

ОПРЕДЕЛЕНИЕ ИНФОРМАТИВНОСТИ ДИАГНОСТИЧЕСКИХ ПАРАМЕТРОВ ДЛЯ ДИФФЕРЕНЦИАЦИИ РАСПРОСТРАНЕННЫХ ДЕРМАТОЗОВ

А.И. ПЕЧЕРСКАЯ

Рассмотрены различные математические подходы к определению информативных признаков. Решены задачи структурной и параметрической идентификации функции важности диагностического показателя. В качестве значимых факторов, влияющих на выходные данные модели, выделены вес и выраженность симптома. Для идентификации вида оператора, устанавливающего связь между входными и выходными данными, выбрана аддитивная форма. Для решения задачи определения информативных показателей выбран информационный критерий, обладающий энтропийными свойствами.

This article presents the various mathematical models for diagnostic character classification. Diagnosis parameter identification system was elaborated. Constitutional guiding symptom regarded to be master factor for finite-extent data processing. Weighting coefficient was computed. The operator format was selected for input-output unit. Side mode was implemented for operator format identification. the test data were analyzed by maximum entropy processing.

ВВЕДЕНИЕ

В настоящее время одними из наиболее распространённых заболеваний кожи являются хронически рецидивирующие патологии кожи [1].

Точность диагноза и быстрота, с которой его можно поставить, зависят от очень многих факторов: от состояния больного, от имеющихся данных о симптомах и признаках заболевания и результатах лабораторных анализов, от общего объема медицинской информации о наблюдении таких симптомов и, наконец, от квалификации самого врача. Своевременно поставленный точный диагноз часто облегчает выбор метода лечения и значительно повышает вероятность выздоровления больного [2].

Дифференциация многих распространенных дерматозов, таких как псориаз, экзема, лекарственная болезнь, весьма сложна, так как этим разным по этиологии и патогенезу патологиям свойственны идентичные клинические проявления. Следовательно, проблема дифференциальной диагностики дерматологических заболеваний является одной из актуальнейших проблем медицины на современном этапе, решение которой позволит сократить срок преддиагностического периода и повысить качество постановки дифференцированного диагноза.

Дифференциация дерматопатологий проводится на основе информации, полученной в результате осмотра пациента, сбора анамнестических данных и проведения лабораторных исследований [3].

В результате лабораторных исследований получают большое количество показателей, многие из которых являются иррелевантными, то есть не относятся к цели исследования [4–5]. Таким образом, при дифференциальной диагностике дерматологических заболеваний одним из ключевых вопросов является получение набора информативных признаков минимального размера,

достаточно полно описывающих каждую из рассматриваемых нозологических единиц.

1. АНАЛИЗ ЛИТЕРАТУРЫ

Исследование информативности диагностических признаков производится путем решения ряда задач статистической классификации на диагностических выборках, включающих различные сочетания признаков, с последующим определением ценности каждого из них. Если задача исследователя не ограничивается только констатацией факта отличия средних, а ставится проблема поиска существующих связей, то на первом, разведочном этапе, применяется исследование коэффициентов корреляций. При этом при анализе непрерывных переменных, используется коэффициент корреляции Пирсона, а для порядковых или категориальных – коэффициент корреляции Спирмена или ранговые корреляции. Следует иметь в виду, что при большом анализируемом количестве параметров, чисто случайно, ряд показателей могут дать достоверные коэффициенты корреляции между собой (например, если анализируется 10 показателей, то в корреляционной матрице 10×10 - 3-4 достоверных показателя могут быть случайностью). Поэтому к полученным коэффициентам следует относиться с осторожностью. С другой стороны, отсутствие достоверной корреляции между исследуемыми показателями может еще не означать отсутствия связи между параметрами. В качестве критерия эффективности разбиения может использоваться максимум функции вероятности правильного распознавания объектов экзаменационной выборки или минимум функции среднего риска [6].

Для выявления возможной связи между параметрами применяется также и дисперсионный анализ. Дисперсионный анализ позволяет определить влияние разных факторов (условий) на исследуемый признак (явление), что достигается путем

разложения совокупной изменчивости (дисперсии, выраженной в сумме квадратов отклонений от общего среднего) на отдельные компоненты, вызванные влиянием различных источников изменчивости. С помощью дисперсионного анализа исследуются угрозы заболевания при наличии факторов риска. Концепция относительного риска рассматривает отношение между пациентами с определенной болезнью и не имеющими ее.

Основной принцип факторного анализа и многомерного шкалирования (метод главных компонент) – снижение размерности выборки (количество переменных задается исследователем и может быть сокращено до 2) на основании создания линейной комбинации всех анализируемых переменных. При этом, ставится задача максимально сохранить информацию, входящую в весь набор данных, для чего вводится характеристика – уровень объясненной дисперсии. Полученные факторы включают информацию, содержащуюся в исходных переменных, при этом количество переменных снижается. Обычно целью факторного анализа является попытка качественно описать большую совокупность имеющихся параметров. Как бы «суммировать» информацию и получить факторы, имеющие некий биологический смысл.

Математическим аппаратом решения задач исследования информативности с момента их возникновения явились комбинаторные методы перебора различных сочетаний признаков (например, методы полного и сокращенного перебора). Однако их недостатками являются либо высокая вычислительная стоимость, либо недостаточная точность [7].

При сокращении размерности пространства диагностических признаков первостепенной задачей является упорядочивание критериев по важности. Важность диагностического показателя – это его способность отражать структурные особенности той или иной дерматопатологии, то есть показатель тем важнее, чем лучше он характеризует изменение дерматологического статуса пациента [8].

Наиболее простым является метод лексикографического упорядочивания критериев. При этом наиболее вероятным является тот диагноз, который имеет более высокую оценку по наиболее важному критерию, что далеко не всегда соответствует действительности. Также этот метод не позволяет учитывать равнозначность некоторых критериев [9].

При выборе существенного для дифференциальной диагностики распространенных дерматозов подмножества признаков одним из методов, которые целесообразно использовать, является метод дискриминантных функций, который учитывает вариабельность, рассматривает совокупность всех признаков, которые берутся во внимание, взятых со своими коэффициентами, которые указывают диагностический вес признака при формировании дерматопатологии [10].

2. ЦЕЛЬ ИССЛЕДОВАНИЯ. ПОСТАНОВКА ЗАДАЧИ

Целью исследования является разработка методики определения информативных диагностических параметров для дифференциации распространенных дерматозов. Это позволит облегчить задачу практического врача, повысить качество постановки дифференцированного диагноза и сократить срок преддиагностического периода.

Анализ литературы [6-10] показывает, что наиболее достоверной является количественная оценка предпочтительности диагностического показателя, в качестве которой выступает обобщенная важность. Чтобы найти эту величину, необходимо обосновать решающее правило, по которому формируется важность показателя в пространстве частных критериев $f_i(x_i)$, то есть решить задачи структурной и параметрической идентификации функции важности:

$$U(x) = G[f_i(x)], \quad (1)$$

где G – оператор, определяющий вид зависимости; $f_i(x)$ – частные критерии, характеризующие выраженность того или иного диагностического показателя у пациента; $i = \overline{1, n}$; n – количество диагностических показателей [11].

3. РЕШЕНИЕ ЗАДАЧИ СТРУКТУРНОЙ ИДЕНТИФИКАЦИИ

Структурная идентификация любой математической модели, в том числе и функции важности, предполагает необходимость решения двух взаимосвязанных задач:

- выделение значимых факторов, влияющих на выходные данные модели;
- определение структуры, то есть вида оператора, устанавливающего связь между входными и выходными данными модели.

Решение задачи структурной идентификации модели связано с выдвижением и проверкой некоторой гипотезы. В случае дифференциации заболеваний кожи вид функции обобщенной важности диагностического показателя $x \in X$ определяется значениями частных критериев $f_i(x)$, $i = \overline{1, n}$, характеризующих решение, и в общем случае эти характеристики не равнозначны, то есть имеют различный вес для эксперта-медика. Это не исключает того, что в частном случае вес некоторых критериев может быть одинаковым.

Особенностью задачи дифференциации дерматопатологий также является то, что признаки, характеризующие состояние пациента, имеют разную природу (различные единицы, интервалы и шкалы измерения), то есть показатели не сравнимы между собой. Следовательно, необходимо либо привести их к единой форме, либо использовать информационный критерий, не зависящий от единиц измерения признаков. Тогда формула (1) может быть записана в виде:

$$U(x) = G[k_i, f_i(x)], \quad (2)$$

где $i = \overline{1, n}$; k_i – параметры изоморфизма, то есть параметры взаимно однозначного соответствия, приводящие разнородные частные критерии $f_i(x)$ к единой метрике, и учитывающие их вес.

Параметры изоморфизма задаются непосредственно экспертом-медиком в численном виде методом экспертных оценок, либо определяются статистическими методами.

Следующим этапом является определение структуры зависимости, то есть вида оператора, устанавливающего связь между входными (весом и выраженностью симптома) и выходными данными (важность). Наиболее широко используемыми для идентификации вида оператора являются аддитивная и мультипликативная формы [9].

Так как параметры изоморфизма являются константами, то мультипликативную форму идентификации оператора G

$$U(x) = \prod_{i=1}^n k_i f_i(x) \quad (3)$$

можно представить в виде

$$U(x) = \prod_{i=1}^n k_i \prod_{i=1}^n f_i(x). \quad (4)$$

$\prod_{i=1}^n k_i$ является постоянным масштабным множителем, а, следовательно, все критерии становятся равно важными, что не соответствует истине. Таким образом, мультипликативная форма не позволяет учесть информацию о предпочтительности частных критериев.

Еще одной особенностью мультипликативной формы является недопустимость полной компенсации одних параметров другими. Если хотя бы один из множителей равен 0, вся мультипликативная функция принимает нулевое значение. А в случае дифференциальной диагностики дерматологических заболеваний невыраженность у пациента того или иного диагностического показателя, характерного для данной патологии, не исключает эту патологию из рассмотрения. Это говорит о том, что мультипликативная форма идентификации оператора G неприменима при определении важности диагностических показателей дерматопатологии.

В отличие от мультипликативной, аддитивная форма оператора зависимости важности показателя от его веса и выраженности у пациента

$$U(x) = \sum_{i=1}^n k_i f_i(x) \quad (5)$$

не имеет вышеназванных недостатков.

Однако формула (5) корректна только в том случае, если k_i учитывают не только вес i -го диагностического показателя, но и являются коэффициентами изоморфизма. В общем случае определение таких коэффициентов затруднено. Это обстоятельство можно преодолеть, если представить аддитивную функцию в форме

$$U(x) = \sum_{i=1}^n \delta_i f_i^H(x), \quad (6)$$

где δ_i – безразмерные коэффициенты, характеризующие вес i -го диагностического показателя для конкретной патологии, для которых выполняются ограничения:

$$0 \leq \delta_i \leq 1, \quad (7)$$

$$\sum_{i=1}^n \delta_i = 1; \quad (8)$$

$f_i^H(x)$ – нормализованные, то есть приведенные к изоморфному виду частные критерии [12].

В случае дерматологической дифференциальной диагностики в качестве критериев $f_i^H(x)$ выступают оценки выраженности того или иного диагностического показателя у пациента. Изоморфизм частных критериев означает, что они имеют одинаковую размерность и интервал возможных значений.

Так как по определению δ_i – безразмерные коэффициенты, то размерность $f_i^H(x)$ должна совпадать с размерностью $U(x)$, то есть $f_i^H(x)$ характеризует локальную важность i -го диагностического показателя x для конкретной патологии:

$$f_i^H(x) = p_i(x). \quad (9)$$

В соответствии с (9) функция (6) примет вид:

$$U(x) = \sum_{i=1}^n \delta_i p_i(x). \quad (10)$$

Метод определения важности частных критериев, основанный на применении аддитивной формы оператора зависимости получил название метод взвешенной суммы.

Как и для всех прямых методов, для метода взвешенной суммы характерны высокие требования к экспертам, особенно на начальных этапах работы при формировании экспертных оценок.

Модель оценивания (10) справедлива только в том случае, если весовые коэффициенты δ_i частных критериев $f_i^H(x)$ заданы точными количественными значениями. Как уже отмечалось выше, носителями этой информации являются эксперты-медики, и, значит, необходимы некоторые процедуры ее получения, то есть решение параметрической идентификации модели.

Если же по каким-либо причинам получение точной количественной информации не возможно, в общем случае оценивание приходится производить в условиях большей или меньшей степени неопределенности. Это возможно в случае дифференциальной диагностики в группе редких малоизученных заболеваний кожи. В этом случае общая модель определения обобщенной важности показателя $x \in X$ имеет вид:

$$U(x) = G[J(\alpha_i), p_i(x)], \quad (11)$$

где $i = \overline{1, n}$; $J(\alpha_i)$ – информация о взаимной важности частных критериев.

4. РЕШЕНИЕ ЗАДАЧИ ПАРАМЕТРИЧЕСКОЙ ИДЕНТИФИКАЦИИ

Параметрическая идентификация заключается в определении конкретных количественных параметров модели.

Для решения этой задачи применяют принцип ассоциативности, метод связей симптомов и синдромов, метод каузальных связей симптомов с синдромами по смежности во времени и другие [13].

Исследование информативности диагностических признаков также производится путем решения ряда задач статистической классификации на диагностических выборках, включающих различные сочетания признаков, с последующим определением ценности каждого из них.

Так как k_i – параметры изоморфизма, приводящие разнородные частные критерии $f_i(x)$ к единой метрике, и учитывающие их вес, следовательно, для их идентификации необходимо использовать информационный критерий, не зависящий от единиц измерения признаков.

Данным требованиям соответствует критерий информативности, обладающий энтропийными свойствами, который не зависит от единиц измерения признака и учитывает только относительные значения распределения признаков на элементах классов распознавания [14].

Для определения информативности диагностических показателей была исследована выборка из 330 пациентов, страдающих дерматозами, которым был поставлен корректный дифференцированный диагноз и проведено эффективное лечение. Состояние каждого пациента описывалось 94 лабораторными показателями (показатели гемограммы, иммунологического, аллергологического, биохимического и общеклинического исследований).

Все признаки были закодированы и поставлены в соответствие 94-мерному вектору, учитывающему присутствие, отсутствие и величину каждого признака.

Пациенты были распределены на 3 класса: псориаз, экзема, лекарственная болезнь.

С математической точки зрения все пациенты рассматриваются как совокупность объектов с переменными количественными и качественными характеристиками.

Пусть $\%$ матрица с положительными элементами $X \in R_+^{n \times m} = \{X = (x_{ij}) \mid x_{ij} > 0; i = \overline{1, n}; j = \overline{1, m}\}$, имеющая n строк ($n = 330$) и m столбцов ($m = 94$). Элементами матрицы являются значения признаков дерматологического статуса каждого пациента. И пусть k – число распознаваемых классов ($k = 3$), $\xi(x_{ij})$ – номер класса, которому соответствует значение E_{ij} j -го признака у i -го пациента и пусть

$p \in \Delta_n = \left\{ p = (p_1, p_2, \dots, p_n)^T \mid \sum_{i=1}^n p_i, p_i \geq 0, \forall i = \overline{1, n} \right\}$. Тогда информативность j -го признака (столбца X_j

матрицы $\%$) можно оценить на основе решения задачи формирования информационного критерия:

$$I(x) = h(x) - h(x/\Omega) = \ln(p^T x \cdot x^{-T} p) - \min_{w \in \Omega} \ln((p/w)^{-T} x \cdot x^{-T} (p \cdot w)),$$

где $h(x) = \ln(x^{-T} p \cdot p^T x) = \ln\left(\sum_{i=1}^n \frac{p_i}{x_i} \cdot \sum_{i=1}^n p_i x_i\right)$ с областью определения Ω в виде:

$$\Omega = \left\{ w \in R_+^{m \times 1} \mid w_u = w_v \Leftrightarrow \xi(x_{uj}) = \xi(x_{vj}) \forall j = \overline{1, m} \right\}.$$

Информативность признака тем выше, чем ближе значение информационного критерия к единице [15 – 16].

В результате из начальных 94 параметров с помощью определения информационного критерия удалось выявить 14 информативных (рис. 1), которые адекватно отражают процессы, проходящие при развитии патологических процессов в коже.

Таким образом, такие распространенные дерматозы, как псориаз, экзема, лекарственная болезнь, можно дифференцировать при помощи следующих лабораторных показателей: содержание в сыворотке крови Т-лимфоцитов (СД3), Т-лимфоцитов хелперов (СД4), Т-цитотоксических клеток (СД8); фактора апоптоза (FASx); полипептида эндотелеина-1; витамина Е, интерлейкина 6 (ИЛ 6), интерлейкина 8 (ИЛ 8), интерлейкина 10 (ИЛ 10), фактора некроза опухоли, молекул средней массы (МСМ), регистрируемых при длине волны 254 нм, титры IgA-антител к трансглутаминазе и глиадину; содержание иммуноглобулина А (IgA); G (Ig G) и общего иммуноглобулина Е (Ig E).

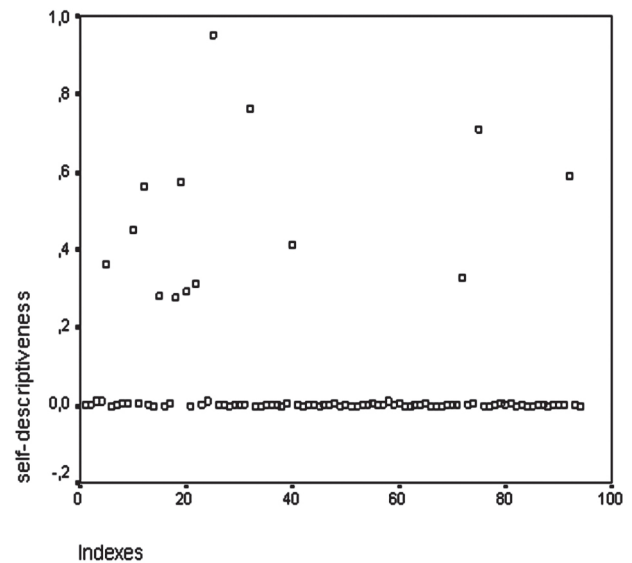


Рис. 1. Диаграмма распределения информативности диагностических показателей

Расчет значений диагностических коэффициентов методом дискриминантных функций подтвердил правильность выбранных 14 существ-

венных для проведения дифференциальной диагностики дерматозов показателей. В 94-мерном пространстве при диагностике 3-х нозологических единиц получили 4 области: точки, свойственные только псориазу, только экземе, только лекарственной болезни и переходная зона.

Оценка степени удачности распределения на группы, полезность дискриминантных функций и количество функций, имеющих реальное содержание при определении отличий между группами, могут быть оценены с помощью коэффициентов канонической корреляции (табл. 1).

Таблица 1

Оценка степени удачности распределения на группы

Функция	Собственное значение	% дисперсии	Совокупный %	Каноническая корреляция
1	11,312	91,3	91,3	0,961
2	2,106	8,7	100,0	0,816

Судя по значениям коэффициентов, равных 0,954 и 0,718 и $P < 0,000$ и $P < 0,000$ расхождения между группами значимое.

Оценка значимости дискриминантных функций также была проверена λ -статистикой Уилкса (табл. 2).

Таблица 2

Оценка значимости дискриминантных функций

Номер функции	Лямбда Уилкса	χ^2	Степени свободы, df	Значимость
1	0,044	418,365	28	0,000
2	0,485	96,719	13	0,000

Полученные собственные значения функций, равные 11,312 и 2,106, указывают на то, что дискриминантные функции подобраны удачно.

Анализ результатов классификации (табл. 3) показал, что прогноз для групп 1 (псориаз), 2 (экзема) и 3 (лекарственная болезнь) сделан верно, и корректно классифицированы, в общей сложности, 96,4% всех больных дерматозами.

Таблица 3

Классификационные результаты

	Диагноз	Прогнозируемая принадлежность к группе			Всего	
		Группа 1	Группа 2	Группа 3		
Первоначально	Количество	Группа 1	148	2	0	150
		Группа 2	8	92	0	100
		Группа 3	0	3	77	80
	%	Группа 1	97,0	3,0	0,0	100,0
		Группа 2	8,0	92,0	0,0	100,0
		Группа 3	0,0	3,8	96,2	100,0

Соответственно, точность дифференциальной диагностики распространенных дерматозов для обследуемых пациентов может достигать 96,4 % в том случае, если выборка пациентов будет идентична тем пациентам, данные которых послужили основой для прогноза.

ВЫВОДЫ

Решены задачи структурной и параметрической идентификации функции важности, которая является количественной оценкой предпочтительности диагностического показателя.

В качестве значимых факторов, влияющих на выходные данные модели, выделены вес и выраженность симптома.

Для идентификации вида оператора, устанавливающего связь между входными и выходными данными, выбрана аддитивная и форма.

Для решения задачи определения информативных показателей оптимальным представляется использовать информационный критерий, обладающий энтропийными свойствами. Целесообразность применения этого критерия обусловлена его эффективностью и нечувствительностью к единицам измерения признаков.

Выделены 14 информативных признаков, которые позволяют разграничивать рассматриваемые нозологические единицы и адекватно отражают процессы, проходящие при развитии патологических процессов в коже. Фактически получено новое медико-биологическое знание из эмпирических данных, что позволит облегчить задачу врача-диагноста, повысить качество дифференцированного диагноза и сократить срок преддиагностического периода.

Литература.

- [1] О.Л. Ананьев, Е. В. Анисимова, Н.В. Иванчикова и др. Кожно-венерические заболевания. Полный справочник. – М.: Эксмо, 2005. – 608 с.
- [2] Н. Бейли. Математика в биологии и медицине. – М.: Мир, 1970.
- [3] Дифференциальная диагностика кожных болезней./ Под ред. Б.А. Беренбейна. – М.: Медицина, 1989. – 672 с.
- [4] Солошенко Э.Н., Высоцкая Е.В., Тихонова А.И. Критерии дифференциальной лабораторной диагностики распространенных дерматозов // Материали науково-практичної конференції «Сучасний менеджмент в дерматовенерології: діагностичні, лікувальні та організаційно-правові аспекти». 2007. – С. 129-130.
- [5] Солошенко Э.Н., Высоцкая Е.В., Тихонова А.И., Порван А.П., Стулий О.Н., Жукова Н.В. Критерии лабораторной дифференциальной диагностики лекарственной болезни и экземы // Имунология та алергологія. 2007. №3 – С. 85-86.
- [6] Павленко В.Д., Фомин А.А. Критерии отбора информативных совокупностей признаков при многоклассовом распознавании. // Труды ОПУ, В.3, 2000. – С. 146-150.
- [7] Медик В.А., Фишман Б.Б., Токмачев М.С. Руководство по статистике в медицине и биологии. – М.: Медицина, 2001. – 352 с.
- [8] А. Bykh, E. Visotska, O. Kozina, A. Tikhonova, A. Porvan, A. Zhook. Application of method of the weighed total for diagnostic index significance calculation in differential

- diagnostics of dermatological diseases // Proceedings of the Fifth International Conference «Information research and applications». Volume 2. 2007. – P. 307-308.
- [9] Э.Г. Петров, М.В. Новожилова, И.В. Гребенник, Н.А. Соколова. Методы и средства принятия решений в социально-экономических и технических системах. – Херсон: ОЛДИ-плюс, 2003. – 380 с.
- [10] Armitage P. Theory and practice in medical statistics / Statistics in Medicine. – 2001. – v. 20. – Issue 17-18. – P. 2537-2548.
- [11] A.Bykh, E.Visotska, O.Kozina, A.Tikhonova, A.Porvan, A.Zhook. Ground of model for the generalized criterion forming at differential diagnostics of dermatological diseases // International Journal «Information Theories & Applications». 2008, Vol. 15. Num. 3 – P. 296-299.
- [12] О.И. Ларичев, Е.М. Мошкович, Ю.П. Иванюков. Количественные методы принятия решений. – М.: Физмалит, 1996. – 208 с.
- [13] Поворозник А.И., Высоцкая Е.В., Козина О.А., Тихонова А.И., Жемчужкина Т.В. Анализ и учет знаний экспертов-медиков в медицинских информационных системах // Технічна електродинаміка. Тематичний випуск «Силова електроніка та енергоефективність». – Ч. 1, 2007. – С.116-119.
- [14] Куренков Н.И., Ананьев С.Н. Информационный критерий и его использование для решения задач обработки многомерных данных. // Информационные технологии. 2006. № 8. – С. 50-55.
- [15] Печерская А.И. Определение информативных показателей при дифференциальной диагностике дерматологических заболеваний // Сборник научных трудов 3-го Международного радиоэлектронного форума «Прикладная радиоэлектроника. Состояние и перспективы развития» МРФ – 2008. Том IV. Конференция «Актуальные проблемы биомедицины». 2008. – С. 260-262.
- [16] Солошенко Э.Н., Высоцкая Е.В., Печерская А.И., Жукова Н.В. Оценка информативности лабораторных показателей при дифференциации клинических вариантов течения и стадий псориаза // Журнал дерматовенерологии и косметологии им. Н.А.Горсуева. 2008, № 1–2 (16). – С. 55-60.

Поступила в редколлегию 15.10.2008



Печерская Анна Ивановна, аспирант кафедры биомедицинских электронных устройств и систем ХНУРЭ. Область научных интересов: математические методы и модели дифференциальной диагностики в медицине.