраст, PQ, QRS, RR, QT, RA1*RA2 + LA1*LA2	> 150 мм <sup>2</sup>	0.69	0.71	0.74	0.71
18	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3 + КСР	>490 мм	0.68	0.71	0.75	0.71
19	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСР + ФВ	45-60%	0.68	0.72	0.75	0.71
20	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСР, ФВ, НТК	-	0.69	0.71	0.75	0.71

Модели на основе СГБ разрабатывали аналогичным образом (табл.5). Увеличение точности прогностической оценки фиксировалось во 2-5 моделях СГБ. Полученные в 5-ой модели 4 метрики качества являются лучшими на текущий момент прогностическими оценками по сравнению со всеми ранее опубликованными результатами: 0,82, 0,76, 0,77 и 0,76 для AUC, ACC, Sen и Spec соответственно. Дальнейшее включение в модели на основе СГБ статистически значимых предикторов в непрерывной форме снижало качество прогноза (модели 6-10). Модели СГБ 11-20, разработанные с использованием предикторов в категориальной форме, достигали пика точности при том же наборе показателей что и модели ЛР (модели 17 табл. 4 и 5). Дальнейшее расширение спектра категориальных предикторов не приводило к изменению качества прогноза. Сопоставимыми по уровню точности были модели на основе ЛР и СГБ.

**Таблица 5. Оценка точности прогностических моделей ПоФП на тестовых выборках для СГБ**

N	Предикторы	Граница предиктора	Метрики точности			
			Sen	Spec	AUC	ACC
1	Возраст	-	0.5	0.59	0.54	0.57
2	Возраст, PQ	-	0.58	0.66	0.65	0.64
3	Возраст, PQ, QRS	-	0.66	0.63	0.71	0.64
4	Возраст, PQ, QRS, RR	-	0.76	0.76	0.8	0.76
5	Возраст, PQ, QRS, RR, QT	-	0.77	0.76	0.82	0.76
6	Возраст, PQ, QRS, RR, QT, RA1*RA2	-	0.72	0.76	0.8	0.75
7	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3	-	0.68	0.78	0.79	0.76

8	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСР	-	0.69	0.75	0.78	0.74
9	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСР, ФВ	-	0.68	0.76	0.78	0.75
10	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСР, ФВ, НТК	-	0.68	0.76	0.78	0.75
11	Возраст	<i>М 55-74 Ж 60-78</i>	0.93	0.21	0.57	0.35
12	Возраст + PQ	<i>170-210 мс</i>	0.25	0.93	0.64	0.8
13	Возраст, PQ + QRS	<i>50-80 мс</i>	0.72	0.46	0.66	0.5
14	Возраст, PQ, QRS + RR	<i>1000-1100 мс</i>	0.56	0.75	0.7	0.71
15	Возраст, PQ, QRS, RR + QT	<i>М &gt; 420 мс Ж &gt; 440 мс</i>	0.67	0.69	0.72	0.69
16	Возраст, PQ, QRS, RR, QT + RA1*RA2	<i>&gt; 160 мм<sup>2</sup></i>	0.63	0.74	0.75	0.72
17	Возраст, PQ, QRS, RR, QT, RA1*RA2 + LA1*LA2	<i>&gt; 150 мм<sup>2</sup></i>	0.69	0.71	0.74	0.71
18	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3 + КСР	<i>&gt; 490 мм</i>	0.68	0.71	0.75	0.71
19	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСР + ФВ	<i>45-60%</i>	0.68	0.72	0.75	0.71
20	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСР, ФВ + НТК	-	0.69	0.71	0.75	0.71

Для моделей на основе ИНС фиксировалось увеличение точности при расширении спектра используемых предикторов как в непрерывной, так и в категориальной форме (табл. 6). Лучшая модель ИНС получена при включении всех значимых предикторов в категориальной форме (Sen - 0,74, Spec - 0,73, AUC - 0,75, ACC - 0,73). Точность последней уступает точности лучшей модели на основе СГБ, но превышает точность модели ЛР, при этом использование ИНС позволяет подтвердить значимость влияние рассматриваемых предикторов.

**Таблица 6. Оценка точности прогностических моделей ПоФП на тестовых выборках для ИНС**

N	Предикторы	Граница предиктор	Метрики точности
---	------------	-------------------	------------------

		a	Sen	Spec	AUC	ACC
1	Возраст	-	0.16	0.86	0.56	0.7
2	Возраст, PQ	-	0.2	0.9	0.59	0.73
3	Возраст, PQ, QRS	-	0.5	0.77	0.71	0.6
4	Возраст, PQ, QRS, RR	-	0.6	0.68	0.63	0.64
5	Возраст, PQ, QRS, RR, QT	-	0.63	0.68	0.68	0.65
6	Возраст, PQ, QRS, RR, QT, RA1*RA2	-	0.65	0.68	0.71	0.67
7	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3	-	0.68	0.7	0.69	0.69
8	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСР	-	0.69	0.72	0.71	0.7
9	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСР, ФВ	-	0.7	0.71	0.71	0.71
10	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСР, ФВ, НТК	-	0.71	0.72	0.72	0.71
11	Возраст	<i>М 55- 74 Ж 60-78</i>	0	1.0	0.55	0.81
12	Возраст + PQ	<i>170-210 мс</i>	0.24	0.94	0.62	0.8
13	Возраст, PQ + QRS	<i>50-80 мс</i>	0.38	0.8	0.67	0.73
14	Возраст, PQ, QRS + RR	<i>1000-1100 мс</i>	0.58	0.74	0.71	0.71
15	Возраст, PQ, QRS, RR + QT	<i>М&gt;420 мс Ж&gt;440 мс</i>	0.68	0.68	0.72	0.68
16	Возраст, PQ, QRS, RR, QT + RA1*RA2	<i>&gt; 160 мм2</i>	0.63	0.75	0.75	0.73
17	Возраст, PQ, QRS, RR, QT, RA1*RA2 + LA1*LA2	<i>&gt; 150 мм2</i>	0.65	0.74	0.75	0.72
18	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСР	<i>&gt;490 мм</i>	0.65	0.74	0.74	0.73
19	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3	<i>45-60%</i>	0.7	0.71	0.74	0.71

	,КСП + ФВ					
20	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСП, ФВ + НТК	-	0.74	0.73	0.75	0.73

#### 4. ЗАКЛЮЧЕНИЕ

В данной работе представлены модели прогнозирования развития ПоФП после КШ, разработанные на основе методов статистического анализа и машинного обучения: межгруппового сравнения, поиска значимых диапазонов, однофакторной ЛР, многофакторных ЛР, СГБ и ИНС. Поиск предикторов выполнялся из набора 100 дооперационных характеристик пациентов с ИБС. Использование статистических методов позволило выделить наиболее значимые факторы. При этом удалось выдвинуть гипотезы о границах диапазонов, формирующих факторы риска. К факторам с наибольшим предиктивным потенциалом относились возраст больных (55-74 года для мужчин, 60-78 лет - для женщин), RR 1000-1100 мс, QRS 50-80 мс, PQ 170-210 мс, QT (>420 мс для мужчин; >440 мс - для женщин), произведение линейных размеров ЛП > 160 мм<sup>2</sup> и ПП > 150 мм<sup>2</sup>, наличие НТК, ФВ 45-60%, КСП ЛЖ > 490 мм. Эти гипотезы были подтверждены однофакторными на основе ЛР и многофакторными ЛР и ИНС моделями. Часть предикторов (ФВ, НТК, КСП) имела нелинейные взаимосвязи с конечной точкой и не повышала точность моделей ЛР. Подтверждением влияния этих факторов на ПоФП являлось пошаговое увеличение метрик качества моделей ИНС. Использование показателей в категориальной форме не вело к снижению точности моделей во всех рассматриваемых случаях.

В тоже время, наилучшую модель авторы получили при использовании СГБ с непрерывными наиболее значимыми факторами. Данная модель для прогнозирования впервые выявленной ПоФП после КШ по 4 метрикам качества (AUC - 0,82, ACC - 0,76, Sen - 0,77 и Spec - 0,76) является наилучшей по сравнению с ранее опубликованными. Расширение предикторов моделей СГБ снижало точность прогноза. Использование показателей в категориальной форме для СГБ также не улучшило качество моделей по сравнению с непрерывными аналогами.

Полученные в ходе исследования предикторы объясняются современными представлениями о патофизиологии ИБС и ФП. Так, изначально (до операции) измененный размер предсердия, нарушенная сократительная функция ЛЖ и отклонения показателей ЭКГ усугубляется на фоне операционного стресса представляются субстратом, способствующим возникновению патологической аритмии. Неоспоримым фактом является сложность клинической интерпретация СГБ и ИНС, но в рассматриваемом эксперименте, некоторые, применяемые для моделирования предикторы, являются в основе своей широко распространенными маркерами функционального состояния больных и их предиктивное значение показано в ряде работ [21]. Однако, были выявлены и ранее не используемые предикторы, такие как, конечный систолический размер, недостаточность трикуспидального клапана, размеры ПП, что является задачей дальнейшего исследования.

#### Литература

1. Kirchhof, P. Разработано при особом участии Европейской ассоциации ритма сердца (EHRA). Одобрено Европейской организацией по изучению инсульта (ESO). Рекомендации ESC по лечению пациентов с фибрилляцией предсердий, разработанные совместно с EACTS / P. Kirchhof, S. Benussi, D. Kotecha, A. Ahlsson. Научное редактирование перевода выполнено Лебедевым Д. С., заведующим НИО аритмологии НМИЦ им. В. А. Алмазова, профессором РАН. Российский кардиологический журнал 2017, 7 (147): 7–86 <http://dx.doi.org/10.15829/1560-4071-2017-7-7-86>
2. Российское кардиологическое общество Национальное общество по изучению атеросклероза Национальное общество по атеротромбозу / "Клинические рекомендации "Стабильная ишемическая болезнь сердца" (утв. Минздравом России) // [Электронный ресурс] / Режим доступа: <https://legalacts.ru/doc/klinicheskie-rekomendatsii-stabilnaja-ishemicheskaja-bolezn-serdtsa-utv-minzdravom-rossii/>
3. Liaqat S, Dashtipour K, Zahid A, Assaleh K, Arshad K, Ramzan N. Detection of Atrial Fibrillation Using a Machine Learning Approach. Information. 2020; 11(12):549. <https://doi.org/10.3390/info11120549>
4. Hill NR, Ayoubkhani D, McEwan P, Sugrue DM, Farooqui U, Lister S, et al. (2019) Predicting atrial fibrillation in primary care using machine learning. PLoS ONE 14(11): e0224582. <https://doi.org/10.1371/journal.pone.0224582>
5. Yong Xia, Naren Wulan, Kuanquan Wang, Henggui Zhang, Detecting atrial fibrillation by deep convolutional neural networks, Computers in Biology and Medicine, Volume 93, 2018, Pages 84-92, ISSN 0010-4825, <https://doi.org/10.1016/j.compbiomed.2017.12.007>.
6. Lown M, Brown M, Brown C, Yue AM, Shah BN, Corbett SJ, et al. (2020) Machine learning detection of Atrial Fibrillation using wearable technology. PLoS ONE 15(1): e0227401. <https://doi.org/10.1371/journal.pone.0227401>

7. Skaria, R., Parvaneh, S., Zhou, S., Kim, J., Wanjiru, S., Devers, G., Konhilas, J., & Khalpey, Z. (2020). Path to precision: prevention of post-operative atrial fibrillation. *Journal of thoracic disease*, 12(5), 2735–2746. <https://doi.org/10.21037/jtd-19-3875>
8. Christalin, Beulah & Jeeva, S.. (2019). Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. *Informatics in Medicine Unlocked*. 16. 100203. 10.1016/j.imu.2019.100203.
9. Шахгельдян К.И., Рублев В.Ю., Гельцер Б.И., Щеглов Б.О. и др. Оценка предиктивного потенциала дооперационных факторов риска фибрилляции предсердий у больных ишемической болезнью сердца после коронарного шунтирования. *Сибирский журнал клинической и экспериментальной медицины*. 2020; 35(4):128–136. <https://doi.org/10.29001/2073-8552-2020-35-4-128-136>.
10. Mathew JP, Fontes ML, Tudor IC, et al. A Multicenter Risk Index for Atrial Fibrillation After Cardiac Surgery. *JAMA*. 2004;291(14):1720–1729. doi:10.1001/jama.291.14.1720
11. Güngör, Hasan & Babu, Abraham & Zencir, Cemil & Akpek, Mahmut & Selvi, Mithat & Erkan, Muhammet & Durmaz, Selim. (2017). Association of Preoperative Platelet-to-Lymphocyte Ratio with Atrial Fibrillation after Coronary Artery Bypass Graft Surgery. *Medical Principles and Practice*. 26. 164-168. 10.1159/000453614.
12. Magee, M. J., Herbert, M. A., Dewey, T. M., Edgerton, J. R., Ryan, W. H., Prince, S., & Mack, M. J. (2007). Atrial fibrillation after coronary artery bypass grafting surgery: development of a predictive risk algorithm. *The Annals of thoracic surgery*, 83(5), 1707–1712. <https://doi.org/10.1016/j.athoracsur.2006.12.032>
13. Zaman AG, Archbold RA, Helft G, Paul EA, Curzen NP, Mills PG. Atrial fibrillation after coronary artery bypass surgery: a model for preoperative risk stratification. *Circulation*. 2000;101(12):1403- 1408. doi:10.1161/01.cir.101.12.1403
14. David Amar, Weiji Shi, Charles W. Hogue, Hao Zhang, Rod S. Passman, Betsy Thomas, Peter B. Bach, Ralph Damiano, Howard T. Thaler, Clinical prediction rule for atrial fibrillation after coronary artery bypass grafting, *Journal of the American College of Cardiology*, Volume 44, Issue 6, 2004, Pages 1248-1253, ISSN 0735-1097, <https://doi.org/10.1016/j.jacc.2004.05.078>.
15. Todorov, Hristo & Janssen, Inka & Honndorf, Stefanie & Bause, Daniela & Gottschalk, Antje & Baasner, Silke & Volkert, Thomas & Färber, Valentin & Stover, John & Westphal, Martin & Ellger, Björn. (2017). Clinical significance and risk factors for new onset and recurring atrial fibrillation following cardiac surgery - a retrospective data analysis. *BMC Anesthesiology*. 17. 10.1186/s12871-017-0455-7.
16. Ad N., Holmes S.D., Patel J., Pritchard G., Shuman D.J., Halpin L. Comparison of EuroSCORE II, Original EuroSCORE, and The Society of Thoracic Surgeons Risk Score in Cardiac Surgery Patients. *Ann Thorac Surg*. 2016;102(2):573-579. doi:10.1016/j.athoracsur.2016.01.105
17. Kolek M.J., Muehlschlegel J.D., Bush W.S. et al. Genetic and clinical risk prediction model for postoperative atrial fibrillation. *Circ Arrhythm Electrophysiol*. 2015;8(1):25-31. doi:10.1161/CIRCEP.114.002300
18. Lin S.Z., Crawford T.C., Suarez-Pierre A. et al. A Novel Risk Score to Predict New Onset Atrial Fibrillation in Patients Undergoing Isolated Coronary Artery Bypass Grafting. *Heart Surg Forum*. 2018;21(6):E489-E496. Published 2018 Dec 13. doi:10.1532/hsf.2151
19. Mariscalco G., Biancari F., Zanobini M., et al. Bedside tool for predicting the risk of postoperative atrial fibrillation after cardiac surgery: the POAF score. *J Am Heart Assoc*. 2014;3(2):e000752. Published 2014 Mar 24. doi:10.1161/JAHA.113.000752
20. Burgos L.M., Seoane L., Parodi J.B. et al. Postoperative atrial fibrillation is associated with higher scores on predictive indices. *J Thorac Cardiovasc Surg*. 2019;157(6):2279-2286. doi:10.1016/j.jtcvs.2018.10.091
21. Ревишвили А.Ш., Попов В.А., Коростелев А.Н., Плотников Г.П., Мальшенко Е.С., Анищенко М.М. Предикторы развития фибрилляции предсердий после операции аортокоронарного шунтирования. *Вестник аритмологии*. 2018;(94):11-16; Revishvili A.S., Popov V.A., Korostelev A.N., Plotnikov G.P., Malysenko E.S., Anishchenko M.M. PREDICTORS OF NEW ONSET OF ATRIAL FIBRILLATION AFTER CORONARY ARTERY BYPASS GRAFTING SURGERY. *Journal of Arrhythmology*. 2018;(94):11-16.(In Russ.) DOI:10.25760/VA-2018-94-11-16.

## USING MACHINE LEARNING MODELS TO PREDICT THE RISK OF ATRIAL FIBRILLATION AFTER AORTOCORONARY BYPASSING

*Boris I. Geltser, Doctor of Medical Sciences, professor,  
Corresponding Member of Russian Academy of Sciences,  
Director of the Department of Clinical Medicine,  
Far Eastern Federal University, Vladivostok  
e-mail: [boris.geltser@vvsu.ru](mailto:boris.geltser@vvsu.ru)*

*Vladislav Yu. Rublev, cardiovascular surgeon, postgraduate  
student, Far Eastern Federal University, Vladivostok  
e-mail: [dr.rublev.v@gmail.com](mailto:dr.rublev.v@gmail.com)*

*Karina I. Shakhgeldyan, Doctor of Technical Sciences, professor,  
Head of the Institute of Information Technologies  
Vladivostok State University of Economics and Service  
Head of the Laboratory of Big Data Analysis in Biomedicine  
and Healthcare, Far Eastern Federal University, Vladivostok  
e-mail: [carinash@vvsu.ru](mailto:carinash@vvsu.ru)*

*Bogdan O. Shcheglov, laboratory assistant,  
Far Eastern Federal University, Vladivostok  
e-mail: [b.shcheglov@mail.ru](mailto:b.shcheglov@mail.ru)*

*Svetlana N. Shcheglova, Candidate of Pedagogical Sciences,  
Associate Professor of the Department of Exact and  
Natural Sciences of North-Eastern State University, Magadan  
e-mail: [schegloff@mail.ru](mailto:schegloff@mail.ru)*

### **Abstract**

In this paper authors perform a comparative analysis of the quality metrics of multivariate logistic regression, stochastic gradient boosting, and an artificial neural network models with continuous and boundary-binarized preoperational factors predicting the risk of atrial fibrillation after CABG in patients who do not have episodes of atrial fibrillation in the preoperative period.

**Keywords: predictive models, stochastic-gradient boosting, artificial neural network, features selection, post-operational fibrillation.**