

Р.В. Ставицкий, В.А. Солодкий, А.Л. Лебедев, О.Н. Плаутин

## СИСТЕМА КОЛИЧЕСТВЕННОЙ ОЦЕНКИ СОСТОЯНИЯ ЗДОРОВЬЯ ЧЕЛОВЕКА

### Аннотация

В качестве критериев оценки состояния здоровья человека взяты показатели периферической крови, которые вводятся в комплекс «АКС-ЭНОФИТ», основанный на теории распознавания образов и кластерном анализе. «АКС-ЭНОФИТ» состоит из обучающей выборки, собранной для здоровых и больных обоих полов, и решающей программы для 11 систем организма (перечень заболеваний, относящихся к отдельным системам организма, опубликован в [1]-[4]). В результате система позволяет проводить диспансеризацию пациентов и контролировать эффективность их лечения.

Одним из классических методов оценки состояния здоровья пациента в медицине является исследование анализа крови, состоящего из набора значений некоторых показателей. Число этих показателей зависит от типа автоматического анализатора крови и составляет от 15 до 25 и более. Было установлено, что на основании динамической картины изменения одного показателя крови нельзя судить об изменениях во всем организме человека. Лишь совокупность показателей (не менее 5) может отражать изменения гомеостаза [2]. Поэтому в 1991 г. группа исследователей приступила к разработке системы информации, получившей название «Автоматизированная классифицирующая система» (АКС-ЭНОФИТ) [1]-[4]. По анализу крови среднестатистический специалист выявляет информацию о состоянии здоровья пациента, которая «лежит на поверхности», т. е. отражает серьезные изменения в крови, состоящие в выходе за пределы допустимого диапазона значений отдельных показателей (табл. 1).

Для оценки состояния здоровья и эффективности лечения необходимо проведение многопараметрического анализа комплекса показателей, объективно отражающего состояние здоровья. Проводились попытки анализировать комплекс показателей при помощи построения нелинейных регрессий. Однако были получены высокие погрешности и достигнута низкая чувствительность. Альтернативным способом решения задачи оценки состояния здоровья наблюдаемых является представление ее как задачи классификации людей по группам. Можно считать, что наиболее естественным математическим аппаратом для решения такой задачи являются методы распознавания образов (РО) [5]-[7].

Таким образом, задачи классификации можно решить путем распознавания образов, т. е. разбиения объектов исследования на классы с предварительным обучением; множество обучающих объектов эксперты заранее разбивают на классы, причем [8], [9]:

- заданные классы не пересекаются;
- заданные классы пересекаются.

Математический аппарат теории распознавания образов обнаруживает несомненное внутреннее родство с вышеописанными задачами классификации по содержанию, результатам и понятийному аппарату. Возможность реализации методов распознавания образов в виде программ для персональных

компьютеров полностью отвечает поставленной задаче автоматизации биофизических исследований и органично дополняет уже практикуемое в данной проблемной области использование техники (например, автоматических анализаторов крови).

Теоретическая база алгебраического подхода, а также обоснование корректности его практического приложения к задачам, подобным решаемым, разработана в Вычислительном центре РАН под руководством академика Ю.И. Журавлева [5]-[7]. Метод динамического кластерного анализа заключается в выявлении «сгустков» обучающих объектов одного класса (кластеров) в пространстве признаков. Отнесение распознаваемого объекта к одному из классов осуществляется в два этапа:

- 1) сначала с помощью некоторой решающей функции определяются оценки принадлежности объекта ко всем кластерам;
- 2) затем на базе полученных на предыдущем этапе результатов определяются оценки принадлежности объекта к классам, анализ которых приводит к принятию окончательного решения.

Вся обучающая выборка дифференцирована по полу.

Подбор данных о периферической крови включает в себя данные о состоянии 11 систем организма (пищеварительной, дыхательной, опорно-двигательной, сердечно-сосудистой, урологической, гинекологической, эндокринной, ЦНС, грудной железы, печени и желчевыводящих путей и общего состояния всего организма).

Работа по отнесению объекта (пациента) к тому или иному классу строится на следующем:

- предполагается возможность получения формализованных описаний пациентов, а формализованное описание каждого из них представляется вектором фиксированных, упорядоченных и нормированных значений его показателей крови, выступающих в роли признаков;
- предполагается, что все множество объектов исследования можно разделить на конечное число классов и что принадлежность каждого объекта к одному из классов в общем случае заранее неизвестна;
- предполагается наличие обучающей выборки, которая содержит конечное число описаний объектов (эталонов), для каждого из которых принадлежность к классу заранее известна.

Таблица 1

## Средние показатели состава периферической крови здоровых мужчин и женщин России

Пол	RBC, 10 <sup>12</sup> /л	HGB, г/л	RET, 10 <sup>11</sup> /л	PLT, 10 <sup>9</sup> /л	WBC, 10 <sup>9</sup> /л	NEUT, 10 <sup>9</sup> /л	NEUT, %	LIMPH, 10 <sup>9</sup> /л	LIMPH, %	BASO, 10 <sup>9</sup> /л	BASO, %	EOS, 10 <sup>9</sup> /л	EOS, %	P, 10 <sup>9</sup> /л	P, %	S, 10 <sup>9</sup> /л	S, %	MONO, 10 <sup>9</sup> /л	MONO, %
Ж	3,9	130	8,0±1,5	2,14± 2,8	4,46	48,6	2,39	1,34	2,47	0,026	0,32	0,035	0...5,9	0,02	0,24	2,31	48,6	0,3	5,72
	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
М	4,3	148	9±2	1,98± 0,016	7,28	63,4	4,39	2,38	3,89	0,08	1,26	0,3	0...5,9	0,14	2,36	4,31	63,4	0,52	8,62
	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Процент колебаний	7	6	11	1,5	39	21	-	44	-	67	-	88	-	107	-	46	-	42	-
	6,5	1,5	7,5	1,0															

Данные о границах классов, полученные в процессе обучения (центры и радиусы гиперсфер в пространстве признаков), представляют собой результаты обучения и используются на этапе принятия решения.

В настоящей работе остановимся на наиболее сложно формируемом этапе обучения.

Этап обучения предполагает некоторый алгоритмический процесс, имеющий целью «научить» распознающую систему правильно определять класс, к которому принадлежит объект исследования, представленный своим формализованным описанием. Наиболее важным подэтапом на этапе обучения является формирование обучающей выборки на базе принятой системы признаков. Обучающая выборка концентрирует в себе интеллектуальный потенциал формирующих ее экспертов (опытных специалистов в данной предметной области). Последние включают в состав обучающей выборки описания таких объектов исследования, которые, с их точки зрения, являются наиболее типичными представителями каждого из заданных классов. Очевидно, что принадлежность объектов обучающей выборки к классам известна как экспертам, так и распознающей системе. Полученная обучающая выборка методами динамического кластерного анализа разбивается на кластеры для того, чтобы обозначить границы классов. Важным свойством обучающей выборки при использовании алгоритмов алгебраического подхода является ее малый объем, что обуславливает высокую скорость работы обучающей системы. Практика показала, что малого числа эталонов достаточно для высокоэффективной классификации.

Крупнейшие гематологи (И.А. Кассирский, А.И. Воробьев, Ш. Бергану, I. Vinatier, Y. Hayashi и др.) отмечают, что любое заболевание организма и его систем отражается в виде изменений показателей крови. Эти изменения трудно выявляются, если рассматривать отдельные показатели крови. Они могут быть установлены при комплексном анализе показателей математическими методами.

Априорно изменения в результате заболеваний каждой из систем разбиты на 4 класса, имеющие количественные выражения в процентах:

- здоровые – 0...20 %;
- начальное отклонение состояния здоровья – 21...40 %;
- выраженное отклонение состояния здоровья – 41...70 %;
- тяжелые заболевания – 71...100 %.

Применение автоматизированного анализатора крови дает большое число показателей (анализируется 1000...3000 клеток крови) и обеспечивает высокую точность получаемых результатов при правильной наладке и контроле анализаторов. Современные анализаторы крови позволяют дополнительно получить ряд расчетных показателей.

Далее будет показано, что ряд показателей у здоровых людей стабилен, в то время как другие показатели самопроизвольно меняются в разных пре-

делах. В связи с этим вторая группа показателей может дать информацию о состоянии организма в целом и его систем только в результате комплексного аналитического анализа. Анализ изменения состояния организма и его систем по показателям периферической крови возможен только сопоставительным путем, когда в качестве эталона (опорных данных) берутся показатели периферической крови здоровых людей (табл. 1).

Следует отметить, что первые пять показателей (табл. 1) оказались практически стабильными. Их расхождение для женщин не превышает 7...15 %, а для мужчин 1...6,5 %. Нестабильность остальных показателей велика (от 21 до 107 %). Это означает, что при выборе аналитической модели, служащей для оценки состояния здоровья, желательно использовать набор показателей, отличающихся относительной стабильностью. Кроме того, что чрезвычайно важно, количественная оценка состояния здоровья должна базироваться на таком диапазоне показателей, который включает в себя «норму» (здоровые) и обеспечивает диапазон изменений используемых показателей у здоровых пациентов. Например, при ограничении нормы интервалом 20...40 % возможно использование следующих показателей: RBC, HGB, RET, PLT, WBC, NEUT, MONO, LIMP. Соответственно необходим подбор расчетных данных. Такие данные были получены путем обработки 1288 образцов крови клинически здоровых мужчин и женщин в возрасте от 18 до 48 лет.

Аналитическим путем, применяя теорию распознавания образов, устанавливается мера сходства между контрольными и исследуемыми объектами. Это позволяет провести кластеризацию исследуемых объектов.

В процессе классификации [8], [9] производится вычисление сходства между контрольными и исследуемыми параметрами. При этом сходство события представляется точками координатного пространства. Меры сходства имеют четыре стандартных критерия:

- 1) симметрия: для двух объектов  $x$  и  $y$  удовлетворено условие расстояния между ними  $d(x, y) = d(y, x) \geq 0$ ;
- 2) неравенство треугольника: для трех объектов  $x, y, z$  расстояние между ними удовлетворяет условию  $d(x, y) \leq d(x, z) + d(y, z)$ ;
- 3) различимость нетождественных объектов: для двух объектов, если  $d(x, y) < 0$ , то  $x < y$ ;
- 4) неразличимость идентичных объектов: для двух идентичных объектов  $x$  и  $x'$ :  $d(x, x') = 0$ , т. е. расстояние между объектами  $x$  и  $y$  равно нулю, т. е. если  $d(x, x') = 0$ , то объекты  $x$  и  $x'$  идентичны.

Для установления мер сходства используется четыре вида коэффициентов сходства: коэффициент корреляции, мера расстояния, коэффициенты ассоциативности и вероятностные коэффициенты сходства.

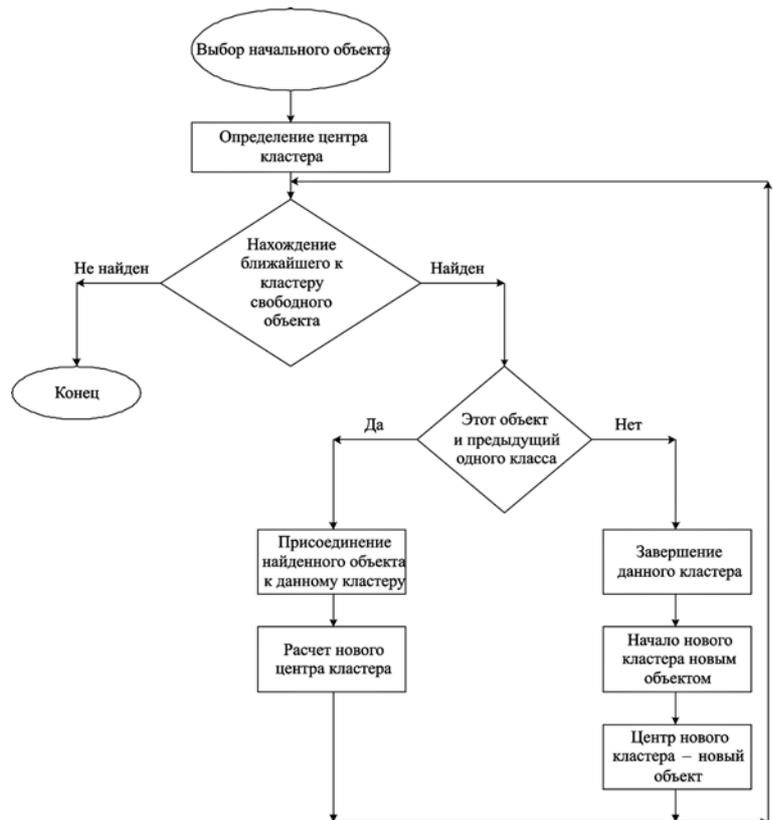


Рис. 1. Блок-схема построения кластеров

Нами использован метод сходства по мерам расстояния. При этом составляющие контрольной (обучающей) выборки распределения кластеров (рис. 1) упорядочиваются по возрастанию евклидова расстояния от начального объекта.

В программе для обучающей выборки использован метод ближайшего соседа, т. е. из состава свободных объектов обучающей выборки формируется начальный объект, с которого начинается процедура разбиения обучающей выборки на кластеры. Все свободные объекты упорядочиваются по возрастанию евклидовых расстояний от начала объекта. Таким образом, динамический кластерный анализ, являясь первым этапом, служит для кластеризации (упорядочивания) параметров обучающей выборки, по сравнению с которой производится установление вида и степени заболевания. Врачами-консультантами выбираются наиболее характерные заболевания всех систем, которые подразделяются на 4 класса.

Всего в обучающей выборке зарегистрировано 118 заболеваний. Для каждого вида заболевания подобрано 20...50 пациентов обоего пола с четко верифицированным диагнозом. Эта верификация проводилась группой из 13 врачей-специалистов в соответствующих областях. Каждый объект (пациент) обучающей выборки вводился в программный комплекс математиками в соответствии с заранее установленными для него системой и классом. Все формализованные описания пациентов, как было показано выше, представлялись фиксированным и

упорядоченным множеством значений показателей крови, полученных в результате лабораторного анализа. Наиболее важными исходными данными для этапа обучения являются объекты обучающей выборки, каждый из которых представлен своим формализованным описанием – указанием принадлежности к одному из классов.

Обучающая выборка формируется экспертами (врачами) и представляется в форме числовой таблицы  $T$ , состоящей из  $M$  строк и  $N$  столбцов. Описание объектов, вошедших в состав обучающей выборки, представляется строками  $S_i$  таблицы  $T$ . Столбцы  $T$  соответствуют признакам  $X_k$ . Строки  $T_{MN}$  разделены на группы, соответствующие данным непересекающихся классов. Обучающая выборка  $T_{MN}$  должна соответствовать следующим основным требованиям:

- число столбцов  $T$  (признаков) теоретически не может быть меньше двух. В противном случае задача теряет смысл, ограничений на максимальное число столбцов нет;
- число строк  $T$  (обучающих объектов) не должно быть менее 10...20 на каждый класс. В противном случае получить эффективное решение затруднительно, ограничений на максимальное число строк практически нет. Особо следует отметить тот случай, когда решается задача собственно классификации, в которой число строк в  $T_{MN}$  равно числу классов (класс представляется одним объектом); в программе АКС в каждом классе используется 40...50 строк для элементов каждого поля, т. е. 80...100 строк для каждого класса;
- число классов, на которые разделены строки  $T_{MN}$ , теоретически не может быть меньше двух; максимальное число классов не может превышать число строк в  $T_{MN}$ ;
- таблица  $T_{MN}$  не может содержать ни одной пары неразличимых строк; если такая пара встречается внутри одного класса, то одна из ее строк – лишняя и не несет новой информации; если строки такой пары встречаются в разных классах, то для одной из строк утверждение о принадлежности к данному классу неверно;
- таблица  $T_{MN}$  не должна содержать ни одного сквозного столбца, т. е. такого столбца, значение которого на всех строках  $T_{MN}$  неразличимы; очевидно, что такой столбец совершенно безболезненно может быть исключен из  $T_{MN}$ ;
- в таблице  $T_{MN}$  не должно быть ни одной пары неразличимых столбцов. Число строк выборки не ограничивается.

Применяя указанную систему к решению задач по оценке состояния организма в целом и его систем на базе набора показателей периферической крови, можно констатировать: наличие обучающей выборки с набором показателей периферической крови, характеризующим каждый из классов всех систем, позволяет по изменению совокупности показателей установить изменение состояния всего организма и его систем. Здесь, по-видимому, играет роль аспект неравномерности изменений разных показателей периферической крови при различных

заболеваниях разной тяжести отдельных систем организма. Степень этих изменений (расстояние от центра гиперсферы (рис. 2-3)) характеризуется соответствующим классом заболевания. Исходя из того, что покрытие осуществляется системой гиперсфер, предполагается:

- возможность получения формализованных описаний пациентов. Формализованное описание представляется вектором фиксированных упорядоченных и нормированных значений его показателей крови, выступающих в роли признаков;
- что множество объектов исследования можно разделить на конечное число классов и что принадлежность каждого объекта к одному из классов в общем случае заранее неизвестна;
- наличие обучающей выборки, которая содержит конечное число описаний объектов (эталонов), для каждого из которых принадлежность к классу заранее известна.

Данные о границах классов, полученные в процессе обучения (центры и радиусы гиперсфер в пространстве признаков), представляют собой результаты обучения и используются на этапе принятия решения.

В результате тщательной подготовки обучающей выборки с помощью математического аппарата возможна оценка состояния организма в целом и его систем по показателям периферической крови.

В качестве примеров приведены паспорта здоровья здорового (рис. 4) и больного (рис. 5) пациентов.

В итоге можно отметить, что использование программного комплекса «АКС-ЭНОФИТ» в лечебных учреждениях дает возможность не только оценить состояние всего организма и отдельных его систем, но и осуществлять динамический контроль эффективности проводимых лечебных мероприятий.

#### Список литературы:

1. Количественные критерии оценки эффективности лечения рака молочной железы / Под. ред. Р.В. Ставицкого, Г.А. Паньшина.
2. Медицинская рентгенология: технические аспекты, клинические материалы, радиационная безопасность / Под. ред. Р.В. Ставицкого. – М.: МНПИ, 2004. С. 48-78.
3. Периферический рак легкого: количественная оценка эффективности радикального химио-лучевого лечения / Под. ред. Р.В. Ставицкого, Г.А. Паньшина. – М.: Гарт, 2008. С. 218.
4. Ставицкий Р.В., Гуслистый В.П. и др. // Международ. мед. журнал. 2001. № 1. С. 55-61.
5. Журавлев Ю.И. // Проблемы кибернетики. 1978. Вып. 33.
6. Журавлев Ю.И., Зенкин А.А., Зенкин А.И. // ЖВМиМФ. 1980. Т. 20. № 5.
7. Журавлев Ю.И., Исаев И.В. // ЖВМиМФ. 1970. Т. 19. № 3.
8. Исаев И.В. // ЖВМиМФ. 1983. Т. 23. № 2.
9. Исаев И.В. // ЖВМиМФ. 1984. Т. 24. № 7.

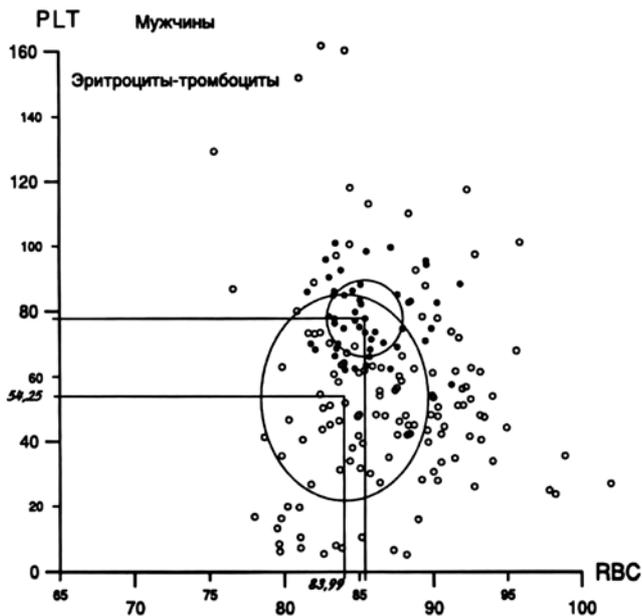


Рис. 2. Проекция распознавания объектов «норма» и «лимфогранулематоз» в пространстве PLT (тромбоциты) и RBC (эритроциты) (мужчины)

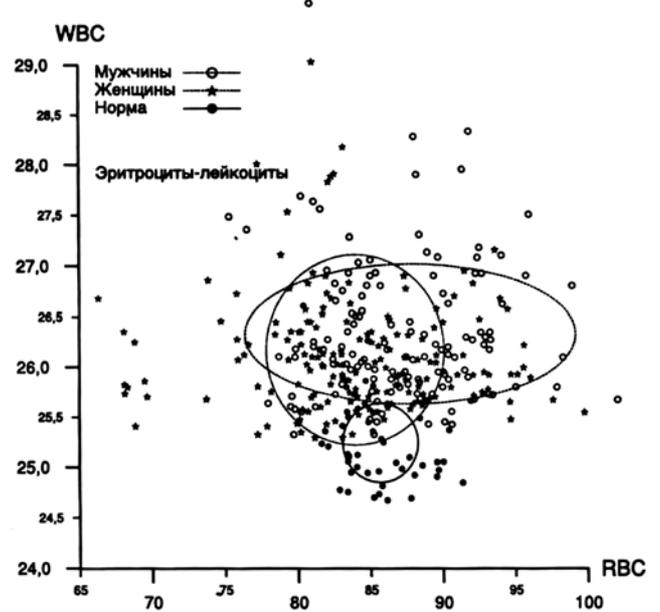


Рис. 3. Распознавание образов «норма» и «лимфогранулематоз» в пространстве WBC (лейкоциты) и RBC (эритроциты) (мужчины и женщины)

### Паспорт здоровья

Дата: 02.03.04

Ф.И.О.:			№ карты:	10
Дата рождения:	1975	Пол:	ж	№ полиса ОМС
Домашний адрес			Тел.:	
Место работы:	«РЖД», служащая		Тел.:	

Оценка состояния здоровья пациента, полученная в результате обработки показателей периферической крови посредством «АКС-ЭНОФИТ»

№ п/п	Наименование системы организма	Степень повреждения, %	Заключение
0	Весь организм	14,0	Здоров
1	Пищеварительная система	12,7	Здоров
2	Органы дыхания	11,5	Здоров
3	Опорно-двигательный аппарат	10,6	Здоров
4	Сердечно-сосудистая система	13,9	Здоров
5	Урологическая система	12,2	Здоров
6	Гинекологическая система	18,3	Здоров
7	ЦНС и органы чувствительности	10,9	Здоров
8	Грудная железа	10,0	Здоров
9	Печень и желчевыводящие пути	13,8	Здоров

Примечание: состояние организма и его систем производится в процентном отношении:

0 – 20 % здоров

21 – 40 % начальное отклонение здоровья

41 – 70 % выраженное отклонение здоровья

71 – 100 % наличие тяжелого заболевания

### Результат анализа крови

WBC $\times 10^9$	RBC $\times 10^{12}$	HGB, г/л	PLT $\times 10^9$	MCV, Мк	MCH, Пг	MCHC, г/л	RDW, %	MPV, Мк	GRA, %	LYM, %	MON, %	PDW, %
5,7	4,5	138	306	85	30,9	363	10,7	7,3	60,2	29,1	10,7	14,3

Рис. 4. Паспорт здоровья пациентки К.

## Паспорт здоровья

Дата: 02.03.04

Ф.И.О.:				№ карты:	38
Дата рождения:	1956	Пол:	ж	№ полиса ОМС	
Домашний адрес				Тел.:	
Место работы:	«РЖД», секретарь			Тел.:	

*Оценка состояния здоровья пациента, полученная в результате обработки показателей периферической крови посредством «АКС-ЭНОФИТ»*

№ п/п	Наименование системы организма	Степень повреждения, %	Заключение
0	Весь организм	47,8	Требуется клиническая диагностика
1	Пищеварительная система	25,1	Требуется клиническая диагностика
2	Органы дыхания	42,9	Требуется клиническая диагностика
3	Опорно-двигательный аппарат	13,0	Здоров
4	Сердечно-сосудистая система	24,4	Требуется клиническая диагностика
5	Урологическая система	17,1	Здоров
6	Эндокринная система	10,1	Здоров
7	Гинекологическая система	19,5	Здоров
8	ЦНС и органы чувствительности	12,9	Здоров
9	Грудная железа	13,5	Здоров
10	Печень и желчевыводящие пути	13,5	Здоров

Примечание: состояние организма и его систем производится в процентном отношении:

0 – 20 % здоров

21 – 40 % начальное отклонение здоровья

41 – 70 % выраженное отклонение здоровья

71 – 100 % наличие тяжелого заболевания

## Результат анализа крови

WBC x10 <sup>9</sup>	RBC x10 <sup>12</sup>	HGB, г/л	PLT x10 <sup>9</sup>	MCV, Мк	MCH, Пг	MCHC, г/л	RDW, %	MPV, Мк	GRA, %	LYM, %	MON, %	PDW, %
11,1	4	135	267	96	33,5	350	12	8,7	76,9	16,8	6,3	14,6

Рис. 5. Паспорт здоровья пациентки Г.

*Роман Владимирович Ставицкий,*

*д-р биол. наук, профессор,*

*зав. лабораторией,*

*Владимир Алексеевич Солодкий,*

*д-р мед. наук, директор, член-корреспондент РАМН,*

*Александр Ларионович Лебедев,*

*канд. техн. наук, научный сотрудник,*

*лаборатория дозиметрических исследований,*

*ФГУ «РНЦРР»,*

*руководитель отдела новых технологий,*

*АНО «ГНТЦЭПТ»,*

*Олег Николаевич Плаутин,*

*инженер по электронике,*

*руководитель группы,*

*лаборатория дозиметрических исследований,*

*ФГУ «РНЦРР»,*

*г. Москва,*

*e-mail: Plautin1966@yandex.ru*