

DOI:

ОЦЕНКА АПОСТЕРИОРНЫХ ВЕРОЯТНОСТЕЙ КЛАССОВ В ЗАДАЧЕ ДИФФЕРЕНЦИАЛЬНОЙ ДИАГНОСТИКИ ЗАБОЛЕВАНИЙ СЛИЗИСТОЙ ОБОЛОЧКИ ПОЛОСТИ РТА

Зенков В.В.

*Институт проблем управления им. В.А. Трапезникова РАН, Россия, г.Москва
ул.Профсоюзная д.65
zenkov-v@yandex.ru*

Молочкова Ю.В., Молочков В.А., Хлебникова А.Н., Глазков А.А.

*ГБУЗ «Московский областной научно-исследовательский клинический институт им. М. Ф. Владимирского», г.Москва ул. Щепкина 61/2 корпус 8
yulia.molochkova@yandex.ru*

Аннотация: Аппроксимация дискриминантной функции Андерсона в точке линейной функцией по обучающей выборке с учителем позволяет оценивать апостериорные вероятности классов в точке без использования условных распределений вероятностей признаков классов и без применения надстроек типа калибратора Платта. Способ оценки апостериорных вероятностей классов применен для разработки компьютерной программы дифференциальной диагностики красного плоского лишая слизистой оболочки полости рта с лейкоплакией и плоскоклеточным раком. Диагностирование по апостериорным вероятностям заболевания у обследуемого пациента повышает обоснованность диагноза и существенно облегчает выбор лечебно-диагностической тактики.

Ключевые слова: дискриминантная функция, оценка апостериорной вероятности класса, красный плоский лишай слизистой оболочки полости рта, лейкоплакия полости рта, плоскоклеточный рак полости рта.

Введение

Апостериорные вероятности (АпоВ) классов [1] при решении задач классификации объектов, диагностики, прогнозирования и т.п. являются исчерпывающей информацией для решения задачи. Они улучшают интерпретируемость результатов, дают возможность применять различные критерии принятия решений, использовать разные стоимости ошибок классификации, намечать стратегию дальнейших действий с объектом.

Методы классификации, ориентированные на минимизацию среднего риска или средней стоимости ошибок классификации, для оценивания АпоВ классов вынуждены усложняться специальными надстройками для пересчета отступов от границ в АпоВ классов, как например, калибратор Платта [2] для метода опорных векторов. Методы, основанные на использовании условных распределений признаков классов и применении формулы Байеса для расчета АпоВ классов, нуждаются в анализе законов распределений признаков и требуют, соответственно, достаточно больших объемов обучающих выборок с учителем. Логистическая регрессия, оперирующая с АпоВ классов, использует сигмоидную функцию для расчета АпоВ класса, предполагая в сущности нормальные условные распределения признаков классов. Метод классификации, использующий дерево решений, дает смещенные оценки АпоВ классов. Для компенсации смещений нужно использовать специальный калибратор.

Весьма перспективным для решения задач классификации является использование дискриминантной функции (ДФ) Андерсона (ДФА), получаемой нами из метода байесовой классификации в изложении Т. Андерсона [1] в виде разности средних стоимостей ошибок классификации при отнесении точки в один из двух классов в многоклассовой задаче. Дискриминантную функцию в таком определении мы называем ДФ Андерсона. Она по определению является функцией регрессии и для ее оценки в заданной точке пространства признаков по обучающей выборке с учителем достаточно взвешенного метода наименьших квадратов. Сама обучающая выборка задачи классификации просто превращается в выборку регрессионного анализа заменой в выборке номеров классов на разности соответствующих стоимостей ошибок классификации. В задаче с двумя классами такое оценивание ДФА в точке тождественно оцениванию АпоВ классов, причем полученная оценка АпоВ не зависит от задаваемых стоимостей ошибок классификации, определяющих ДФА. Пересчитать оценку ДФА в точке в оценку АпоВ класса и наоборот также просто, как пересчитать градусы Фаренгейта в градусы Цельсия и наоборот.

Для оценки ДФА в точке достаточно использовать линейную зависимость ее от признаков, поскольку нужно оценивать ДФА не во всей области обучающей выборки. Оценивается лишь касательная к ДФА в заданной точке. Простота метода позволяет использовать его на малых выборках,

о чем свидетельствует данный доклад. Обучающая выборка в силу объективных причин насчитывает всего 36 точек в задаче с тремя классами. Трех-классовая задача заменяется тремя двух-классовыми по принципу один класс против остальных.

В данной работе способ оценки АпоВ классов в точке по линейной аппроксимации ДФА применен для разработки компьютерной программы дифференциальной диагностики красного плоского лишая слизистой оболочки полости рта (КПЛ СОПР) с лейкоплакией и плоскоклеточным раком (ПКР). КПЛ СОПР наряду с лейкоплакией той же локализации относятся к наиболее частым потенциально злокачественным заболеваниям этой области. В свою очередь, при длительном существовании КПЛ СОПР может приобретать клиническое сходство с лейкоплакией. Частота трансформации КПЛ СОПР в одно из самых распространенных и агрессивных онокологических заболеваний полости рта ПКР по данным разных источников варьирует от 0 до 5,3%, в то время как частота трансформации лейкоплакии в ПКР оценивается выше и составляет от 0.13 до 17.5% [3]. Зачастую, эти заболевания обладают клиническим сходством, что сильно усложняет выбор верной тактики для обследования и лечения таких пациентов [4].

В настоящее время, для выявления ранней злокачественной трансформации потенциально предраковых поражений СОПР уникальная роль отводится применению надежных биомаркеров. Как правило, для злокачественных процессов, по сравнению с предраковыми и воспалительными, характерно существенное повышение пролиферативной активности клеток. Одним из наиболее надежных индикаторов пролиферативной активности клеток является антиген Ki-67, представляющий собой ядерный негистоновый белок, состоящий из двух полипептидных цепей, с молекулярными массами 320 и 359 кДа [5]. Экспрессия белка строго связана с клеточной пролиферацией, антитела к Ki-67 экспрессируются делящимися клетками, находящимися в G1, S, M, G2 стадиях клеточного цикла [6]. Экспрессия белка Ki-67 значительно возрастает при злокачественных процессах, по сравнению с доброкачественными и предраком. Аналогичную закономерность ранее наблюдали при поражениях СОПР (при ПКР СОПР и КПЛ СОПР) [7]. Также известна ассоциация злокачественного потенциала опухолей с особенностями экспрессии в клетках адгезивных молекул, в частности - E-кадгерина-кальций-зависимой адгезивной молекулы клеток эпителиальных тканей [8]. Нарушение E-кадгерин-опосредованных связей приводит к увеличению клеточной инвазии и миграции [8]. По данным различных исследований, аномальная экспрессия E-кадгерина характерна для злокачественных эпителиальных опухолей: колоректального рака, рака пищевода и желудка, печени, легкого, молочной железы, предстательной железы, злокачественных опухолей эндометрия [9]. Нарушение экспрессии E-кадгерина наблюдали и в клетках предраковых состояний, в частности при лейкоплакии и цервикальной интраэпителиальной неоплазии [10]. Таким образом, этот маркер может применяться также для раннего выявления злокачественной трансформации КПЛ СОПР. На настоящий момент не существует диагностического алгоритма интерпретации результатов иммуногистохимического исследования (ИГХ) с анализом экспрессии маркеров адгезии E-кадгерина и пролиферации Ki-67 для дифференциальной диагностики КПЛ, лейкоплакии и ПКР.

В связи с этим, целью работы явилась разработка компьютерной программы с применением метода оценки АпоВ классов в точке по линейной аппроксимации ДФА для решения задачи дифференциальной диагностики КПЛ СОПР, лейкоплакии и ПКР той же локализации по данным ИГХ исследования в инцизионном панч-биоптате очага поражения уровня пролиферации с использованием маркера Ki-67 и адгезии с использованием E-кадгерина. В работе приведены результаты гистологических исследований и ИГХ изучения экспрессии маркеров пролиферации Ki-67 и адгезии E-кадгерина в 36 биоптатах СОПР больных КПЛ, лейкоплакией, ПКР и результаты расчетов апостериорной вероятности (АпВ) этих заболеваний СОПР с близкой клинической симптоматикой.

Был осуществлен сбор и обработка экспериментальных данных (обучающей выборки, Таблица 1), состоящей из строк признаков и достоверно установленных диагнозов заболеваний у пациентов, участвовавших в обследовании.

Таблица 1. Обучающая выборка

N	K1	Шифр	Ki-67	E-кадгерин				
1	2	3	4	5	6	7	8	9
1	1	10573	52	0	0	10	60	30
2	1	10572	55	0	0	0	100	0
3	1	32276	35	0	10	10	0	80
4	1	35280	60	0	10	0	70	20
5	1	35281	80	0	0	0	80	20
6	1	59657	35	30	20	0	20	30

N	Кл	Шифр	Ки-67	Е-кадгерин				
				0	5	0	25	70
7	1	45376	85	0	5	0	25	70
8	2	25838	30	90	10	0	0	0
9	2	54827	25	90	10	0	0	0
10	2	826	28	80	20	0	0	0
11	2	47878	15	85	15	0	0	0
12	2	48244	21	85	15	0	0	0
13	2	22402	20	40	10	0	0	50
14	2	18747	25	90	10	0	0	0
15	2	16959	28	93	7	0	0	0
16	2	54314	15	100	0	0	0	0
17	2	37774	17	100	0	0	0	0
18	2	14308	17	100	0	0	0	0
19	2	20501	17	30	20	50	0	0
20	2	42866	10	60	20	0	0	20
21	3	37837	10	80	20	0	0	0
22	3	7773	7	100	0	0	0	0
23	3	1695	8	100	0	0	0	0
24	3	37837	7	100	0	0	0	0
25	3	7652	6	100	0	0	0	0
26	3	6914	12	100	0	0	0	0
27	3	6916	11	100	0	0	0	0
28	3	30831	7	100	0	0	0	0
29	3	16156	8	100	0	0	0	0
30	3	31390	6	100	0	0	0	0
31	3	8652	8	100	0	0	0	0
32	3	8389	10	100	0	0	0	0
33	3	1353	7	100	0	0	0	0
34	3	8387	13	100	0	0	0	0
35	3	1917	8	100	0	0	0	0
36	3	8854	12	100	0	0	0	0

Столбцы таблицы по номерам: 1 – порядковый номер строки выборки N ; 2-9 – обучающая выборка, где Kl –заболевание, обозначаемое номерами классов: 1 – ПКР, 2 - Лейкоплакия, 3 – КПЛ; столбец 3 – шифр пациента; столбец 4 – индекс Ки-67; столбцы 5-9 – различные показатели экспрессии Е-кадгерина по столбцам: 5 – мембранная , 6 – мембранно-редуцированная, 7 – мембранно-цитоплазматическая, 8 – цитоплазматическая, 9 – отсутствие иммунореактивности.

2 Математическая постановка задачи

2.1 ДФ Андерсона и ее свойства

Определение. ДФ Андерсона, как разность средних стоимостей ошибок классификации по Т. Андерсону [1] при отнесении точки x в один из пары классов r, s , $k=(1, \dots, K)$, K – количество классов, есть

$$(1) f_{rs}(x, C) = g_r(x, C) - g_s(x, C) = M_{k|x}(C_{rk} - C_{sk}), \quad C_{rk|x} - C_{sk|x} = f_{rs}(x, C) + \varepsilon_{rs}(x, C),$$

где $g_i(x, C) = \sum_k C_{ik} p(k|x)$ - средняя стоимость ошибок классификации при отнесении точки x в класс i , вычисляемая по апостериорным вероятностям классов $p(k|x)$ и матрице стоимости ошибок C , $M_{k|x}(\cdot)$ – математическое ожидание по k в точке x , $\varepsilon_{rs}(x, C)$ – ошибка оценки ДФА в точке x , имеющая нулевое среднее, C_{rk} – стоимость ошибки, если точка из класса k ошибочно относится в класс r , $C_{rk} > 0$, $C_{rr} = 0$. Если $f_{rs}(x, C) \leq 0$, то точку x предпочтительнее отнести в класс r , чем в класс s .

Из определения (1) видно, что ДФ Андерсона есть функция регрессии. Для ее аппроксимации по обучающей выборке с учителем (k_n, x_n) , где k_n – номер класса в строке n , x_n – вектор признаков в той же строке, нужно всего лишь номера классов в строке выборки заменить на разности $C_{rk_n} - C_{sk_n}$. Эти разности есть оценки ДФА в выборке регрессионного анализа. В случае двух классов ДФА принимает вид

$$(2) \quad f_{12}(x, C) = g_1(x, C) - g_2(x, C) = C_{12}(1 - p(1|x)) - C_{21}p(1|x) = C_{12} - (C_{12} + C_{21})p(1|x).$$

Для ее аппроксимации в строке выборки вместо номера класса 1 следует поставить $-C_{21}$, а вместо класса 2 поставить C_{12} . Из (2) получается выражение для АпоВ первого класса

$$(3) \quad p(1|x) = (C_{12} - f_{12}(x)) / (C_{12} + C_{21}).$$

В точках на границе между классами (там ДФА равна нулю) АпоВ первого класса $p^* = p(1|x)$ постоянна и определяется только стоимостями двух ошибок классификации. Если стоимости ошибок выбирать из условия, что их сумма равна единице, что не приводит к потере общности, то на границе между классами

$$(4) \quad p^* = C_{12},$$

и (3) примет вид

$$(5) \quad p(1|x) \equiv p^* - f_{12}(x, p^*).$$

Иными словами, задание C_{12} в этом случае для построения ДФА равнозначно заданию АпоВ первого класса p^* на границе между классами. В области точек первого класса, (там $f_{12}(x, p^*) \leq 0$) АпоВ первого класса $p(1|x) \geq p^*$, как следует из (5). В выражение для ДФА введен аргумент p^* , отражающий зависимость ДФА от стоимостей ошибок классификации $C_{12} = p^*$, $C_{21} = 1 - p^*$. АпоВ классов от p^* не зависят, являясь объективной характеристикой решаемой задачи. Но стоимости ошибок задает пользователь, привнося в задачу элемент субъективности. Субъективность, в частности, приводит к тому, что при некотором выборе стоимостей ошибок классификации задача собственно классификации теряет свой смысл: любую точку надо относить в один и тот же класс, поскольку знак ДФА теряет зависимость от x . Но оценка АпоВ в заданной точке никак не зависит от того, существует или нет граница между классами в пространстве признаков при заданных стоимостях ошибок классификации. Для метода оценивания АпоВ классов стоимости ошибок классификации можно задавать любые, в частности, можно всегда задавать $p^* = 0.5$, заменив в (5) переменную константой.

2.2 Исходные данные

В работе использовали 36 биопанов СОПР: 7 больных плоскоклеточным раком, класс №1, 13 – лейкоплакией, класс №2, и 16 больных КПЛ, класс №3 в таблице 1, столбец 2. Иммуногистохимическое исследование проводили в соответствии со стандартным протоколом. Определение уровня пролиферации оценивалась с помощью индекса, который вычислялся как среднее от числа меченных ядер на 100 учтенных ядер (при учете 500-1000 опухолевых клеток). Подсчет меченных ядер проводился в репрезентативных полях зрения с относительно равномерным распределением опухолевых клеток, сверху вниз и слева направо. Клетки на периферии срезов, где чаще всего наблюдается фоновое окрашивание, не учитывались.

Для оценки характера экспрессии Е-кадгерина выделяли следующие типы: Мембранный – равномерное распределение маркера по всей клеточной мембране; Мембранно-редуцированный – распределение маркера на отдельных участках клеточной мембраны; Цитоплазматический – распределение маркера в цитоплазме (мелкогранулярный в виде мелких гранул); Мембранно-цитоплазматический (сочетание мембранного и цитоплазматического); Отсутствие иммунореактивности.

2.3 Подход к решению задачи

Метод, использованный в данной работе, по любой заданной вероятности p^* , что не влияет на результат, строит линейную аппроксимацию ДФА в заданной точке и по тождеству (5) вычисляет АпоВ в этой точке. Метод позволяет наращивать объем обучающей выборки, удалять и добавлять строки. Суть метода изложена в следующем разделе.

Диагноз устанавливается по наибольшей вычисленной для пациента АпоВ класса заболевания. Сравнения АпоВ классов заболевания для пациента позволяют судить и о степени надежности диагноза. При низкой надежности (близкие вероятности двух заболеваний) возможно дальнейшее обследование пациента для уточнения диагноза с использованием, может быть, другой обучающей выборки, но с использованием той же программы.

Для оценки работоспособности компьютерного метода, устанавливающего диагноз по максимальной АпоВ болезни, нужно подсчитать количество несовпадений компьютерных решений с решениями учителя в столбце 2 Таблицы 1 или Таблицы 2. Для улучшения обобщающей способности метода (generalization ability), т. е. для борьбы с переобучением (overfitting) при расчете АпоВ классов

пациентов, входящих в обучающую выборку из последней исключались строки с данными для рассчитываемых пациентов - метод leave-one-out. В результатах решения задачи в Таблице 2 в столбцах 4,5,6 представлены расчеты вероятностей классов без отбрасывания из обучения рассчитываемых точек, а в столбцах 7,8,9 – с отбрасыванием.

На полученных данных ограниченного объема, нужно было оценить перспективу применения ИГХ в качестве скринингового метода диагностики КПЛ, ПКР и лейкоплакии СОПР и обосновать необходимость продолжения исследований с целью получения более надежных результатов, например, на выборках большего объема.

3 Решение задачи

Основу метода составляет ДФ Андерсона как регрессионная зависимость, для аппроксимации которой имеется выборка регрессионного анализа, получаемая из обучающей выборки с учителем задачи классификации простой заменой в выборке номеров двух классов значениями -0,5 и 0,5, т.е. используется значение параметра $p^*=0,5$. Линейная аппроксимация ДФ Андерсона в точке выполняется взвешенным методом наименьших квадратов.

Имеется обучающая выборка с учителем, состоящая из строк (k_n, x_n) , где в строке n находятся номер класса k_n и вектор-строка признаков x_n .

Задача решается для двух классов. Если классов больше (в нашем случае их три), то последовательно для заданной точки в пространстве признаков (результатов анализов) пациента решается вначале задача, например, для первого заболевания против остальных и находится АпоВ первого заболевания для пациента, потом для второго заболевания и т.д.. Затем выполняется нормировка, чтобы в сумме АпоВ классов давали единицу.

В качестве аппроксимации ДФА используется линейная комбинация $(1, x)\lambda$ признаков (в общем случае и функций от признаков x), отобранных для этой цели. Компонент вектора-строки – единица необходим для свободного члена линейной комбинации. Вектор-столбец коэффициентов λ подбирается так, чтобы точнее аппроксимировать ДФА в окрестности заданной точки.

Критерий взвешенного метода наименьших квадратов для вычисления λ :

$$(6) Q(\lambda) = \min_{\lambda} \sum_{n=1}^{n=N} \{ [C_{1k_n} - C_{2k_n} - (1, x_n)\lambda]^2 \exp(-W \|x - x_n\|) \},$$

где вектор-строка $(1, x_n)$ кроме строки признаков x_n обучающей выборки содержит компонент-единицу: N – количество строк в выборке; k_n – номер класса 1 или 2 в строке n ; x_n – вектор признаков в строке выборки n ; $W \geq 0$ – скалярный весовой коэффициент, задающий спад весовой функции по мере удаления точки выборки от заданной точки. В качестве весовой функции может использоваться не только экспонента.

Оптимальный вектор λ получается путем решения системы линейных уравнений – условий минимума критерия (6). Размерность матрицы определяется количеством компонент вектора признаков, она не зависит от объема обучающей выборки. При больших значениях W матрица задачи становится вырожденной – слишком мало соседних точек влияют на результат – эффект переобучения. При малых W влияющих точек становится много и оценка ДФА в точке ухудшается. Оптимальное значение W находится простым перебором по минимуму ошибок классификации в каждой из трех задач - один класс против двух других, рассматриваемых в задаче как второй класс.

АпоВ класса вычисляется по (5), где вместо f_{12} нужно использовать скалярное произведение $(1, x)\lambda$.

Диагноз по каждому пациенту устанавливается по максимальной из трех АпоВ классов заболеваний. За ошибку классификации считается несовпадение компьютерного диагноза и номера болезни в обучающей выборке.

Для борьбы с переобучением использовался метод leav-one-out. При расчете АпоВ классов из числа обучающих исключались данные пациента, для которого выполнялся расчет. Его данные переводились в ранг тестовой точки. Результаты расчетов представлены столбцами 4-9 Таблицы 2. В столбцах 4,5,6 – результат, полученный без отбрасывания из обучающих тестовой точки. В столбцах 7,8,9 – с отбрасыванием.

Выбор лучшей величины коэффициента W для каждой из трех задач производился перебором из нескольких значений такой величины коэффициента, которой соответствует минимальная сумма ошибок классификации. Ошибка классификации - это несовпадение номера класса P1-P3 с максимальной АпоВ в строке с номером класса Kl в той же строке. Строки ошибок в Таблице 2 выделены полужирным шрифтом

Таблица 2. Результат решения задачи

N	K1	Шифр	p1	p2	p3	P1	P2	P3
1	2	3	4	5	6	7	8	9
1	1	10573	1,00	0,00	0,00	1,00	0,00	0,00
2	1	10572	1,00	0,00	0,00	1,00	0,00	0,00
3	1	32276	1,00	0,00	0,00	0,00	1,00	0,00
4	1	35280	1,00	0,00	0,00	1,00	0,00	0,00
5	1	35281	1,00	0,00	0,00	1,00	0,00	0,00
6	1	59657	1,00	0,00	0,00	0,47	0,53	0,00
7	1	45376	1,00	0,00	0,00	1,00	0,00	0,00
8	2	25838	0,00	1,00	0,00	0,00	1,00	0,00
9	2	54827	0,00	0,99	0,01	0,00	0,99	0,01
10	2	826	0,00	1,00	0,00	0,00	1,00	0,00
11	2	47878	0,00	0,89	0,11	0,00	0,47	0,53
12	2	48244	0,00	0,99	0,01	0,00	0,94	0,06
13	2	22402	0,00	1,00	0,00	0,35	0,64	0,01
14	2	18747	0,00	0,99	0,01	0,00	0,99	0,01
15	2	16959	0,00	1,00	0,00	0,00	0,99	0,01
16	2	54314	0,00	0,71	0,29	0,00	0,55	0,45
17	2	37774	0,00	0,98	0,02	0,00	0,96	0,04
18	2	14308	0,00	0,98	0,02	0,00	0,96	0,04
19	2	20501	0,00	1,00	0,00	0,00	1,00	0,00
20	2	42866	0,00	1,00	0,00	0,07	0,15	0,79
21	3	37837	0,00	0,10	0,90	0,00	0,68	0,32
22	3	7773	0,00	0,00	1,00	0,00	0,00	1,00
23	3	1695	0,00	0,01	0,99	0,00	0,01	0,99
24	3	37838	0,00	0,00	1,00	0,00	0,00	1,00
25	3	7652	0,00	0,00	1,00	0,00	0,00	1,00
26	3	6914	0,00	0,18	0,82	0,00	0,22	0,78
27	3	6916	0,00	0,12	0,88	0,00	0,15	0,85
28	3	30831	0,00	0,00	1,00	0,00	0,00	1,00
29	3	16156	0,00	0,01	0,99	0,00	0,01	0,99
30	3	31390	0,00	0,00	1,00	0,00	0,00	1,00
31	3	8652	0,00	0,01	0,99	0,00	0,01	0,99
32	3	8389	0,00	0,08	0,92	0,00	0,09	0,91
33	3	1353	0,00	0,00	1,00	0,00	0,00	1,00
34	3	8387	0,00	0,30	0,70	0,00	0,38	0,62
35	3	1917	0,00	0,01	0,99	0,00	0,01	0,99
36	3	8854	0,00	0,18	0,82	0,00	0,22	0,78

Из Рис. 1 (проекция точек выборки на плоскость из двух координат) видно, что точки класса Лейкоплакия принципиально нельзя отделить линейной поверхностью от точек объединенного класса из КПЛ и плоскоклеточного рака. Но для аппроксимации ДФА в точке линейной функцией не имеет значения, какой вид имеют границы классов.

В силу особенностей выборки, Рис. 1, для определения АпоВ первого класса использованы признаки в столбцах 4 и 9 Таблицы 1, для второго и третьего – признаки в столбцах 4 и 5. Причина - малое количество строк выборки. Все признаки, кроме первого, изменяются мало и они между собой находятся в сильной зависимости. При определенных сочетаниях используемых признаков отбрасывание из обучения некоторой строки, используемой в качестве тестовой, приводило к вырождению матрицы решения задачи для этой строки.

Программа создана инструментальными средствами МАТЛАБ и ПИТОН. Время решения каждой из двух задач в программе с сохранением тестовой точки в выборке и с ее удалением составило 1,3 сек на ноутбуке Prestigio Smartbook 116C, Процессор: Intel 1.1® Atom™ x5-Z8350 CPU @ 1.44 GHz 1.44 GHz, ОЗУ 2 Гб. Операционная система 32-х разрядная, процессор x64, Windows 10 Home.

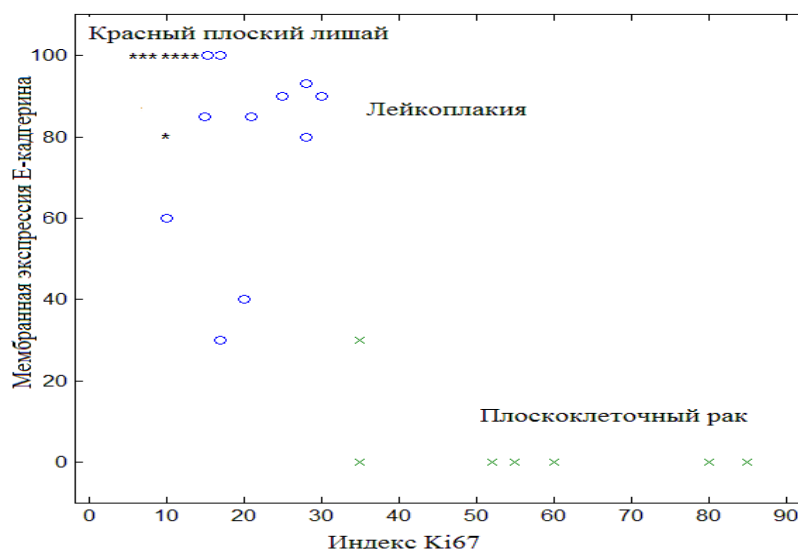


Рис. 1. Точки выборки

Работа выполнена в рамках комплексного проекта Министерства Образования и Науки России №10.331-17.

Заключение

1. Для постановки диагноза заболевания пациента вычислялись по обучающей выборке с учителем апостериорные вероятности трех заболеваний. Решение принималось в пользу той болезни, для которой вероятность была наибольшей. Наличие вероятностей для рассмотренных заболеваний позволяет выносить более обоснованное решение.
2. Метод диагностики использует компьютерную программу для расчета вероятностей заболеваний по совокупности анализов обследуемого пациента. Имеется возможность без изменения программы пополнять обучающую выборку новыми строками и удалять сомнительные строки, не внося изменений в программу.
3. Важность оценки перспективы применения описанного метода дифференциальной диагностики заболеваний слизистой оболочки полости рта связана с их широкой распространенностью, частым клиническим сходством, возможностью малигнизации данных патологических состояний и неблагоприятным прогнозом одного из самых агрессивных онкологических заболеваний – плоскоклеточным раком слизистой оболочки полости рта.

Литература

1. Anderson T.W. An Introduction to Multivariate Statistical Analysis. Third Edition. – Wiley. – 2003. – P. – 721.
2. Platt J. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods (PDF). – Advances in large margin classifiers. 10 (3): 61–74.
3. Garcia-Pola M.J., Llorente-Pendas S., Gonsales-Garcia M., Ggarcia-Martin J.M. The development of proliferative verrucous leukoplakia in oral lichen planus. A preliminary study // Med Oral Patol Oral Cir Bucal. – 2016. – V.21. – №3. – 328-334.
4. Speight P.M., Khurram S.A., Kujan O. Oral potentially malignant disorders: risk of progression to malignancy // Oral Surg Oral Med Oral Pathol Oral Radiol. – 2017. – pii: S2212- 4403(17)31248-8.
5. Duhrow M., Schluter C., Wohlenberg C., Flad H.D., Gerdes J Molecular characterization of the gene locus of the human cell proliferation-associated nuclear protein defined by monoclonal antibody Ki-67 // Cell Prolif. – 1996. – Vol.29. – № 1. – P.1–12.
6. Molino A., Micciolo R., Turazza M., Bonetti F., Piubello Q., Bbonetti A., Nortilli R., Pelossi G., Cetto G.L. Ki-67 immunostaining in 322 primary breast cancers: associations with clinical and pathological variables and prognosis // Int J Cancer. – 1997. – Vol. 74. – №4. – P.433-7.
7. Zargarman M., Jamshidi S., Eshghran N., Moghimbeigi A. Suitability/unsuitability of cell proliferation as an indicator of malignant potential in oral lichen planus: an immunohistochemical study // Asian Pac J Cancer Prev. – 2013. – Vol. 14. – № 11. – P. 79–83.

8. Tsalikidis C., Papachristou F., Pitiakoudis M., Asimakoupoulus B., Tripsianis G., Bolanari E., Syrigos K.N., Simopoulos C. Soluble E-cadherin. as a diagnostic and prognostic marker in gastric carcinoma // *Folia Med (Plovdiv)*. – 2013. – Vol. 55. – №3–4. – P. 26–32.
9. *Wojcik-Krowiranda K., Forma E., Zaczek A., Brys M., Anna M.K., Bienkiewich A.* Expression of E-cadherin and beta1-integrin mRNA in endometrial cancer // *Ginekol Pol.* – 2013. – Vol. 84. – № 11. – P. 910-914;
10. *Vizirianakis I.S.* Desmoglein-3/ γ -catenin and E-cadherin/ β -catenin differential expression in oral leukoplakia and squamous cell carcinoma // *Clin Oral Investig.* – 2014. – Vol. 18. – №1. – P. 199–210.