КОМПЬЮТЕРИЗАЦИЯ ПРОЦЕССА ВИДОВОЙ ИДЕНТИФИКАЦИИ МИКРООРГАНИЗМОВ

Дерябин Д.Г., Фот Н.П., Бравичева О.С.

Оренбургский государственный университет, г.Оренбург

На основе процедуры кластерного анализа разработана компьютерная программа видовой идентификации микроорганизмов, основными преимуществами которой перед существующими аналогами являются возможность получения требуемого результата при любой комбинации дифференцирующих тестов, а также совместимость со всеми используемыми для этой цели диагностическими тест-системами. С использованием данной программы проанализирована диагностическая эффективность пяти наиболее распространенных коммерческих тест-систем для идентификации стафилококков и определены причины, определяющие предел их диагностических возможностей. Полученные результаты позволили автоматизировать процедуру видовой идентификации микроорганизмов, а также разработать рекомендации по оптимизации используемых для этих целей диагностических тест-систем

Идентификация микроорганизма с установлением бинарного наименования, состоящего из латинского наименования его родовой и видовой принадлежности (например – Staphylococcus aureus) является стандартным требованием при лабораторном исследовании. Однако, происходящее в последние десятилетия лавинообразное увеличение количества описанных видов микроорганизмов с резким увеличением числа дифференцирующих тестов, сделало невозможным использование традиционных дихотомических подходов к их идентификации. Реальностью стала необходимость проведения многоальтернативной дифференциации в условиях, когда принадлежность к каждому из потенциально возможных видов определяется сочетанием положительных и отрицательных результатов значительной по численности совокупности равноправных идентификационных тестов.

Указанные обстоятельства определили актуальность задачи компьютеризации процесса видовой идентификации микроорганизмов с акцентом на разработку вероятностно-статистических подходов к обработке и анализу результатов комплекса дифференцирующих тестов, что является основой для принятия обоснованного решения об отнесении микроорганизма к тому или иному определенному виду. При этом в связи с многоплановостью и сложностью исследуемых объектов (для каждого объекта регистрируются значения множества переменных), на первый план выходят проблемы построения группировок и классификаций по многомерным данным.

Электронно-вычислительная техника позволяет воспользоваться разработанным аппаратом многомерного статистического анализа, среди которых следует выделить методы распознавания объектов «с учителем» (например - дискриминантный анализ) и «без учителя» (автоматическая классификация или кластерный анализ). При этом если на «входе» задачи располагают не только классифицируемыми данными, но и обучающими выборками, то для решения задачи обращаются к методам дискриминантного анализа («классификация с обучением»); в противном случае речь идет о «классификации без обучения» и применяют методы кластер-анализа [1,3].

Попытки решения поставленной задачи с использованием с использованием дискриминантного анализа продемонстрировали его неприменимость при имеющейся форме исходных данных вследствие несоблюдения условия о нормальности их распределения. С другой стороны, использование кластерного анализ позволило вплотную приблизиться к решению классификации имеющихся многомерных наблюдений, каждое из которых описывалось набором независимых и равнозначимых исходных переменных.

При применении методов кластерного анализа для целей видовой идентификации стафилококков исходные данные были представлены в виде матрицы размерностью 36X44 (36 известных видов стафилококков, охарактеризованных по 44 таксономически значимым характеристикам). Проведение дальнейших вычислений с использованием обычного Евклидово расстояния, принимающего все таксономически значимые признаки как равноценные для целей идентификации, по формуле: $d_{ij}^2 = \Sigma_{k=1}^m (x_{ik} - x_{jk})^2$; где $x_{ik} - x_{jk}$ – расстояние между i и j видом стафилококка, принадлежащие к k – му классу, позволяет представить каждый из вновь идентифицируемых микроорганизмов как некую «точку» в 44-мерном признаковом пространстве [7], где геометрическая близость двух точек означает однородность их видовой принадлежности. Соответственно, при изменении исходных данных (расширении перечня идентифицируемых видов или расширении числа идентификационных признаков) общая структура расчетов не меняется, а новые данные просто добавляются в уже сложившуюся картину дифференциации.

На данной основе в среде программирования Delphi 5 была разработана программа для персонального компьютера [8], принципиальная болк-схема которой представлена на рис.1. На «входе» программы вводятся исходные данные, характеризующие наличие или отсутствие у идентифицируемого микроорганизма ряда его таксономически значимых характеристик. В дальнейшем при решении задачи идентификации происходит обращение к методу кластерного анализа, с использованием которого производится расчет «расстояния» объекта до каждого из 36 известных видов стафилококков. Соответственно, на «выходе» программы представляется заключение о степени подобия вновь идентифицируемого микроорганизма каждому из известных видов стафилококков по мере убывания подобной вероятности и с указанием возможной ошибки идентификации.

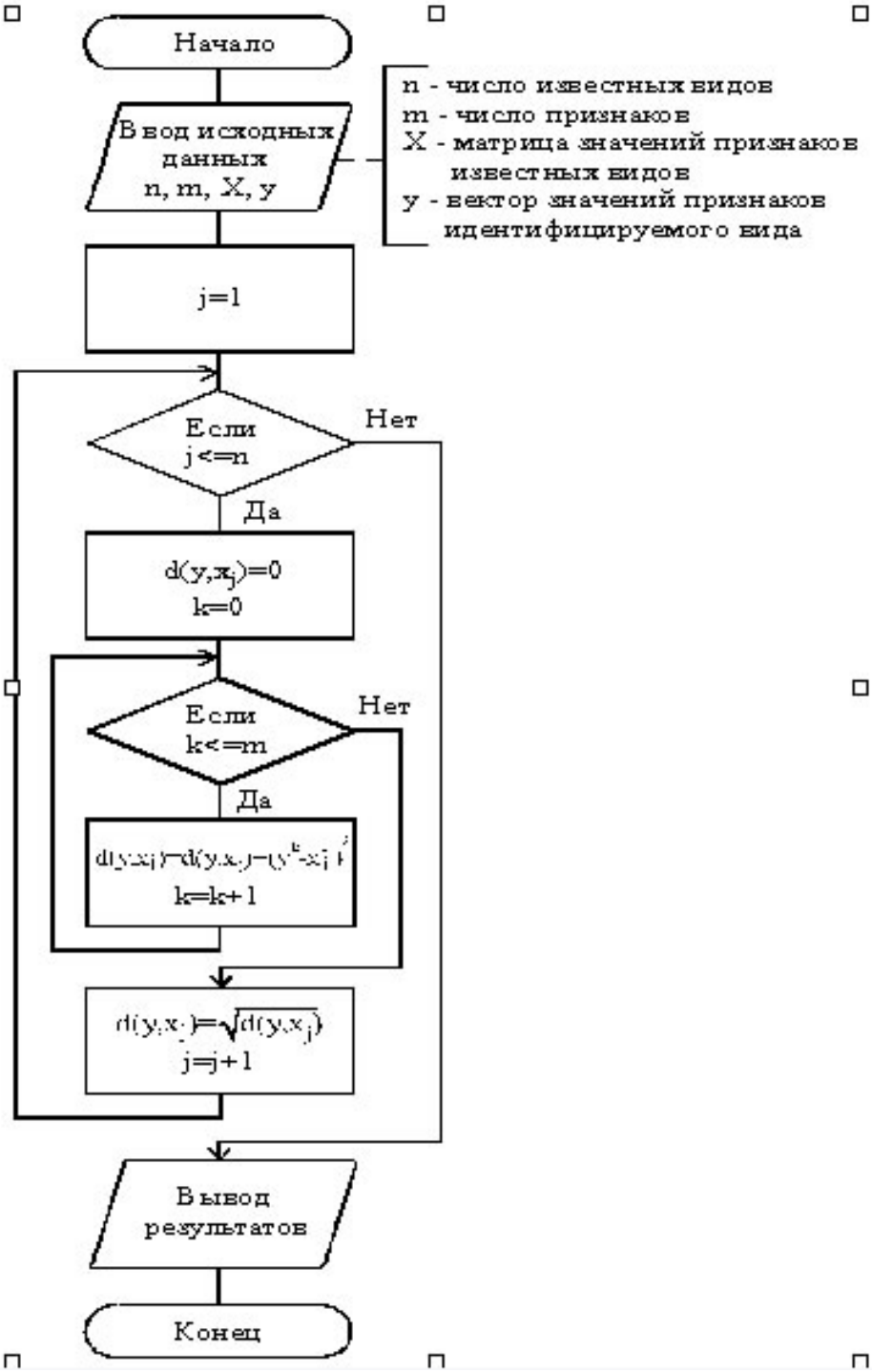

Рис. 1. Блок-схема программы для видовой идентификации микроорганизмов

После запуска программы на экране появляется основная форма «Универсальной программы для видовой идентификации микроорганизмов», в которой из нескольких групп микроорганизмов на основании предварительно проведенных исследований (морфологические, тинкториальные и др. свойства) предлагается выбрать одну для дальнейшей идентификации.

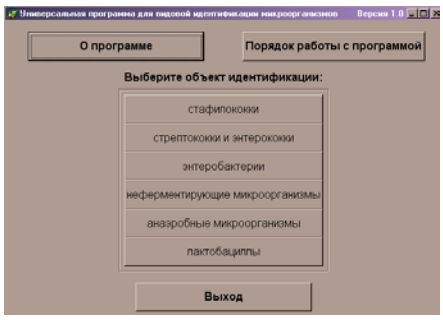

Рис. 2.1. Стартовое окно программы.

После нажатия соответствующей кнопки на экране появляется форма ввода данных, представляющая из себя таблицу с перечнем таксономически значимых характеристик, используемых для видовой идентификации определенной группы микроорганизмов (рис.2.2). При этом в ячейки, соответствующие отдельным биологическим характеристикам, использованным для идентификации исследуемого микроорганизма, вводятся результаты тестов: «+» – наличие признака; «-» - отсутствие признака. В случае, если какие-либо тесты не выполнялись или при их учете был получен нечеткий результат, соответствующие ячейки таблицы не заполняются. По окончании ввода значений признаков следует нажатие кнопки «Ввести данные», после чего становится активной кнопка «Расчет».

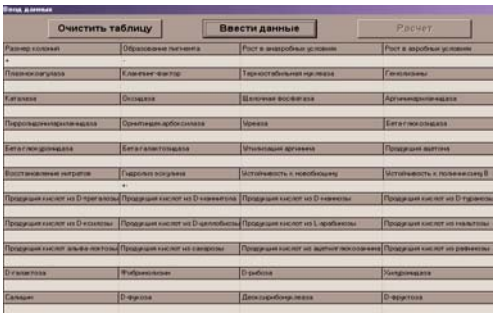

Рис. 2.2. Таблица ввода исходных данных.

Итоговая форма «Результат идентификации» состоит из четырех таблиц (рис.2.3). В первой (верхней) приводятся значения Евклидова расстояния между идентифицируемым штаммом и известными видами микроорганизмов с указанием вероятности подобного заключения. Во второй и третьей таблицах приводятся результаты тестов «за» и «против» отнесения идентифицируемого штамма микроорганизма к текущему виду из первой таблицы. При этом под второй и третьей таблицами указывается общее число совпавших и несовпавших тестов соответственно. В четвертой (нижней) таблице приводятся рекомендуемые значения дополнительных тестов для более точной дифференциации между тремя наиболее вероятными видами микроорганизмов.

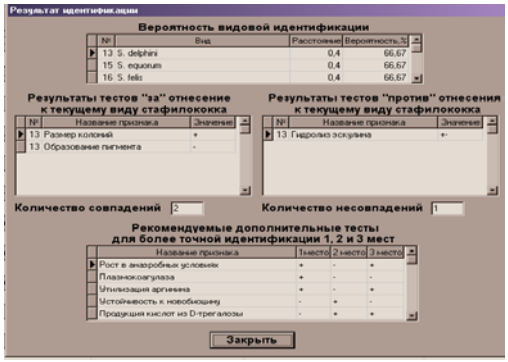

Рис 2.3. Форма представления результатов идентификации.

Разработанная «Универсальная компьютерная программа идентификации микроорганизмов» зарегистрирована в качестве оригинального программного продукта для ЭВМ (сертификат № 2004610330 от 2 февраля 2004г.). При этом исследование ее диагностической эффективности на примере 23 эталонных и 150 клинических штаммов стафилококков, предварительно идентифицированных до вида референс-методом в соответствии с «Определителем бактерий Берджи» [6], свидетельствовало о не менее чем 95% точности при вероятности идентификации каждого штамма на уровне не менее 85%. Основным же отличием данной программы от известных аналогов является то, что она жестко не связана ни с одной из известных коммерческих диагностических тест-систем и представляет собой «открытую систему», позволяющую получать требуемый результат при использовании любой комбинации конвенциональных таксономически значимых тестов, в том числе самостоятельно выполняемых пользователем.

Последняя особенность определила возможность использования разработанной программы для проведения «виртуального эксперимента», сущностью которого стала оценка пределов диагностической эффективности пяти наиболее распространенных коммерческих тест-систем, используемых для идентификации представителей рода Staphylococcus [5]. При этом были изучены пять наиболее широко представленных на российском рынке тест-систем: ПБДС (пластина биохимическая дифференцирующая стафилококки), производимая НПО «Диагностические системы» (Россия), «STAPHYtest-8» и «STAPHYtest-16» (Lachema, Чехия), а также API-Staph и ID32 («bioMerieux», Франция) [9-17].

Результаты идентификации изученной выборки стафилококков по совокупности тестов, предусмотренных при использовании системы ПБДС, позволили констатировать совпадение с результатами, полученными с использованием референс-метода в57,14±1,06 % случаев (рисунок 3.А). При этом причины неправильной видовой идентификации стафилококков при использовании тест-системы ПБДС могли быть сведены к двум основным моментам. Ошибки первого рода были связаны с тем, что диагностические возможности данной тест-системы принципиально не позволяли получить верное заключение, так как она изначально предусматривала возможность идентификации только 19 из 36 известных видов и подвидов стафилококков (14,34±0,98 % от общего количества диагностических заключений). Ошибки второго рода определялись тем, что несмотря на заявленные диагностические возможности тест-системы как достаточные для идентификации определенного спектра видов стафилококков, полученный результат все же оказывался неверным. На долю последней причины приходилось более половины всех ошибочных результатов идентификации и 28,52±1,27 % от общего количества полученных диагностических заключений.

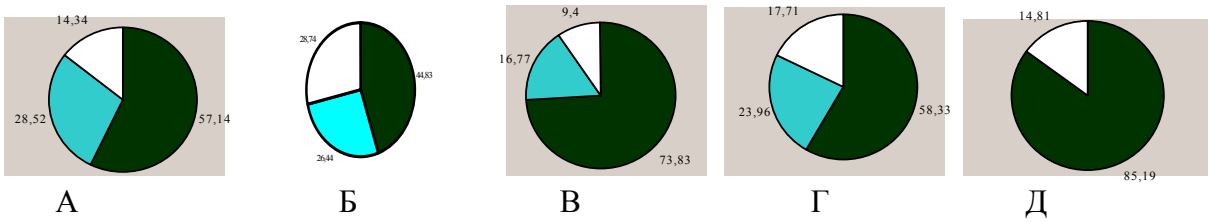

Рис.3. Диагностическая эффективность тест-систем для видовой идентификации стафилококков: А- ПБДС; Б – «STAPHYtest-8»; В- «STAPHYtest-16»; Г – API-Staph; Д - ID32.

Следующим объектом исследования стали диагностические тест-системы для идентификации стафилококков, выпускаемых фирмой «Lachema» (Чехия) под коммерческим названием «STAPHYtest» Данная тест-система достаточно широко представлена на российском рынке лабораторного оборудования и материалов, что объясняется традициями ее использования еще со времен СССР. Следует отметить, что за этот период произошла

определенная эволюция тест-системы от использования 8-ми к использованию 16-ти дифференцирующих тестов, что представило дополнительный интерес для анализа в плане получения представлений о результативности подобной модификации.

Проведенные исследования позволили констатировать, что результаты идентификации стафилококков по трехзначному цифровому коду, предусмотренному при использовании тест-системы «STAPHYtest-8», совпали с результатами, полученными с использованием референс-метода, лишь в 44,83±5,33% случаев (рисунок 3.Б). При этом причины неверной идентификации вновь могли быть сведены к двум основным моментам. Ошибки первого рода были связаны с тем, что несмотря на заявленные диагностические возможности тест-системы как достаточные для идентификации определенного спектра видов стафилококков, полученный результат все же оказывался неверным (26,43±4,72 % от общего количества диагностических заключений). Ошибки же второго рода были связаны с тем, что диагностические возможности данной тест-системы принципиально не позволяли получить верное заключение, так как она изначально предусматривала возможность идентификации только 13 из 36 известных видов стафилококков (10,34±3,2% от общего количества диагностических заключений). Последняя причина являлась ведущей – на ее долю приходилось более половины всех ошибочных результатов идентификации и 28,74±4,85% от общего количества полученных диагностических заключений.

Совершенствование данной системы с удвоением перечня идентифицирующих тестов («STAPHYtest-16») результировалось в значительном повышении ее диагностической эффективности, выразившемся в совпадении с результатами, полученными с использованием универсальной компьютерной программы, в 73,83±3,6% случаев (рис. 3.В). Природа возникающих ошибок и в этом случае была аналогична таковым у описанных выше тест-систем и преимущественно связана с тем, что при заявленных диагностических возможностях как достаточных для выявления 31 вида и подвида стафилококков, полученный результат при их идентификации оказывался неверным (16,78±3,06 % от общего количества диагностических заключений). На долю же причины, определяемой изначальной невозможностью идентификации оставшихся 5 видов и подвидов стафилококков приходилось 9,40±2,39% от общего количества полученных диагностических заключений.

Результаты идентификации изученной выборки стафилококков по совокупности тестов, предусмотренных при использовании системы API-Staph, позволили констатировать совпадение с результатами, полученными с использованием референс-метода только в 58,33±5,05% случаев (рисунок 3.Г), что несколько выше аналогичного показателя у ПБДС, но существенно ниже, чем у STAPHYtest-16, несмотря на большее количество используемых дифференцирующих тестов. При этом ведущей причиной возникновения ошибочных результатов при использовании тест-системы API-Staph (17,71±3,89 % от общего количества диагностических заключений) являлась изначальная невозможность идентификации не входящих в заявленный перечень 16 видов и подвидов стафилококков.

Результаты идентификации изученной выборки стафилококков по совокупности тестов, предусмотренных при использовании системы 32 ID, совпали с результатами, полученными с использованием референс-метода в 85,0±6,87% случаев (рис. 3.Д), что является самым высоким показателем из всех исследованных тест-систем. Причины же неверной идентификации (15,0 ± 6,87% от общего числа исследованных штаммов) в данном случае были связаны только с недостаточностью заявленных диагностических возможностей тест-системы 32 ID для идентификации определенного спектра видов и подвидов стафилококков.

Таким образом, у пяти сравниваемых коммерческих тест-систем для идентификации стафилококков была продемонстрирована различная диагностическая эффективность, неоднозначно связанная с количеством используемых в них дифференцирующих тестов. В этой связи на следующем этапе работы с целью анализа причин, определивших границы идентификационных возможностей сравниваемых тест-систем, был осуществлен анализ

информативности входящих в них идентификационных тестов, а также отклонений в их интерпретации от значений, приведенных в «Определителе бактерий Берджи».
При проведении подобного исследования информативность каждого из конвенциальных таксономически значимых признаков стафилококков была рассчитана и выражена в числовой форме (т.н. «мера Кульбака»), каковая оказалась максимальной для признака образования кислоты в аэробных условиях из маннозы (I=30,08), а минимальной для способности к росту в аэробных условиях (I=0,001). При этом в группе из 44 конвенциональных тестов, используемых для целей межвидовой дифференциации стафилококков, были выявлены 20 признаков с максимальной информативностью (I>20), в наибольшей степени пригодные для решения соответствующих задач.

В состав системы ПБДС входили тесты с достаточно низкой информативностью, что привело к относительно невысокой суммарной информативности данной тест-системы (ΣI=346,40) и результировалось в ее минимальной диагностической эффективности (57,14±1,06% правильных результатов идентификации).

Входящие в состав панели «STAPHYtest-8» тесты имели достаточно высокую информативность, однако их небольшое количество определило суммарную информативность всей тест-системы на уровне всего ΣI=171,68, что также результировалось в ее невысокой диагностической эффективности.

На этом фоне суммарная информативность тест-системы «STAPHYtest-16» составила уже ΣI=402,31, что закономерно привело к почти двукратному повышению ее диагностической эффективности до уровня 73,83±3,6%.

Суммарная информативность системы API-Staph (ΣI=365,95) оказалась выше чем у ПБДС, но ниже чем у «STAPHYtest-16», что соответствует указанным выше промежуточным значениям ее диагностической эффективности.

Наконец, в состав панели 32ID входили 25 тестов, основная масса которых является высокоинформативными, что среди сравниваемых тест-систем позволяет констатировать ее наибольшую суммарную информативность(ΣI=507,06) и соответствующую ей диагностическую эффективность (85,0±6,87% правильных результатов идентификации).

Таким образом, проведенный анализ позволил сделать предположение о связи диагностической эффективности тест-системы с суммарной информативностью входящих в нее дифференцирующих тестов, каковая была выражена логарифмической зависимостью вида ДЭ=90,99lg(ΣI%)–88,23; где ДЭ – диагностическая эффективность (в %), ΣI% – суммарная информативность диагностической панели в процентах от суммарной информативности всех 44 конвенциональных тестов.

Информативность признаков, используемых в тест-системах для межвидовой дифференциации стафилококков.

Дифференцирующие признаки	Информативность	ПБДС	STAPHY test-8	STAPHY test-16	API-Staph	32ID
Размер колоний (крупный)	27,24	-	-	-	-	-
Образование пигмента	19,09	-	-	-	-	-
Рост в анаэробныхусловиях	17,04	-	-	-	-	-
Рост в аэробных условиях	0,001	-	-	-	-	-
Плазмокоагулаза	18,25	+	-	-	-	-
Клампинг-фактор	12,39	-	-	-	-	-
Термостабильная нуклеаза	20,80	-	-	-	-	-
Гемолизины	21,07	-	-	-	-	-
Каталаза	7,50	-	+	-	-	-
Оксидаза	14,64	-	-	+	-	-
Щелочная фосфатаза	29,75	+	+	+	+	+
Аргининариламидаза	9,69	-	-	-	-	+
Аргининдегидролаза	25,38	+	-	+	+	+
Пирролидониларилами дазa	21,73	-	-	+	-	+
Орнитиндекарбоксилаза	10,53	-	-	+	-	+
Уреаза	27,47	+	+	+	+	+
β-глюкозидаза	21,65	-	-	-	+	-
β- глюкуронидаза	21,40	-	-	+	-	+
β- галактозидаза	24,04	-	-	+	-	+
Продукция ацетоина	25,87	+	+	+	+	+
Восстановление нитратов	19,39	+	+	+	+	+
Гидролиз эскулина	17,11	-	-	+	-	+
Устойчивость к новобиоцину	29,66	-	-	+	-	+
Устойчивость к полимиксину	18,79	-	-	-	-	-
Арабиноза	14,48	+	-	-	-	+
Галактоза	13,29	+	-	-	-	-
Ксилоза	20,16	+	-	+	+	-
Ксилитол	6,15	-	-	-	+	-
Лактоза	18,11	+	-	-	+	+
Мальтоза	23,40	+	-	+	+	+
Маннитол	20,67	+	+	+	+	+
Манноза	30,08	+	-	+	+	+
Мелезитоза	16,52	-	-	-	+	-
Раффиноза	13,27	-	-	-	+	+
Рибоза	20,74	-	-	-	-	+
Салицин	13,71	+	-	-	-	-
Сахароза	19,58	+	+	+	+	+
Трегалоза	21,45	+	+	+	+	+
Тураноза	22,15	-	-	-	-	+
Фруктоза	5,36	+	-	-	+	+
Целлобиоза	14,06	-	-	-	-	+
N – ацетилглюкозамин	21,69	-	-	-	+	+
Суммарная информативность (ΣI)	775,33	346,40	171,68	402,31	365,95	507,06

Дополнительной причиной возникновения диагностических ошибок, выявленной при проведении настоящего исследования, явились отклонения в интерпретации результатов дифференцирующих тестов от таковых, приведенных в «Определителе бактерий Берджи». При этом наибольшее количество подобных отклонений, обнаружено у тест-системы 32ID (105 значений), а наименьшее у ПБДС (14 значений). Соответственно, при определении связи между анализируемыми параметрами была выявлена линейная зависимость вида ДЭ=0,7845*(%отк.)+5,36, где ДЭ – диагностическая эффективность (в %); %отк – процент отклонений от значений, приведенных в «Определителе бактерий Берджи».

Найденное на этой основе уравнение множественной регрессии (рис.4) имело вид: ДЭ=0,8536*(ΣI%)+0,17*(%отк)+2,9863 и при расчете через коэффициент детерминации свидетельствовало о том, что диагностическая эффективность изученных тест-систем для идентификации стафилококков на 75% определяется суммарной информативностью входящих в них дифференцирующих тестов, а на 20% – осутствием отклонений от значений, приведенных в «Определителе бактерий Берджи». В целом же анализ выявленных зависимостей позволил констатировать, что от современных коммерческих тест-систем с ограниченным количеством идентификационных тестов ожидать диагностической точности на уровне 95-99% вряд ли представляется возможным. Достижение таких высоких результатов возможно только с использованием референс-метода, доступного лишь для узкого круга специализированных лабораторий. В то же время, в условиях увеличения перечня идентифицируемых видов стафилококков, постоянный мониторинг информативности тестов, используемых для их межвидовой дифференциации, должен стать эффективным инструментом для определения оптимального перечня тестов, позволяющего при минимизации затрат получить максимально возможный диагностический эффект.

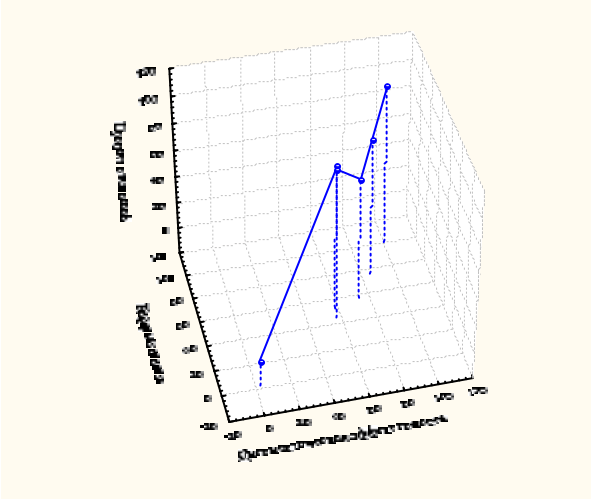

Рис.4. Графическое выражение зависимости диагностической эффективности тест-системы от суммарной информативности входящих в нее тестов и процента отклонений от значений, приведенных в «Определителе бактерий Берджи».

Таким образом, решение задачи компьютеризации видовой идентификации микроорганизмов позволило не только автоматизировать процедуру обработки и анализа результатов тестирования, но и обосновать подходы к оптимизации перечня дифференцирующих тестов, используемых в составе коммерческих тест-систем для идентификации микроорганизмов, а также определить иные направления их совершенствования.

Литература

1. *Айвазян С.А., Мхитарян В.С.* Прикладная статистика и основы эконометрики", - Москва: ЮНИТИ, 1998 г.;
2. *Дерябин Д.Г.* Стафилококки: экология и патогенность. – Екатеринбург: УрО РАН, 2000. – 239с.;
3. *Дубров А.М., Мхитарян В.С., Трошин Л.И.* Многомерные статистические методы – Москва: "Финансы и статистика", 2000 г.;
4. Определитель бактерий Берджи. В 2-х т. Пер с англ. / Под ред. Дж.Хоума, Н.Крига, П.Снита и др. – М.: Мир, 1997;
5. *Нехорошева А.Г., Скала Л.З., Поликарпова С.В., Винокуров А.Е., Скаженик В.Ю.,* Использование усовершенствованных микро-ЛА-тестов для идентификации микроорганизмов различных групп в клинической микробиологии // Клин. лаб. диагностика. – 2000. - №3. – С.51-54.
6. *Савицкая К.Н., Черненькая Т.В.* Использование микротест систем для идентификации изолированных клинических штаммов// Клиническая лабораторная диагностика. – 1996 - №5 – с.29-35;
7. *Фот Н.П.* Возможность использования статистических методов при идентификации стафилококков// Учебная научно-производственная инновационная деятельность высшей школы в современных условиях - Международная юбилейная научно-практическая конференция, посвященная 30-летию ОГУ. – 2001. - С. 60;
8. *Фот Н.П., Бравичева О.С.* Видовая идентификация микроорганизмов рода стафилококков с применением методов кластер-анализа // Вестник ОГУ. – 2002. - №3 – С.132-135.;
9. *Gahrn-Hansen B., Heltberg O., Rosdahl V.T., Sogaard P.* Evaluation of conventional routine method for identification of clinical isolates of coagulase-negative Staphylococcus and Micrococcus species. Comparison with API-Saph and API-Staph-Ident // Acta Pathol. Microbiol. Immunol. Scand. [B]. – 1987. – V.95. – N5. – P.283-292.;
10. *Gemmell C.G., Dawson J.E.* Identification of coagulase-negative staphylococci with the API-Staph system // J.Clin.Microbiol. – 1982. – V.16. – N5. – P.874-877.;
11. *Grant C.E., Sewell D.L., Pfaller M. Et al.* Evaluation of two commercial systems for identification of coagulase-negative staphylococci to species level // Diagn. Microbiol. Infect. Dis. – 1994. – V.18. – N1. - P.1-5.;
12. *Ieven M., Verhoeven J., Pattyn SR., Goossens H.* Rapid and economical method for species identification of clinically significant coagulase-negative staphylococci// J-Clin-Microbiol. 1995 May, 33(5): 1060-3;
13. *Perl T.M., Rhomberg P.R., Bale M.J. et al.* Comparison of identification system for Staphylococcus epidermidis and other coagulase-negative Staphylococcus species // Diagn. Microbiol. Infect. Dis. – 1994. – V.18. – N3. – P.151-155.;
14. *Petras P. A* ORIDES – orientational identification of coagulase-negative staphylococci from human clinical materialu// Epidemiol-Microbiol-Imunol. 1995 March; 44(1): 15-9;
15. *Polian D., Zeman J.* Discrimination value of test performed with STAPHYtest Lachema from the aspect of clinical microbiology// Cesk. Epidermid., Imunol. – 1994 –V. 43 – p. 26-28;
16. *Renneberg J., Rieneck K., Gutschik E.* Evaluation of Staph ID 32 system and Staph – Zym system for identification of coagulase-negative staphylococci // J-Clinic-Microbiol. 1995 May; 33(5): 1150-3;
17. *Sedlacek I., Kocur M.* Identification of Staphylococcus and Micrococcus species with the STAPHYtest system// Folia Microbiol (Praha). – 1991. - V.36.-p.401-405.