

## Информационная система прогнозирования восстановления ритма у больных с фибрилляцией предсердий (ДИСК\_3)

Баллод Б.А., канд. техн. наук, Муромкина А.В., канд. мед. наук, Ковалев Д.Е., студ.

Приведены результаты многомерного интеллектуального анализа варибельности ритма сердца у больных с фибрилляцией предсердий. Разработана модель принятия решений и её программная реализация, позволяющая диагностировать пароксизмальную форму нарушения ритма сердца.

*Ключевые слова:* варибельность ритма сердца, диагностика, дерево решений.

## Information System for Forecasting Heart Rhythm Restoration Patients with Fibrillation of Auricles (DISK\_Z)

B.A. Ballod, Candidate of Engineering, A.V. Muromkina, Candidate of Medical Science, D.E. Kovalev, Student

The results of multidimensional intellectual analysis of heart rhythm variability of patients with fibrillation of auricles are given. Decision-making model and its program implementation allowing to diagnose Bouveret's disease the form of disturbance of heart rhythm are developed.

Keywords: heart rhythm variability, diagnostics, a tree of decisions.

К наиболее трудоемким задачам медицины относятся постановка диагноза и выбор курса лечения. Традиционно врачи решали эти задачи, полагаясь лишь на собственную интуицию и опыт. Сегодня в их арсенал все чаще входят способы, основанные на высоких технологиях и позволяющие обрабатывать большие потоки информации. Для этой цели применяются разнообразные регрессионные модели, основанные на применении технологии интеллектуального анализа данных (Data Mining): деревья решений, нейронные сети, карты Кохонена и др.

Анализ варибельности ритма сердца (ВРС) является неотъемлемой частью обследования кардиологических больных. В последние годы предпринимаются попытки оценки ВРС у больных с фибрилляцией предсердий (ФП). Разброс кардиоциклов объясняется отсутствием единого водителя ритма и особенностями атриоventрикулярного поведения при ФП.

Цель исследования – разработка интеллектуальной системы принятия решений о вероятности восстановления синусового ритма у больных с фибрилляцией предсердий.

На базе Data Mining легко создавать экспертные системы для диагностики заболеваний, которые используют накопленные данные клинических исследований, автоматически выявляют значимые признаки и моделируют сложные зависимости между симптомами и заболеваниями.

Для проведения анализа была составлена обучающая выборка, содержащая значения факторов ВРС пациентов кардиологического диспансера (объемом более 100). После исключения выбросов и пропусков объем выборки, фрагмент которой представлен в табл. 1, составил  $N = 85$ .

Таблица 1. Фрагмент обучающей выборки

ФИО	Пет***	Афе****	Ста*****	Фед****	Абр*****	и т.д.
Группа	А	А	В	А	В	
TP	11379	21513	29072	32528	24624	
VLF	1611	1726	2175	5393	2124	
LF	2094	4568	6572	9532	6790	
HF	7675	15219	20325	17602	15710	
RRNN	704	517	709	903	769	
SDNNf	131	149	190	195	175	
SDNNf  RRNNf	0,186	0,288	0,268	0,216	0,228	

**Примечания.** Группа А – пароксизмальная форма нарушения ритма; группа В – постоянная форма нарушения ритма; TP – общая мощность спектра ритмограммы; VLF – очень медленные волны; LF – медленные волны; HF – быстрые волны; RRNN – средняя продолжительность интервала между кардиоциклами; SDNNf – стандартное отклонение, квадратный корень из разброса интервалов RR; SDNNf |RRNNf – отношение стандартного отклонения к средней продолжительности интервала между кардиоциклами.

Задачи классификации с большим успехом решаются одним из методов Data Mining – при помощи деревьев решений. Деревья решений – классификатор, полученный из обучающего множества, содержащего объекты и их характеристики, на основе обучения. Дерево состоит из листьев, указывающих на класс, и узлов, содержащих правила классификации. Оно может использоваться для классификации объектов, не вошедших в обучающее множество. Получаемая модель – это способ представления правил в иерархической, последовательной структуре, где каждому объекту соответствует единственный узел, дающий решение.

Результат построения дерева приведен на рис. 1.

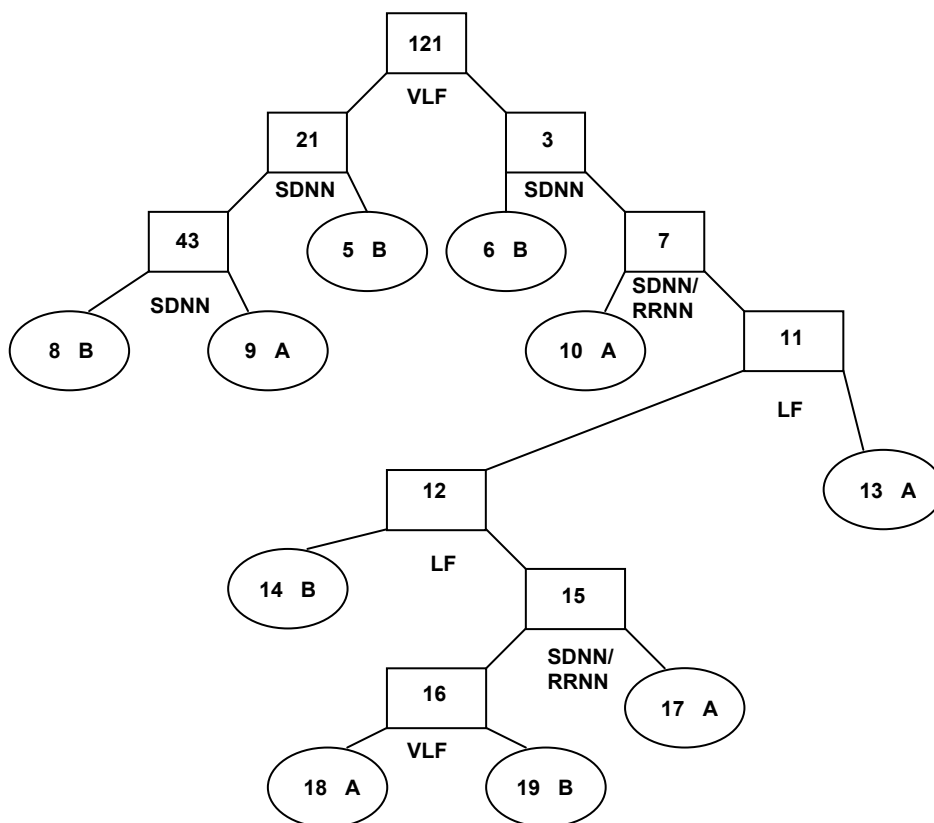


Рис. 1. Пример построения дерева решений (число узлов – 9; число терминальных вершин – 10)

Полученную модель можно использовать при определении класса заболевания, к которому следует отнести вновь поступившего больного. Среди множества алгоритмов построения деревьев решений часто используется алгоритм C&RT (сокращение от Classification And Regression Tree – «Дерево Классификации и Регрессии»). Основными отличиями алгоритма CART являются:

- бинарное представление дерева решений;
- функция оценки качества разбиения;
- механизм отсечения дерева;
- алгоритм обработки пропущенных значений;
- построение деревьев регрессии.

В алгоритме C&RT каждый узел дерева решений имеет двух потомков. На каждом шаге построения дерева правило, формируемое в узле, делит заданное множество примеров (обучающую выборку) на две части – часть, в которой выполняется правило (потомок right), и часть, в которой правило не выполняется (потомок left). Для выбора оптимального правила используется функция оценки качества разбиения.

В алгоритме C&RT оценка качества формализована в индексе *Gini*. Если набор данных *T* содержит данные *n* классов, тогда индекс *Gini* определяется как

$$G_{init}(T) = 1 - \sum_{i=1}^n p_i^2,$$

где  $p_i$  – вероятность (относительная частота) класса *i* в *T*.

При построении дерева все известные ситуации обучающей выборки сначала попадают в верхний узел, а потом распределяются по узлам, которые, в свою очередь, также могут быть разбиты на дочерние узлы. Критерий разбиения – различные значения какого-либо входного фактора. Для определения фактора, по которому будет происходить разбиение, используется показатель, называемый *энтропией* – мера неопределенности. Выбирается тот фактор, при разбиении по которому устраняется больше неопределенности. Неопределенность тем выше, чем больше примесей (объектов, относящихся к различным классам) находятся в одном узле. Энтропия равна нулю, если в узле будут находиться объекты, относящиеся к одному классу.

Построение дерева решений по алгоритму C&RT (полный перебор) реализовано соответствующим модулем пакета статистического анализа STATISTICA, диалоговое окно которого показано на рис. 2.

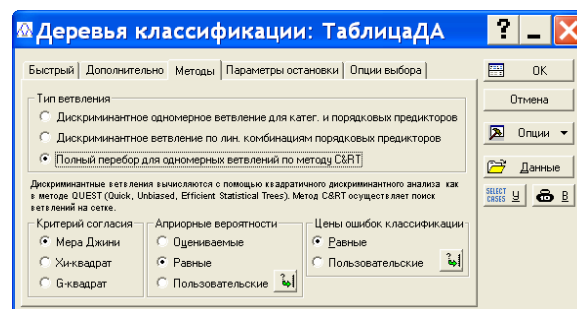


Рис. 2. Диалоговое окно системы STATISTICA

В табл. 2 и на рис. 3 представлены ошибки классификации на обучающей выборке объемом N = 85. Количество ошибок в группе А равно 5, в группе В – 3, что составляет менее 10 % от числа диагностируемых. Такой результат вполне пригоден для практического использования.

Таблица 2. Ошибки классификации

		Ошибки классификации на обуч. выборке (ТаблицаДА)	
		Предсказ. (строки) и наблюдаемые (столбцы)	
		Объем обуч. выб. N= 85	
Класс	Класс	а	в
а	а	5	
в	в	3	

Оценка значимости исходных факторов ВРС приведена в табл. 3, где среди статистических наиболее значимыми являются факторы SDNNf и SDNNARRNNf, а среди частотных – TP, HF, LF и VLF.

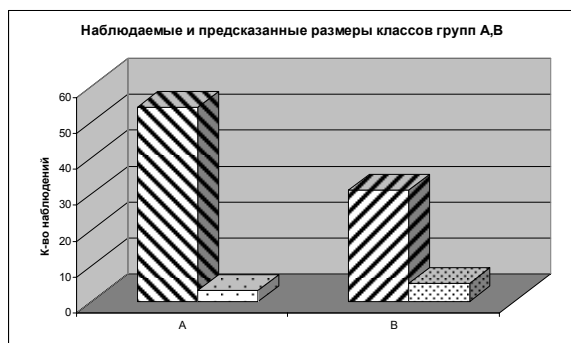


Рис. 3. Предсказываемые и наблюдаемые размеры классов А, В

Таблица 3. Ранги значимости факторов ВРС

		Ранги значимости предикторов (ТаблицаДА)	
		Одномерные ветвления	
		0 = низкая значимость; 100 = высокая	
Перемен.	Ранг		
TP	91		
VLF	78		
LF	84		
HF	86		
SDNNf	100		
SDNNARRNNf	83		

Общая структура дерева, содержащего 9 узлов и 10 вершин, представлена в табл. 4.

Таблица 4. Структура дерева решений

Структура (ТаблицаДА)							
Дочерние вершины, наблюдаемые, предсказанные классы, условия ветвления							
Узел	Левая ветвь	Правая ветвь	Класс а	Класс в	Предск. класс	Ветвл. по постоян.	Ветвл. по перемен.
1	2	3	54	31	а	-1462,5	VLF
2	4	5	12	20	в	-127,5	SDNNf
3	6	7	42	11	а	-128,0	SDNNf
4	8	9	12	12	в	-94,5	SDNNf
5			0	8	в		
6			0	2	в		
7	10	11	42	9	а	-0,2	SDNNARRNNf
8			2	10	в		
9			10	2	а		
10			14	0	а		
11	12	13	28	9	а	-10395,0	LF
12	14	15	14	8	а	-4727,5	LF
13			14	1	а		
14			5	0	а		
15	16	17	9	8	в	-0,3	SDNNARRNNf
16	18	19	6	8	в	-2191,5	VLF
17			3	0	а		
18			5	2	а		
19			1	6	в		

Для практического применения модель дерева решений была реализована в EXEL популярного офисного пакета MS Office. Интерфейс программы приведен на рис. 4.

Технология диагностирования предусматривает следующие этапы.

**Этап 1.** С помощью программы «Полиспектр» производится запись и первичный анализ кардиограммы больного (фрагмент протокола приведен на рис. 5).

**Этап 2.** Данные предварительного анализа заносятся в диалоговую форму программы ДИСК\_3, представленную на рис.5.

**Этап 3.** После выполнения расчета в окне «прогноз» выводится решение интеллектуальной системы: «ВОССТАНОВЛЕН» (или «АРИТМИЯ») (рис.5), что означает прогнозируемую оценку восстановления ритма сердца после соответствующего лечения. Такое решение принято, потому что в процессе диагностирования по результатам предварительного анализа больной был отнесен к группе В – с пароксизмальной формой нарушения ритма сердца, т.е. с исходом в восстановление ритма.

Одобренное врачом решение заносится в базу данных, которая служит обучающей выборкой (табл. 5) для дальнейшей корректировки модели дерева решений.

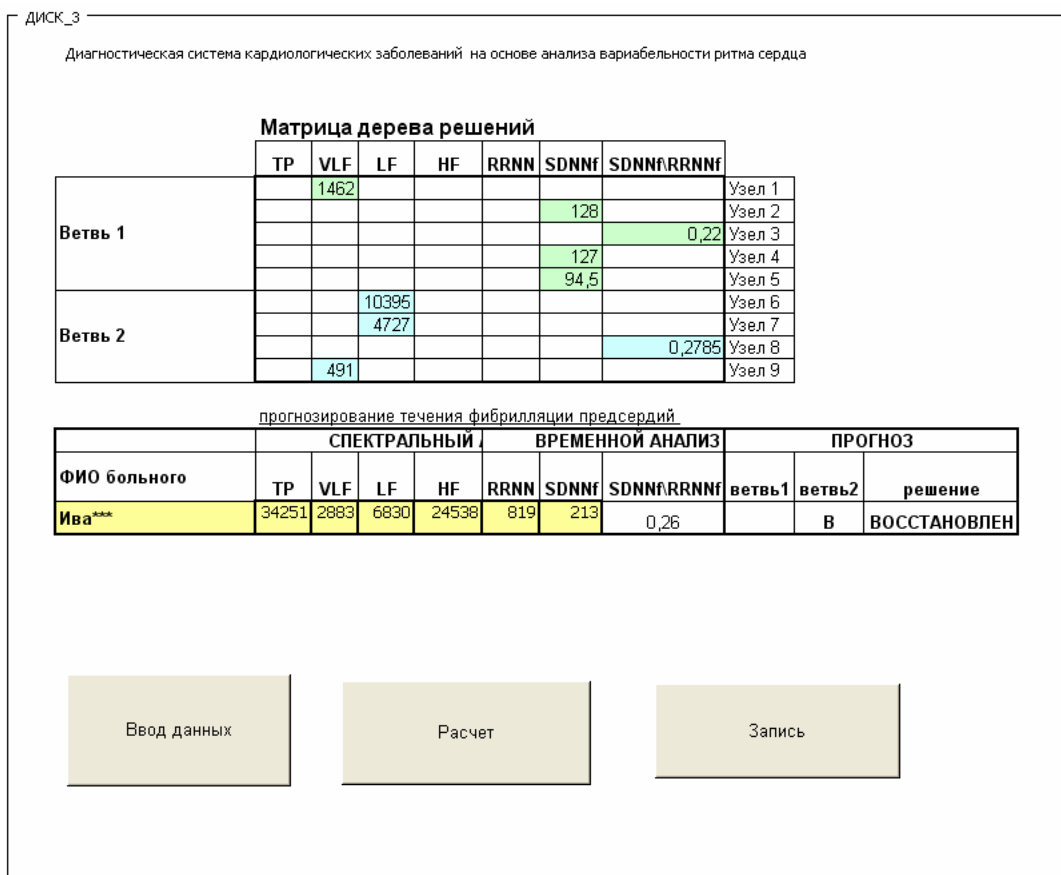


Рис. 4. Интерфейс программы диагностирования

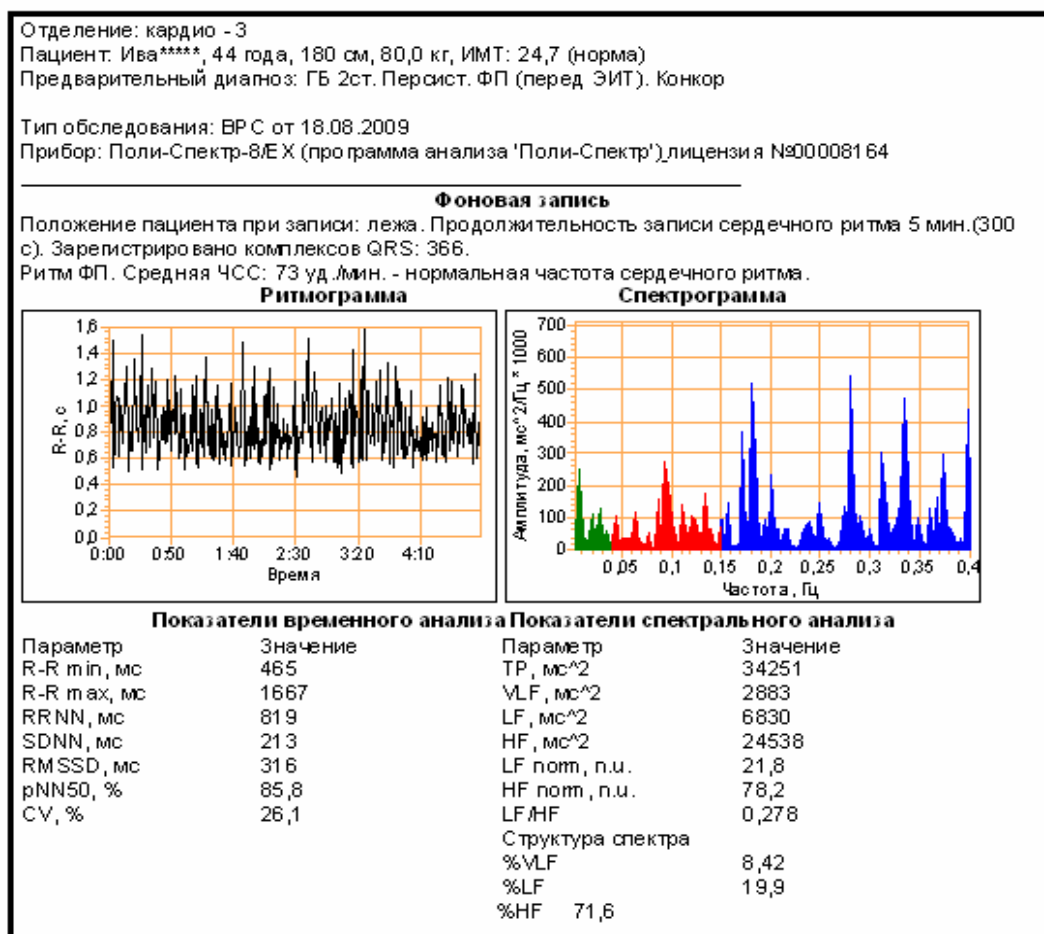


Рис. 5. Фрагмент протокола

Таблица 5. Хранилище прогнозов

ФИО	Спектральный анализ				Временной анализ			Прогноз		
	TP	VLF	LF	HF	RRNN	SDNNf	SDNNfRRNNf	Вет1	Вет2	Решение
Ива***	34251	2883	6830	24538	819	213	0,26		В	Восстановлен
Пет***	11379	1611	2094	7675	704	131	0,186	А		АРИТМ
Афе****	21513	1726	4568	15219	517	149	0,288		А	АРИТМ

**Этап 4.** На основании предварительного анализа ВРС и решения интеллектуальной системы врач делает заключение о вероятности восстановления синусового ритма у конкретного пациента с ФП.

### Заключение

Методом интеллектуального анализа данных (Data Mining) на основе алгоритма деревьев решений построена модель диагностирования формы ФП (пароксизмальная или постоянная) и прогнозирования восстановления сердечного ритма.

Баллод Борис Анатольевич,  
ГОУВПО «Ивановский государственный энергетический университет имени В.И. Ленина»,  
кандидат технических наук, доцент кафедры информационных технологий,  
e-mail: ballod@it.ispu.ru

Муромкина Анна Владимировна,  
ГУЗ Кардиологический диспансер,  
врач, кандидат медицинских наук,  
e-mail: ballod@it.ispu.ru

Ковалев Денис Евгеньевич,  
ГОУВПО «Ивановский государственный энергетический университет имени В.И. Ленина»,  
студент кафедры информационных технологий,  
e-mail: ballod@it.ispu.ru

Модель реализована в виде программного продукта ДИСК\_3 и может использоваться в виде интеллектуального помощника врача-кардиолога.

### Список литературы

1. Баллод Б.А., Елизарова Н.Н. Методы и алгоритмы принятия решений: Учеб. пособие. – М.: Финансы и статистика; ИНФРА-М, 2009.
2. Михайлов В.М. Вариабельность ритма сердца. – Иваново: НейроСофт, 2000.
3. Муромкина А.В. Оценка вариабельности желудочковых сокращений на фоне фибрилляции предсердий // Вестник ИГМА. – 2009. – № 2. – С. 22–24.