

Использование многопараметрических методов анализа информации в онкологии.

Боженко В.К. , Сотников В.М., ФГУ «Российский научный центр рентгенорадиологии федерального агентства по здравоохранению и социальному развитию Российской Федерации» (г. Москва).

Аннотация

Обзор применения многопараметрических методов анализа информации в онкологии для получения диагностической и прогностической информации.

Ключевые слова: многопараметрический анализ, математические методы обработки информации, нейронные сети, статистика.

Clinical evaluation of multivariate analysis in oncology.

Bojenko V.K., Sotnikov V.M.

Summary

The review of the multivariate method analysis medical information in oncology.

Оглавление:

Введение

Математические методы анализа используемые в клинической онкологии

Проблемы размерности

Использование многомерных методов анализа для получения прогностической информации

Использование моделей регрессии в медицинской практике

Метод искусственных нейронных сетей

Список литературы

Введение

В большинстве клинических ситуаций одновременно действует множество факторов (параметров). Между ними существуют сложные взаимные влияния. Общий эффект от двух факторов может превышать их индивидуальную сумму. Поэтому важным аспектом получения прогностических и диагностических данных в последнее время стали достижения методов обработки информации. Специалисты в области статистики отмечают, что многие методы обработки информации в этой области возникли и получили дальнейшее развитие благодаря задачам, поставленным медиками и биологами (Славин, 1989, Кант, 1987). Методы многопараметрического анализа и построенные на их принципе системы классификации и прогноза получили применение в кардиологии,

экстренной медицине, спортивной медицине, онкологии и др. Многие методы статистики, кластерный анализ, нейронные сети, экспертные системы, тройные матрицы, главные компоненты, неточная логика (fuzzy) возникли или развивались для решения задач медицинской классификации и прогноза (Кендалл, Стьюарт 1976, Welstead , 1994). Многофакторный анализ – единственный реальный способ обработки многих данных (переменных), когда на этапе анализа приходится иметь дело сразу со многими переменными. (Рандомизация также контролирует множество переменных, но на этапе планирования и проведения исследования (Флетчер и др. 1998). Многофакторный анализ (multivariate analysis) – это совокупность методов одновременного рассмотрения воздействия многих переменных. Они используются для того, чтобы корректировать эффекты множества исследуемых переменных для выявления независимого действия одного фактора, позволяют выделить из большого числа переменных малое их подмножество, которое вносит независимый и существенный вклад в исход, и упорядочить переменные по силе их влияния на исход. Прогресс в области медицины и физиологии связан с новыми возможностями компьютерных исследовательских систем, поскольку они являются не только средствами статистической обработки данных, но и инструментом медико-биологического познания (Генкин ,1999). Смена парадигмы обработки клиничко-лабораторных данных на парадигму их глубокого многостороннего анализа – одна из основных задач современной медицинской информатики (Эмануэль, Генкин, 1997). Фактически ставится задача автоматизированного извлечения нового медико-биологического знания из эмпирических данных и предоставления врачу-клиницисту возможность использовать его для решения практических задач (Плескун и др. 1979, Гланц, 1978, Дубров и др. 1978, медик и др. 2001). Похожие тенденции имеются и в лабораторной медицине. "Доминантой в оценках прогресса в лабораторной медицине служит не столько степень улучшения аналитических возможностей, хотя без него нет и самого прогресса, сколько совершенствование вклада лабораторной информации в обоснование диагностических и лечебных решений. " (Меньшиков, 2003).

Одной из предпосылок для нового подхода в получении диагностических данных стало распространение систем регистрации, хранения и представления информации о пациенте в виде электронных историй болезни – образующих, фактически, централизованные базы данных и позволяющих накапливать и проводить целенаправленный их анализ (Сусликов, 1999, Sim et al. 20001). Другой важный аспект, позволяющий использовать новые подходы в анализе клинических данных – стандартизация получаемой

информации (Медицинская лабораторная диагностика, 1997). Это касается как процессов формализации медико-биологических данных, заносимых в историю болезни, их соответствие общепринятым Российским и международным стандартам, так и процессов стандартизации методов получения медико-биологической информации. Это включает в себя такие процедуры как внутри лабораторный и внешний контроль качества, процедуры регулярного внешнего аудита и др. Не последнее значение в качестве получаемой информации имеют и средства диагностики. Использование современных технологий существенно повышает достоверность получаемых данных и создает предпосылки извлечения новой диагностической информации, ранее недоступной из-за методических и методологических проблем. Внедрение единых подходов к компьютерному ведению заболевания прежде всего выявляет их преимущества для больного (Панкратова и др., 1997)

Планирование исследования и адекватный анализ полученных данных составляет неперемную часть научной деятельности, а статистические методы являются математическим аппаратом, применяемым для этого. Достижения в области технического обеспечения (на компьютерах первых поколений некоторые задачи с применением итерационного алгоритма могли решаться несколько дней) и создание современных математических и статистических пакетов, оснащенных большинством новейших методов обработки, позволили значительно расширить области применения и внедрения этих методов в медицине. Одним из главных направлений работы исследователей стало не только исследование информационной ценности новых методов диагностики и лечения, но возможность по-новому оценить информативность традиционных (рутинных) показателей. В современных исследованиях в начальный список изучаемых признаков входят практически все клинические и лабораторные показатели, доступные исследователю. Так, например, в работе были исследованы более 200 клинических и лабораторных показателей для оценки их прогностической важности при раке легкого (Buccheri, Ferrigno, 1994). Неудивительно, что при таком объеме информации на первое место стали выходить методы, применимые для обработки многопараметрических данных. Даже такой традиционный метод, как критерий Стьюдента – метод сравнения достоверности отличия средних, из-за этих отличий (особенностей) уступает место методу анализа дисперсий, позволяющему работать с многопараметрическими таблицами (MANOVA) и в большинстве зарубежных журналов, результаты, обработанные без использования этого метода, не принимаются к публикации.

Математические методы анализа используемые в клинической онкологии

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) (Боровиков, 1997, 2001). 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 (Леман, 1979, Медик и др., 2001). Для выявления возможной связи между параметрами применяется также и дисперсионный анализ, позволяющий не только оценить групповые отличия, но и обнаружить связи между переменными различной природы (непрерывными, порядковыми или категориальными).

Дополнительная задача корреляционного анализа (являющаяся основной в регрессионном анализе) состоит в оценке уравнений регрессии, где в качестве результативного признака выступает признак, являющийся следствием других признаков (факторов) – причин. Причинно-следственная связь устанавливается из вне статистических соображений, например из аргументов, касающихся физической (физиологической) природы явлений (Боровиков, Ивченко, 1999).

Проблемы размерности

Рассматривая применение методов многомерного анализа в медицине (прогнозирования) в медицине необходимо отметить ряд принципиальных ограничений имеющих у них. Т.к. основные принципы в методах прогноза основаны на методах линеаризации, то для них свойственна чувствительность к размерности выборки (количеству исследуемых объектов, которое зависит уже не только от дисперсии анализируемого признака, как это было при сравнении средних). Линейные методы аппроксимации (регрессии, прогноза) требуют, чтобы на каждый анализируемый признак (переменную) в группе было около 10 объектов (в

случае клинических исследований – пациентов). Т.е. если исследуется 10 показателей, например 10 биохимических показателей крови, то в группе должно быть 100 пациентов (Бейли, 1970, Боровиков, Боровиков, 1997, Denne, 2001). Это ограничение является принципиальным не только для методов линейной регрессии (и других методов линеаризации), но и таких методов? как дискриминантный и факторный анализ. Хотя в ряде работ говорится, что кластерный анализ менее чувствителен к размерности выборки, при его использовании также надо учитывать это ограничение (Боровиков, Ивченко, 1999, Гланц, 1999, Gould, 2001).

Внедрение в медицину современных молекулярных технологий также сталкивается с проблемой размерности. Например, одно из важных применений метода анализа профиля экспрессии генов с использованием микрочипов – создание новых подходов в классификации опухолей (Bustin, Draidí, 1998, Golub, 2000, Park et al., 2002). Возможность измерять большое количество генов приводит к тому, что количество переменных (генов) значительно превосходит количество образцов. Т.е. наступает условие, когда стандартные статистические методы классификации не работают или имеют существенные ограничения. Т.е. возникает необходимость разработки новых методов для анализа таких данных при решении задач классификации и предсказания.

Проблемы такого рода называются проблемами размерности и могут быть решены с помощью специальных методов. К таким методам можно отнести методы факторного анализа, многомерного шкалирования, генетического алгоритма – одного из алгоритмов, используемых в методе искусственных нейронных сетей.

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 (например, до 2). Обычно целью факторного анализа является попытка качественно описать большую совокупность имеющихся параметров. Как бы "суммировать" информацию и получить факторы, имеющие некий биологический смысл. Интересной особенностью факторного анализа является его "непредвзятость" с точки зрения анализа объектов, т.к. фактически основной анализ производится над корреляционной матрицей, не включающей информацию о каждом объекте.

Это важно, если исследователь хочет использовать собственные значения факторов для дальнейшего анализа методами классификации. Использование таких подходов и помогает решать современные задачи, имеющие проблемы размерности (например, в современной молекулярной биологии).

Описанные методы применялись для классификации образцов опухолей (варианты лейкоза и лимфом) на основании профиля экспрессии генов (Nguyen, Rocke, 2002). Полученные результаты имели стабильное решение и работали в условиях рандомизации образцов. Сравнение различных статистических методов для анализа такого рода данных также анализируются в работах (Weitzman 2000, Raychaudhuri et al., 2001, Clare, King 2001, Olshen, Jain 2002).

Рассматриваются результаты полученные с применением кластерного анализа, нейронных сетей, вариантов дискриминантного анализа, метода главных компонент и факторного анализа. Показано, что стандартные методы классификации могут удовлетворительно решать подобные задачи.

Использование многомерных методов анализа для получения прогностической информации

Подходы для получения прогностической информации могут существенно отличаться как по набору анализируемой информации, так по способу получения окончательной информации. Они могут быть организованы в виде таблиц, опросников, простых или сложно организованных компьютерных программ (Сусликов 1999, Batchelor, Dua, 1995, Мэйндоналд 1988). Большинство распространенных индексов, пытается ориентироваться на широкий круг потребителей, который может включать кроме врачей также и пациентов.

Распространенный вариант построения прогностического признака – таблица или текст, включающие соответствие набора качественно описанных свойств с определенной количественной величиной – примером таких прогностических величин может быть индекс физического состояния пациента – индекс Карновского, оценка этого же свойства на основании классификации ECOG.

В последние годы качество жизни больного и экономические вопросы, связанные с лечением становятся такими же важными как и продолжительность жизни и считаются дополнительными клиническими критериями при лечении онкологического заболевания. В настоящее время стоят очень сложные методологические и статистические проблемы сравнения полученных данных и оценки их предсказательной мощности сделанные на основании используемых в настоящее время способов измерения качества жизни (QoL) для различных контингентов пациентов.

Нерешенными являются вопросы интерпретации «величины» регистрируемой по разным шкалам и статистической значимости регистрируемых различий. Высокая субъективность оценки состояния больного по анкетам приводит к низкой достоверности получаемых результатов. В исследовании Национального противоракового института Канады (National Cancer Institute of Canada) приводятся результаты сравнения оценки функционального состояния пациентов на основании различных шкал

(Physical functioning item of the subjective significance scale and the physical functioning scale of the QLQ-C30). Полученный коэффициент корреляции между ними имел величину 0.23 и являлся совершенно неудовлетворительным (Special Paper Methodological and Statistical Issues of Quality of Life (QoL) 1998). Приведенные результаты иллюстрируют существующую проблему объективности информации при использовании анкетных (или табличных) методологий получения количественных оценок.

Решение похожих задач, в других областях медицины (оценка общего состояния больного и прогноз заболевания) достигается с использованием несколько других подходов. Хотя окончательные оценки также получаются на основании табличных данных, исходная информация включает существенно большее количество показателей, основанных большом количестве функциональных и лабораторных исследований. Такие индексы прогноза широко распространены, например, в медицине неотложных состояний (APACHE, TISS, ASA, CRIB). При этом индексы могут быть ориентированы на конкретный тип травмы (например черепно-мозговой) (Зильбер, 1995, Knaus et al. 1991, Cockburn et al. 1993, Cullen et al. 1974). Эти индексы имеют высокую степень предсказательной способности. Например, APACHE позволяет с вероятностью более 95% предсказать исход травмы.

Возвращаясь к способам оценки физического состояния онкологического больного на основании индексов Карновского или ECOG (и других подобных), следует подчеркнуть, что, несмотря на всю их простоту, они продолжают оставаться адекватным инструментом в руках врача. В большом количестве работ, посвященных прогностической значимости различных признаков у онкологических больных они остаются одними из наиболее информативных (Ship et al. 1993, Mendenhall et al. 1994, Maritinow et al. 1998, Mori et al. 1999, Montoto et al. 2002, Bierman et al., 2002). Таким образом, индексы, основанные на несложных математических принципах, занимают важное практическое место в клинической практике и, в тоже время, имеющие методологические проблемы при их определении ставят задачи по повышению объективизации способов их регистрации. На перспективность

использования просто организованных интегральных показателей крови указывают авторы работы (Бельских и др. 1996). Они использовали два варианта такой оценки: лейкоцитарный индекс интоксикации Кальф-Калифа (Кальф-Калиф, 1941) и показатель, названный авторами – энтропия лейкоцитарной формулы (Номограммы для определения некоторых интегральных показателей крови человека / Борисов и др., 1989). Оба показателя оказались информативными в оценке прогноза развития патологического процесса. Авторы делают вывод, что использование интегральных показателей крови в диагностическом мониторинге позволяет повысить информативность традиционных диагностических тестов, улучшить оценку эффективности терапии, наметить пути прогнозирования исхода и индивидуализации лечения. Использование интегральных показателей крови перспективно при компьютерной оценке результатов лабораторных исследований и разработке экспертных диагностических систем.

Другой, наиболее распространенный способ получения прогностической информации можно проиллюстрировать на примере программы, предложенной (Valet et al., 1984, Valet, Hoffkes, 1997, Valet et al. 1998). Авторы предлагают систему классификации - SMDC – standardized Multiparameter Data Classification.

Программа классификатор CLASSIF1 использующая стандартизованные, инструментальные и лабораторные данные цитометрии и другие многопараметрические данные, включая клиническую химию и биомедицинские и клинические данные открыта для использования через интернет по протоколу TELNET. Любой желающий может использовать свои данные поиска решения своей задачи. В авторском варианте программа используется для решения таких задач как: оценка риска инфаркта миокарда на основании данных экспрессии активационных антигенов тромбоцитов (используемые данные для классификации CD62, CD63, тромбоспондин, иммуноглобулин G); оценка риска перетренировки для профессиональных велосипедистов на основании уровня лимфоцитарных антигенов; прогноз выживаемости у больных меланомой (делается дискриминация на два класса до 10 лет и более 10 лет) на основании клинических данных. Принцип алгоритма построен на вычислении пары перцентилей 5/95%, 10/90%, 15/85% на основании кривой распределения значений в обучающей выборке. (Какую пару перцентилей использовать решается на практике.) Далее, конкретные значения заменяются на: 0- если оно внутри выбранного интервала, (+) - для значения больше и (-) – для значения меньше выбранных перцентилей. Далее, на основании уже известных клинических данных формируется

матрица – сколько каких комбинаций признаков соответствует какой патологии.

Клин. \ лабораторная классиф.	Данные	Всего	N	R	D
Норма, N		17	100	5.9	0
Риск инфаркта, R		97	2.1	96.9	3.1
Диаб. Тромбоз, D		14	57.1	7.1	42.9

В клетках показан процент пациентов попавших в данную категорию для выбранного уровня перцентилей. Процесс оптимизации классификации – увеличение суммы значений диагональной компоненты матрицы – 100% классификация для каждого класса. Далее, единичные колонки (признаки) или их комбинации исключаются и определяется результат классификации. Если он не ухудшается, в конечный алгоритм эти данные не включаются.

Данная классифицирующая система больше зависит от внутри лабораторной точности, чем от межлабораторной правильности. Инструментальная точность может быть скорректирована (инструментальную ошибку можно исключить) нормализовав перцентили на среднее значение референтных образцов (в качестве которых могут быть образцы крови доноров). К тому же, биохимические свойства многих клеточных систем в течение болезни можно немедленно сравнить, когда классификаторы стандартизованы на общие, доступные клетки, такие как клетки периферической крови.

Похожую идеологию получения диагностической информации использовали отечественные авторы создания "информационной технологии анализа медицинских данных" (Генкин 1993, 1996, 1999, Генкин, Степанова 1997). Их подход также основывается на создании интервалов для каждого из исследованных клинико-лабораторных данных, однако используются не стандартные статистические интервалы, а их значения подбираются таким образом, чтобы частоты попадания объектов, качественно различаемых экспертом максимально отличались. Таким образом, формируются интервальные оценки с соответствующими вероятностями попаданию объектов в определенные интервалы. Если теперь перейти к двум признаком, то образуется набор прямоугольников – со сторонами, соответствующими интервалам разбиения для каждого признака. Множество таких прямоугольников, вместе с оценкой вероятности попадания в соответствующие интервалы авторы назвали "бинарными" структурами. Важной особенностью предлагаемого метода анализа данных, подчеркиваемой авторами является возможность "выявлять диагностическую и прогностическую ценность и для значений, находящихся в пределах

нормы". Это также позволяет открывать новые интервалы, которые прежде в клинической практике не рассматривались, как не несущие информации. Это очень важное свойство, на которое обратили внимание авторы. Очень часто, использование "общепринятых" норм не позволяет извлечь важную диагностическую информацию из полученных экспериментальных и клинических данных. Само понятие "нормы" и "лабораторной нормы" – чрезвычайно сложно и требует осторожного обращения в том смысле, что анализ данных в любом виде исследований, по-видимому, необходимо начинать исключая интервальные "нормальные" ограничения и проводить статистический анализ в конкретных, исследуемых группах (Генкин 1998). Полученные результаты затем качественно можно сравнивать с имеющимися интервалами нормы. (Если это необходимо). Важной особенностью предлагаемого анализа – это возможность исследования зависимости признаков, совпадающего, по существу, с методом анализа матриц сопряженности. Получаемая таким образом информация позволила авторам решить ряд интересных диагностических задач. Были исследованы особенности гемопоэза и метаболизма у больных лимфогранулематозом, а также показана принципиальная возможность дифференциальной диагностики между лимфогранулематозом и другими лимфопролиферативными заболеваниями на основании лабораторных показателей крови. (Генкин, Степанова, 1997, Генкин, Корицова, Степанова, 1997, Генкин 1998).

Принципиально другой подход для построения системы получения информации описан в работе (Robinson, Bobcock, 1989). Авторами рассматривалась вероятность заболевания СПИДом у пациентов с ВИЧ инфекцией. Если у ВИЧ инфицированного пациента развился СПИД, то прогноз неблагоприятен, время выживания короткое. Однако выяснилось, что часть пациентов живет значительно дольше других. Было проведено исследование, какие характеристики пациентов влияют на выживаемость. С выживаемостью оказалось связано несколько физиологических характеристик. Используя сочетание этих факторов, исследователи разработали многоступенчатую прогностическую систему, оценивая в 1 балл присутствие каждого из 7 факторов: сильная диарея или содержание сывороточного альбумина менее 2 г.л, любые неврологические расстройства, насыщение крови кислородом менее 50 мм.рт.ст., гематокрит менее 30 %, число лимфоцитов менее 1500 мл⁻¹, число лейкоцитов менее 2500 мл⁻¹, число тромбоцитов менее 140 000 мл⁻¹. Были выделены три прогностические стадии в зависимости от суммы баллов: I-стадия – 0 баллов, II-стадия – 1 балл, III- стадия – 2 балла и более. При этом авторы отмечают, что медиана выживаемости в первой группе – 11,6, второй группе – 4,5, а в 3-й 2,1 месяц.

Следует отметить, что этот подход основан на интервальной оценке параметров – для формирования групп имеющих фиксированную вероятность наступления какого-либо события (это может быть рецидив заболевания или смерть больного или, например, дожитие до определенной величины). Именно такой подход был использован в многоцентровом, международном исследовании для формирования прогностического индекса у больных злокачественными лимфомами (Ship et al., 1993) и аналогичного индекса для больных лимфомой Ходжкина (Hasenclever, Diehl, 1996). На основании анализа большой группы пациентов оценена вероятность дожития до 5-летнего периода в зависимости от набора 5 прогностических факторов. Авторам удалось найти уровни факторов (интервалы их значений), при которых они вносят одинаковый вклад в модель. Таким образом, вероятность дожития определяется не значением факторов, а их количеством. Это существенно упрощает использование модели, хотя сами оценки и получены на основании использования модели логит-регрессии.

Следует отметить, что такой подход является стандартной процедурой анализа клинических данных и построения прогностических моделей (Vuccheri et al. 1994). Например в работе (Piccini et al. 1984) исследовались прогностическая значимость основных гематологических показателей у 146 пациентов с раком почки. Пациенты делились на группы 1- опухоль в пределах капсулы (66), 2- выход за капсулу (39), 3- с отдаленными метастазами (27), 27 – с кахексией и без операции. Наиболее часто встречаются: анемия, в основном нормохромная - гемоглобин менее 120 г/л (41% в 1-й группе, 59% - во второй, 74% в третьей и 79% - в 4-й); лейкоцитоз (более 8 тыс.: 38% в 1-й группе, 51% - во второй, 37% в третьей и 50% в 4-й). Тромбоцитоз средней степени (350-500 тыс. обнаруживался с заметной частотой во 2 и 3-й группах (20% и 18% соответственно). С другой стороны лейкопения (лейкоциты мене 3.5 тыс.) и тромбоцитопения (тромбоциты менее 140) встречались значительно реже. Таким образом, авторы попытались на основании клинического анализа крови оценить распространенность опухолевого процесса. Приведенные результаты показывают ту же логику поиска диагностической информации: – попытка классификации на основании попадания признаков в определенный интервал. Подобные подходы также активно используются отечественными авторами (Нечеснюк 2000, Доценко 2003). Однако большинство исследователей осознает недостаточную предсказательную мощь такого рода индексов и ищет пути увеличения их информативности (Axtdorph et al. 2000).

Далее по частоте использования следуют варианты кластерного и дискриминантного анализа. Например, авторами (Кашин, Фирсов 1983)

приводятся результаты по использованию диагностической функции вида $N = \sum k_i \cdot x_i \{x_i = 0, 1\}$ для дифференциальной диагностики туберкулеза - пневмонии - рака легкого. Были найдены коэффициенты для трех функций - соответственно трех пар заболеваний и пороговые значения для этих пар. Используются 15 показателей, среди которых 5 показателей крови. Расчет был произведен на основании анализа данных 300 историй болезни, а контрольное исследование на 50 и показано совпадение диагнозов в 84% случаев. Этот подход, можно представить как вульгаризацию дискриминантного метода. Диагностическая функция имеет вид упрощенной дискриминантной функции, в которой вместо значения параметра используется дискретное значение признака (1 - если признак больше определенного уровня или 0 если признак отсутствует). Авторы рекомендуют использовать метод на поликлиническом уровне, для постановки предварительного диагноза и решения направления больного в специализированную клинику. Дискриминантный подход при анализе многопараметрических показателей крови для диагностики в онкологии активно применяется группой исследователей из Белорусского онкологического института им Н.Н.Александрова (Машевский А.А, 1994, Доста и др. 1999, Смолякова и др. 2000, Прохорова и др. 2001). Использование дискриминантного анализа комплекса биохимических и физико-химических характеристик крови позволяет, по мнению авторов, оценивать такие важные характеристики при лечении онкологических больных как прогноз и мониторинг эффективности лечения при раке предстательной железы, прогноз осложнений при применении многокомпонентной химиотерапии у детей, оценка уровня эндогенной интоксикации при лимфогранулематозе. На эффективность использования дискриминантного анализа клинико-лабораторных данных при решении дифференциально-диагностических задач указывается в работе (Савченко и др., 2001). Одной из областей, где активно используется метод дискриминантных функций является рентгенология. Использование этого способа классификации формализованных рентгенологических данных позволяет стандартизовать получаемые диагнозы и, в некоторых случаях, повысить диагностическую эффективность по сравнению с экспертной (Brake et al. 2000, Ng et al. 2001, Jiang et al. 2001, Lo et al. 2002).

Наравне с дискриминантным анализом, активно исследуется возможность практического использования различных вариантов кластерного анализа для решения медицинских задач (Yang et al. 1994, Sim et al. 2001, Subdo et al. 2000, Pauler et al. 2002, Legler et al. 1995, Lefkopoulou, Ryan et al. 1993, Aitchison et al. 1995). Например, в монографиях (Ставицкий 1999, Ставицкий и др., 2000)

рассматривается применение метода кластерного анализа для решения задач прогноза в медицине. В качестве основных трудностей использования метода указываются: низкий уровень формализации предметной области, а также тот факт, что для получения необходимого количества информации может потребоваться высокий уровень затрат. Эти причины, по мнению авторов, ориентируют исследователя при выборе методов анализа в сторону методов классификации основанных на методах распознавания образов. Преимуществами этих методов является их ориентация на результат принятия, а не анализ процесса и механизма принятия решения. Во многих случаях такой подход упрощает поиск решения задачи. Предложенная авторами система классификации на основании гематологического анализа крови, предлагает новые подходы к применению метода распознавания образов для задач классификации. Интересен также вывод авторов о возможности снижения количества объектов в обучающей выборке. Как уже говорилось, этот показатель представляется наиболее критическим в большинстве систем и методов, решающих задачи классификации. Это ограничение следует учитывать при планировании исследований, т.к. на практике бывает трудно организовать набор необходимого количества пациентов, имеющих кроме всего прочего одинаковое качество диагностической информации.

Вариант кластерного анализа – деревья классификации (разновидность принятия решения метода графов) часто используется в областях, где информация носит качественный характер и может быть преобразована в порядковые переменные. Это, например, такие дисциплины как цитология, гистология, психология. Классификация объектов (в случае цитологии или гистологии это постановка диагноза) производится на основании последовательного принятия решений при регистрации определенного уровня признаков. Например, в работе (Sudbø et. al. 2000) применяется метод деревьев классификации для оценки степени злокачественности на основании гистологических признаков для рака языка. Авторы отмечают, что для эффективного использования метод оценки должен быть воспроизводим. Этого нельзя сказать о рутинных гистологических методах, для которых межлабораторная проверка дает большое количество расхождений. Использование математических методов классификации информации, получаемой с помощью методов анализа изображения позволило авторам найти новые свойства гистологической архитектуры, имеющую высокий коэффициент корреляции с выживаемостью. Таким образом, внедрение технологии автоматической регистрации параметров и их математический анализ, повысили правильность оценки степени злокачественности для этого варианта рака.

Хотелось бы также упомянуть группу методов, которые используют принцип получения информации об объекте (в данном случае пациенте) исследуя динамику изменения какого-либо показателя. Например, оценивая некие параметры до и после воздействия (лечения) можно исследовать как параметры самого объекта (*in vivo*), так и модельные системы (например реакция крови на облучение *in vitro*, реакция опухолевых клеток на химиопрепараты в культуре, и т.п.). Такой подход очень распространен в экспериментальной биологии, т.к. позволяет более углубленно исследовать свойства объекта. Эти методы условно можно назвать "релаксационными". Т.е. система выводится из равновесия каким либо способом и исследуются процессы возвращения ее в состояние равновесия мы получаем информацию о системе.

Например, авторы (Pauler, Laird, 2002) указывают на то, что внедрение новых лабораторных показателей в клиническую практику расширяет круг биомаркеров, динамические исследования которых (приводится пример показателя MCV) на фоне применения фармпрепаратов могут быть источником важной информации для предсказания осложнений и коррекции программы лечения. Именно математический анализ динамики изменений на этапе проведения испытаний нового препарата позволяет сделать адекватные выводы о наступлении нежелательных осложнений. В работе (Fumagalli, 2003) показано, что именно мониторинг содержания лейкоцитов в процессе лечения П-2 больных раком почки является независимым фактором прогноза как для общей выживаемости, так и для оценки ответа на проводимую терапию. Таким образом, использование динамических исследований на фоне тестирующего (примером может служить "провокационные пробы") или лечебного воздействия может дать существенно более ценную информацию для построения прогностических моделей и такого рода подходы необходимо более интенсивно внедрять в клиническую практику.

Использование моделей регрессии в медицинской практике

Необходимость и важность адекватного прогноза продолжительности жизни для онкологического пациента обусловлена рядом причин. Во-первых это необходимо для адекватного выбора программы лечения (особенно с точки зрения возможных побочных эффектов, связанных с применением высокотоксичных схем лечения), во-вторых – это может позволить правильно распределять финансовые ресурсы, направляя их более эффективно, в третьих – правильная оценка может помочь пациенту и его семье осуществить адекватно планирование на этот период. Кроме того, в ряде стран, для пациентов с прогнозируемой скорой гибелью предусмотрены

специальные финансовые и социальные программы (Health service Guidelines 95).

Построение прогноза (развитие болезни, различные исходы) (наступление события через конкретный момент времени) можно попытаться описать на основании известных, на настоящий момент, данных. Решение таких задач осуществляется в рамках различных моделей регрессии. Наиболее изученной является модель линейной регрессии (McCullagh, Nelder, 1989), однако ее использование для описания большинства, клинических важных показателей (продолжительности жизни и других событий) неприменимо из-за нелинейной природы изучаемых явлений. Неудовлетворенность линейными методами для построения предсказательных моделей привела к разработке альтернативных подходов. Например, в работе (Batchelor, Dua, 1995) показано, что комбинация линейных моделей существенно снижает ошибку оценки.

Поскольку наиболее важным показателем при сравнении различных схем лечения является продолжительность жизни, способы ее описания и сравнения играют принципиальную роль. Наиболее важны следующие семейства распределений, которые используются для описания продолжительности жизни экспоненциальное (в том числе, линейное экспоненциальное) распределение, распределение Вейбулла экстремальных значений и распределение Гомперца. В принципе, для описания средних времен жизни и сравнения нового метода лечения, со старыми, можно было бы использовать стандартные параметрические и непараметрические методы. (Следует сказать, что, Наиболее простым способом описания выживаемости в выборке является построение таблиц времен жизни.) Техника таблиц времен жизни - один из старейших методов анализа данных о выживаемости. Таковую таблицу можно рассматривать как "расширенную" таблицу частот. Построение таблиц времен жизни, оценивание функции выживания с помощью процедуры Каплана-Мейера являются описательными методами исследования цензурированных данных (Гланц, 1999).

Основные статистические методы для сравнения схем лечения или оценки прогностической значимости различных признаков используемые в настоящее время в онкологии это регрессионная модель Кокса, экспоненциальная модель и модель логнормальной регрессии (Cox, 1972, Cox, Oakes, 1984, de Bruijene et al., 2001). Регрессионная модель Кокса основана на предположении о лог-линейной зависимости между функцией описывающей параметры (например выживаемость) и ковариатами, а также на том, что эта зависимость мультипликативна (последнее предположение называется также предположением пропорциональности).

В работе (Mori et al., 1999) дан пример современной технологии применения регрессионных моделей для построения прогностических критериев. Основной задачей авторов было исследование влияния возраста на выживаемость у больных злокачественными неходжкинскими лимфомами. Известно, что возраст является одним из важнейших независимых признаков, связанных с плохим прогнозом у пациентов со злокачественной неходжкинской лимфомой. В тоже время, количество пожилых людей страдающих этой патологией постоянно увеличивается. Поэтому, выделение среди этих больных группы с хорошим прогнозом (почему-то написано так) прогнозом, для проведения более агрессивной терапии очень важно. В данном исследовании была предпринята попытка разработать модель для предсказания исхода у пожилых пациентов до предпринятого лечения. Для этого в предварительном исследовании, 443 пациента с НХЛ старше 65 лет были проспективно рандомизированы и пролечены по схемам T-COP (T, pirarubicin), СНОР или T-COEP (E;VP16) в сроки между 1990 и 1992 гг и 425 больных старше 70 были проспективно рандомизированы и пролечены по схемам T-COP с высокой стандартной или низкой дозой (между 1993-97 гг) в мультицентрическом исследовании. Все 861 имели возраст между 65-92 с медианой 75. Эффективность полной ремиссии была 48,2%. Прогностическими факторами для полной ремиссии были, (по убыванию важности) физическое состояние больного (performance status PS), клиническая стадия, ЛДГ, и уровень альбумина ($P < 0.01$). Прогностическими факторами для общей выживаемости были (over all survival), PS (0 или 1 если больше), стадия (I или II против III или IV), ЛДГ (нормальное или повышенное), альбумин (до или более 3,5), гистологическая классификация (низкая, промежуточная (intermediate), высокая) и тип (B или T) (Вилкоксон $P < 0.01$). Произвели анализ по модели Кокса. Четыре параметра были идентифицированы как прогностические для выживаемости: PS, клиническая стадия, ЛДГ, альбумин. Никакие другие не имели независимой прогностической значимости, включая экстранодальное вовлечение ($p = 0.147$). Каждому из этих показателей присваивалось значение 1 и общее количество баллов считалось. Пациенты были разбиты на три группы риска: низкий риск (сумма 0-1), промежуточный (2), 3 или 4 – высокий риск. Безрецидивная общая выживаемость для этих трех групп составляли соответственно: 59.2%, 45.7% или 37.6%, и 43%, 22% или 18%. Пример, приведенный выше описывает классический путь поиска прогностических факторов с использованием стандартных средств статистики, принятых для регрессии параметров, имеющих сложную временную зависимость от большого количества признаков. Полученная регрессионная модель имеет сложное математическое описание и далее, для

более наглядного описания, выбираются диапазоны параметров, формируются группы, определяются значения искомым параметров (выживаемость в данном случае) имеющие значения внутри диапазонов. В данном примере это три группы, но приведенные значения выживаемости, как безрецидивной, так и общей для промежуточной и высокой группы риска имеют очень близкие значения, которые, скорее всего, лежат внутри ошибки определения, а полученные отличия в 22 и 18 могут считаться не значимыми. Таким образом, предлагаемое разделение на группы риска можно считать не целесообразным и ограничиться, в данном случае только двумя. Это, скорее всего, позволит еще более сократить количество признаков и максимально упростить предлагаемый прогностический показатель. С другой стороны, такое упрощение, может сказаться на ошибке. В другой работе (Montoto et al. 2002) анализировались прогностические факторы для безрецидивной выживаемости (survival from progression) - для 90 пациентов с фолликулярной лимфомой (больные, пролеченные в течение 20 лет в гематологической клинике Биомедицинского института Барселоны). Показано, что наиболее значимы – возраст, состояние больного, ЛДГ, бета-микроглобулин,, вовлеченность костного мозга, стадия и IPI. Многопараметрический анализ (с использованием методов регрессии Кокса и др. общепринятых методов) показал, что наиболее важны состояние (performance status) и RD (response duration) – продолжительность ремиссии.

Другим примером может служить исследование (Kloess et al. 1999) в котором авторы провели многопараметрический анализ прогностической значимости методом логистической регрессии у 571 больного после 2905 циклов химиотерапии для оценки гематотоксичности. Анализировались стандартные клинические и лабораторные данные, а также характеристики курсов химиотерапии. Были обнаружены следующие прогностические факторы для лейкоцитопении (в скобках дана отношение несогласия): высокое значение ЛДГ(3.0), возраст более 60 (2.0), женский пол (2.0), плохое состояние (2.1), использование этопозиды (6.1), интервал для G-CSF более 3 недель. Для прогноза тромбоцитопении не играли роли пол и состояние, дополнительно к тем же признакам были важны некоторые параметры курсов химиотерапии (доза винкристина, длительность перерыва и самих курсов). Анемия имела следующие прогностические показатели: изначально низкий гемоглобин (10.5), низкий альбумин (1.3), высокую ГГТ (2.1), и те же показатели, что и для лейкопении (ЛДГ не имел значимого влияния). Интересно, что приводимые цифры дискриминации на группы высокого и низкого риска осложнений, соответственно: лейкопении, тромбоцитопении и анемии, имели цифры ошибочной классификации

(предсказание риска развития осложнения при его отсутствии) и правильной (предсказание попадания в группу риска и развитие осложнения) соответственно 16% и 97%, 0% и 54%, 0% и 28%. Очевидно, что дискриминирующее значение необходимо изменить для тромбоцитов и анемии, что должно увеличить процент правильной классификации. Из приводимых цифр очевидно, что предсказание тромбоцитопении эквивалентно угадыванию, а предсказание анемии (согласно модели) говорит о хорошем прогнозе(!) (только у 27% пациентов данной группы разовьется анемия). Далее авторы подчеркивают, что параметры, вошедшие в разработанную ими модель сходны с параметрами для международного прогностического индекса (PI): это ЛДГ, состояние пациента, возраст. Они также обращают внимание на чрезвычайную гетерогенность гемочувствительности пациентов к эквивалентным курсам химиотерапии.

Данный пример показывает, во-первых, попытки оценки некоторых осложнений, которые, возможно значимы для врача и пациента в процессе лечения, и по-видимому сказываются на качестве жизни, могут не иметь значения для общей и безрецидивной выживаемости, во-вторых - приводимые оценки скорее говорят о необходимости дальнейшего развития модели (по крайней мере это касается тромбоцитопении и анемии).

На проблемы, возникающие при использовании методов регрессии, основанных на модели Кокса указывается в ряде работ (Henderson, 2000, Hendersson et al. 2001). В тоже время, очень часто оценка продолжительности жизни пациента осуществляется на основании индивидуального опыта врача. Оценка точности таких оценок была темой ряда исследований общий вывод которых - предсказательная ценность таких оценок низка. (MacKillop, Quirt, 1997, Christakis, Lamont, 2000, Parkers, 1972, Heyes-Moore, Johnson-Bell, 1987, Forster, Lynn, 1988). В частности (Christakis, Lamont, 2000) указывают, что только 33% оценок продолжительности жизни пациентов при такой оценке адекватны. Почему же формальные статистические модели не используются более часто в практике оценки продолжительности жизни? В тоже время анализ выживаемости очень широко используется для определения прогностических факторов (ковариат), сравнения различных групп пациентов, построения прогностических индексов. В работах (Henderson, 2000, Hendersson et al. 2001) приводится две основные причины, обуславливающие низкую практическую значимость используемых моделей оценки продолжительности жизни. Существующие модели регрессии (сравнивались модели регрессии Кокса, Вейбулла, логнормальная и Аалена) на основании используемых предикторов (например таких как пол, возраст, те или иные

клинические симптомы, стадия процесса) также дают большую ошибку оценки абсолютной величины продолжительности жизни. Эта ошибка сравнима с величиной получаемой при экспертной оценке продолжительности жизни. Математическим отражением низкой практической значимости общепринятых моделей является низкий уровень объясняемой дисперсии отмечаемый в большинстве работ. Автор подчеркивает, что даже знание статистически высоко значимых ковариат не может повысить предсказательную способность на практически значимую величину. Таким образом, по мнению автора, на основании имеющихся моделей врач может сказать больному "прошу прощения, Вы проживете 3 месяца, но есть 60% вероятность, что я ошибаюсь" или "есть 80% шансы, что Вы умрете между 2 неделями и 2 годами". Автор делает вывод, что хотя много усилий затрачивается на сбор материала, общепринятые модели полезны для сравнения групп и анализа предикторов, адекватной оценки абсолютной величины периода жизни они дать не могут и предполагает, что возможно, это является необходимым (обязательным) следствием реальных данных используемых при анализе выживаемости.

Метод искусственных нейронных сетей

Искусственные нейронные сети, отличаются от других методов математического анализа тем, что при поиске решения не следуют строго фиксированному набору математических методов, а повторяют дедуктивные методы поиска решения, свойственные работе эксперта (Боровиков 2001). Более того, нейронные сети могут включать практически неограниченное число факторов для построения прогноза, и эти факторы можно добавлять в модель, если экспериментатор решит, что новые факторы будут полезны для улучшения результата. Метод может также оценить степень прогностической значимости и взаимодействие новых факторов с уже имеющимся набором. Это отличает его от традиционных статистических методов. Особенностью метода ИНС является то, что освоить и применять технологии нейронных сетей значительно проще, чем изучать математическую статистику или нечеткую логику. (Ежов, 1977, Круглов, Борисов, 2001, Haykin, 1994, Ripley, 1996) (Основные принципы метода искусственных нейронных сетей рассматриваются в главе "Материалы и методы").

Количество публикаций по применению нейротехнологии в медицине трудно оценить точно. Однако если в 1988-89 годах их были единицы, то с 1995 года ежегодно появляются сотни. Нейросети уже довольно давно активно

применяются в анализе геномных последовательностей ДНК, в частности для распознавания промоторов - участков, предшествующих генам и связываемых с белком РНК-полимераза, который инициирует транскрипцию. Их используют для дифференциации кодирующих и не кодирующих участков ДНК (экзонов и интронов) и предсказания структуры белков (Ripley 1996) Диагностика и лечение онкологических заболеваний, а также разработка новых медикаментозных средств несомненно представляют собой важнейшую область применения нейросетевых технологий. Диагностика одной из форм меланомы была реализована с помощью нейросетевого симулятора Multineuron, разработанного в ВЦ СОАН в Красноярске под руководством А.Н.Горбаня (Горбань, Россиев, 1996, Горбань и др. 1998)

Одно из главных направлений, в котором сейчас идут работы по использованию нейронных сетей, - диагностика рака молочной железы. Ряд групп использует метод нейронных сетей дифференцировать рентгенологические изображения для постановки диагноза на основе особенностей, с которыми обычно имеют дело радиологи. Оказалось, что ИНС способны решать поставленную задачу более эффективно (Jiang 2001, Tourassi 2001, Brake 2000). В работе авторов (Shia et al., 1999) сравнивается эффективность метода искусственных нейронных сетей с другими методами прогноза и с традиционными линейными моделями. Полученные выводы позволяют авторам сказать, что НН, как правило, дает лучшее предсказание, чем традиционные методы и могут использоваться вместо них, а также как часть сложных экспертных систем для повышения предсказательной эффективности.

В исследовательском медицинском центре Университета Ньюкасла, Великобритания в предварительных пилотных исследованиях (Snow et al., 1994) оценивались диагностические критерии биопсии у мужчин с подозрением на рак предстательной железы и вероятность рецидива после радикальной простатэктомии. Было показано, что применение метода искусственных нейронных сетей позволяет получить модель обладающую высокой чувствительностью и специфичностью. Особо следует подчеркнуть, что использованные в работе показатели - хорошо известные, традиционные клинические и биохимические критерии.

В серии работ (Naguib, Hamdy, 1997, 1998, Naguib et.al. 1996) анализировалось применение варианта метода искусственных нейронных сетей для оценки прогностических факторов. В качестве исходных данных для анализа использовались как традиционные прогностические клинические показатели - возраст, стадия, наличие отдаленных метастазов,

коэффициент Глисона, уровень простатического сывороточного антигена, а также наличие или отсутствие экспрессии p53 и bcl-2 и вариант примененного лечения. Всего 8 показателей. Результат проведенного лечения или течения заболевания делился на три категории, которые и оценивались моделью: 1- нет ответа на лечение, 2- рецидив после эффективного лечения и/или прогрессия у не леченных пациентов, 3- стойкий положительный эффект после лечения и/или отсутствие прогрессии (роста опухоли, распространения заболевания) у не леченных пациентов. Всего проанализировано 38 пациентов, которые делились на две равномерные группы случайным образом (по 19 в каждой и это при 8 входных признаках) одна из которых использовалась как группа обучения, а вторая – как группа проверки результатов обучения модели. В качестве выхода использовали трехуровневый вариант (три класса): 1-нет ответа на лечение, 2- рецидив вслед за успешным лечением или прогрессия болезни в случае нелеченных пациентов, 3- стойкая ремиссия или отсутствия прогрессии болезни в не леченных случаях. Использовалась модель GRNN, которая является обобщением более специфических вероятностных нейронных сетей, использующихся для целей классификации. Одними из главных достоинств этого типа сетей является их быстрая обучаемость, их сходимость к оптимальному решению при увеличении числа объектов и их эффективность при работе с переменными (sparse data) данными, которые отличаются большой ошибкой (большой сигмой). Структура использованной сети – 8 входных нейронов, 19 – в промежуточном слое и 3 выходных. Для внутреннего слоя использовалась функция суммации по методу городских кварталов, а передаточная функция – Функция Гаусса. Чувствительность оценивалась как процент правильных предсказаний рецидивов, а специфичность – как процент правильного предсказания ответа на лечение. Были получены следующие результаты – специфичность – 80 %, чувствительность – 92,9 %, общая эффективность модели – 89,5 %.

В другой работе эта группа авторов (Nagib et. al. 1999) оценивала наличие метастазов в региональные лимфоузлы, когда клинически они еще не определяются. В качестве источника информации использовались цитологические данные. Полученные результаты позволили сделать вывод, что разработанный метод лучше предсказывает вероятность развития метастазов чем гистологический.

Приведенные результаты исследований продемонстрировали применимость метода ИНС для решения задач классификации в случае оценивания прогноза, а также показали, что принципиальное ограничение для методов классической статистики – "кошмар размерности", не так строг в случае алгоритма нейронных сетей. Оценивая результаты, подчеркивается,

что традиционные статистические методы: анализ Кокса (Cox's proportional hazards) и метод логлинейной регрессии (logistic regression) (Erllichman et. al., 1990) применимы, однако в ситуации, когда анализируются данные, зависящие от времени, и когда прогностические факторы взаимодействуют сложным образом и эти зависимости сложно оценить и выявить, обычные статистические методы могут иметь сильные ограничения на стадии оценки полученных результатов.

Следует заметить, что авторы не рекомендуют совсем отказаться от статистическим методов, однако считают, что метод нейронных сетей хороший дополнительный метод для более аккуратного построения прогноза как эффекта лечения, так и продолжительности жизни. (Boracchi et. al., 2001) также сравнивали информативность стандартных методов статистики и метода нейронных сетей для решения задач прогноза на группе 2233 больных раком легкого. Авторы показали, что стандартные методы проигрывают в эффективности.

В работе (Shi et al. 1999) авторы разбирают важную проблему ограничения методов линейной регрессии, встающих при решении сложных задач прогнозирования. Они подчеркивают, что линейная комбинация очень часто не может описать сложные временные зависимости, особенно если анализируемые переменные содержат большую ошибку. Именно с такими данными приходится иметь исследователю - клиницисту. Авторы убедительно доказывают, что применение методов искусственных нейронных сетей часто оказываются более информативными для решения такого рода задач. По мнению авторов, повышение качества прогнозирования можно добиться комбинацией метода ИНС (искусственных нейронных сетей) и других методов прогнозирования (Markov Gated Expert, high order statistics, fuzzy logic).

Большое количество работ по нейронным сетям подчеркивают их преимущество в решении проблемы уменьшения размерности входных данных. В качестве таких методов могут быть предложены компоненты нейронных сетей как анализ чувствительности (Engelbrecht et. al. 1998) сети Кохонена (Sung 1998). Важным механизмом формирования и выбора информативных параметров, разработанных в последнее время, является генетический алгоритм. Следует подчеркнуть, что большинство авторов считает, что особую важность для выбора информативных, прогностических факторов для построения модели, особенно на первых, разведочных этапах играет мнение эксперта. Однако бывают случаи, когда его знания неполны по объективным причинам, а также варианты, когда построение экспертной системы может быть очень сложным и дорогостоящим.

Другой "ахиллесовой пятой" нейронных сетей является тот факт, что достигнутое решение, пусть и очень эффективное, остается для исследователя черным ящиком оно мало понятно для человеческого ума является мало применимым для областей, требующих объяснения полученного решения. (Shyn, Park, 1999).

Хотелось бы также отметить перспективность применения метода нейронных сетей для решения задач прогноза, связанных с определением таких величин как продолжительность жизни. Как уже отмечалось (Henderson, 2001) использование стандартных подходов регрессии дает неудовлетворительные результаты, характеризующиеся большой ошибкой. В ряде работ показано, что ИНС могут решать подобные задачи регрессии лучше традиционных методов. Так в работе (Lundin, 1999) на большом клиническом материале (951 случай) исследовалась возможность предсказания продолжительности жизни у больных раком молочной железы. В качестве предикторов использовались 8 стандартных характеристик: размер опухоли, местная распространенность, гистологический тип, митотический индекс, ядерный полиморфизм, формирование тубулярных структур, уровень некроза в опухоли, возраст пациентов. Получаемые оценки выживаемости были выше чем получаемые на основании стандартного метода логистической регрессии. Количество ошибок в контрольной выборке составляло 40 из 300 (13%). Важной особенностью полученного решения была его низкая чувствительность к уменьшению количества предикторов. Выбрасывание даже наиболее информативных показателей (например, местной распространенности) только незначительно снижало предсказательную эффективность. Авторы также указывают на устойчивость во времени полученной модели и интерпретируют свой вывод одинаковой эффективностью предсказания для 5, 10 и 15 – летнего периода. Они также делают вывод о перспективности применения метода нейронных сетей для оценки продолжительности жизни у онкологических больных. К недостаткам данной работы можно отнести тот факт, что производилось предсказание интервала (5, 10 или 15 лет) а не более точная оценка величины продолжительности жизни. Другим примером использования метода нейросетей может быть работа (De Laurentiis et. al. 1999) посвященная оценке величины периода ремиссии и риска рецидива. Авторам удалось получить модель, позволяющую более эффективно предсказывать наступления рецидива, чем стандартные методы. Авторы подчеркивают, что такой подход позволяет дифференцировать пациентов по уровню риска рецидива внутри групп с одинаковой стадией по TNM.

Таким образом, анализ литературы показал, что использование многомерных методов становится обязательной процедурой анализа медицинских и биологических данных. В мировой литературе этим вопросам посвящены как периодические медицинские и биологические издания, так и большое количество монографий. Имеется хорошее обеспечение и прикладными программами. В тоже время, используемые, в настоящее время в медицине и в онкологии, в частности, методы оценки прогноза обладают рядом недостатков и являются неудовлетворительными по величине регистрируемой ошибки. Разработка новых подходов для решения этих проблем, является важной практической задачей.

Список литературы

1. Аббасова С.Г., Кушлинский Н.Е., Липкин В.М., Трапезников Н.Н. Факты и перспективы изучения Fas-FasL-системы в норме и при патологии // Успехи совр. Биол.-2000.-Т. 120.-№ 3.- С. 303-318.
2. Аспекты клинической дозиметрии.- Ставицкий Р.В., Лебедев Л.А., Гуслистый В.П. и др.; под. ред. Р.В.Ставицкого.- М.: МНПИ, 2000 г.- с.388.
3. Бахлаев И.Е. Паранеопластические синдромы у больных раком легкого.// Вопр. Онкол.-2001.-т.47.-№4.-С.440-442.
4. Бейли Н. Математика в биологии и медицине.- Москва: Мир, 1970.
5. Бельских А. Н., Костюченко А. Л., Жибурт Е. Б., Сизов Д. Н. Новые возможности оценки эффективности экстракорпоральных методов гемокоррекции в лечении больных острыми гнойно-деструктивными заболеваниями легких и плевры // Клин.лаб.диагн.1996,№1,с.42-43.
6. Боровиков В. П. Программа STATISTICA для студентов и инженеров: КомпьютерПресс, Москва, 2001.
7. Боровиков В.П., Боровиков И.П. STATISTICA – Статистический анализ и обработка данных в среде Windows .- М.: Филин,1997.- 608с.
8. Боровиков В.П., Ивченко И.Г. Прогнозирование в системе Statistics в среде Windows.-М: Финансы и статистика, 1999.- 381 с.
9. Возный Э.К., Харченко В.П., Гуров С.Н., Добровольская Н.Ю., Большакова С.А. Некоторые прогностические факторы в комплексном

- лечении местнораспространенного рака молочной железы.//Вопр.онкол.-2000.-т.46.-№ 6.-С.732-735.
10. Генкин А.А. Парные коэффициенты корреляции клинико-лабораторных признаков: артефакты и трудности интерпретации.-Клин. лаб. диагн.-1993.-№6.-с.71-73.
 11. Генкин А.А. Коэффициенты корреляции клинико-лабораторных данных как признаки механизмов регуляции.-Клин. лаб. диагн.-1996.-№3.-с.44-48.
 12. Генкин А.А. Опухолевые заболевания крови: возраст, концентрация холестерина и число эритроцитов // Тер. архив.- 1998.-№3.- с.68-77.
 13. Генкин А.А. О последовательной стратегии Байеса и механизме принятия решений в интеллектуальной системе ОМИС // Клиническая лабораторная диагностика.- 1998.- №4.- С. 42-49.
 14. Генкин А.А. Новая информационная технология анализа медицинских данных.-Политехника, Спб. –1999.- с.192.
 15. Генкин А.А. Программный комплекс ОМИС как инструмент системного анализа клинико-лабораторных данных.//Клин. лаб. диагн. – 1999.-№7.-с.38-48.
 16. Генкин А.А., Корытова Л.И., Степанова И.А. Лимогранулематоз: системный анализ возрастных и половых особенностей гемопоэза и метаболизма.- Тер. Архив.-1997.-№2.-с.37-45.
 17. Генкин А.А., Степанова И.А. Распознавание неходжкинских лимфом и лимфогранулематоза без морфологической информации.- 2-й симпозиум "Неинвазивная диагностика".- под. ред Ю.М.Лопухина.- М.1997.-с.21.
 18. Герштейн Е.С., Кушлинский Н.Е.Тканевые маркеры как факторы прогноза при раке молочной железы.// Практ.онкол.-2002.-3.-№1.-С.38-44.
 19. Гланц С. Медико-биологическая статистика.-М.,Практика,1999.-459 с.
 20. Горбань А.Н., Дунин-Барковский В.Л., Миркес Е.М. Нейроинформатика.. – Новосибирск.: Наука. Сиб. предприятие РАН, 1998. - 296с.
 21. Горбань А.Н., Россиев Д.А. Нейронные сети на персональном компьютере. - Новосибирск: Наука, 1996. - 276 с.

22. Гулятьев А. Имитационное моделирование в среде Windows. М.-1998.- с.29.
23. Гуслистый В.П., Лебедев Л.А., Ставицкий Р.В., Блинов Н.Н., Жанина Т.В. Автоматизированная классифицирующая система (АКС) //Кровь-индикатор состояния организма и его систем: Под ред.Р.В.Ставитского.-М.:МНПИ, 1999.-С.21-34.
24. Давыдов М.И., Аксель Е.М.Злокачественные новообразования в России и странах СНГ в 2001 г.-М.: МИА, 2003
25. Давыдов М.И., Нормантович В.А., Киселевский М.В., Волков С.М., Казанова Г.В., Родионов Л.М., Ковригина А.М., Полоцкий Б.Е., Бычков М.Б., Платинский Л.В., Богатырев В.Н., Оразгельдыев Г.Р. Адоптивная иммунотерапия при опухолевых плевритах: клинико-лабораторное исследование.// Рос.онкол.ж.-2000.-№6.-С.14-17.
26. Доста Н.И., Машевский А.А., Прохорова В.И., Муравский В.А. Новый лабораторный комплекс в диагностике, прогнозе и контроле эффективности лечения рака предстательной железы: Новые технологии в современной медицине. -Сб. науч. работ.:Минск. -1999. стр.82-85.
27. Дубров А.М, В.С.Мхитарян, Л.И.Трошин. Многомерные статистические методы:Учебник.-М.:Финансы и статистика,1998.-352 с.
28. Ежов А. Нейронные сети в медицине.- Институт инновационных и термоядерных исследований: Троицк, 1977, с.120.
29. Зильбер А.П. Медицина критических состояний.Общие проблемы. -Петрозаводск: Изд.Петрозаводского университета,1995.- 360 с.
30. Иванов С.Д. Прогнозирование лейкопении при лучевой и химиотерапии больных лимфогранулематозом.// Матер.Конф."Болезнь Ходжкина":.Петрозаводск.-1999.-С.73-74.
31. Иванов С.Д. Прогнозирование лейкопении на начальных этапах лучевой и химиотерапии больных лимфогранулематозом.//Вопр.онкол.-2000.-№2.-т.46.-с.129-135.
32. Исследование системы крови в клинической практике./ Под ред. Г.И. Козинца, В.А.Макарова.-М., 1997.-157с.

33. Кальф-Калиф Я. Я. // Врач. дело. - 1941. - №1. - С. 31-35.
34. Кант В.И. Математические методы и моделирование в здравоохранении.-Мю:Медицина, 1987.-224с.
35. Кашин А.В., Фирсов В.Г. Применение математических диагностических функций для предварительной диагностики некоторых заболеваний легких.1983.Применение новых технических средств в грудной хирургии.Алма-Ата с.34-38.
36. Кендалл М.. Стьюарт А. Многомерный статистический анализ и временные ряды.:М.-Наука.-1976.-736с.
37. Киселев А.В. Прогностические факторы у детей с болезнью Ходжкина.// Вестн. РАМН.-1996.-т.10.-С.26-28.
38. Колыгин Б.А. Отдаленные результаты полихимиотерапии лимфомы Ходжкина у детей.//Вопр. Онкол.-2003.-т.49.-№ 4.-С.501-505.
39. Круглов В.В., Борисов В.В. Искусственные нейронные сети.Теория и практика.-М.: Горячая линия- Телеком, 2001.- 382с.
40. Кровь-индикатор состояния организма и его систем.- под ред. Р.В.Ставицкого.-М.:МНПИ, 1999.-с.22.
41. Кушлинский Н.Е., Дегтярь В.Г., Бабкина Т.В., Соловьев Ю.Н., Алиев Ю.Н., Трапезников Н.Н. Метаболизм андрогенов в злокачественных и доброкачественных опухолях костей.// Бюлл.Эксп.Биол. Мед.-2000.- №1.-с.90-92.
42. Леман Э. Проверка статистических гипотез.-М.Наука,1979.- 408 с.
43. Машевский А.А. Диагностика основных форм злокачественных опухолей по биохимическим и биофизическим характеристикам крови: Автореф.... д-ра мед наук.-Минск, 1994. -35 с..
44. Медик В.А., Фишман Б.Б., Токмачев М.С. Руководство по статистике в медицине и биологии. В 2-х томах/Под редЮ.М.Комарова.Т.2. Прикладная статистика здоровья./Медик В.А., Фишман Б.Б., Токмачев М.С.-М.:Медицина, 2001.-352с.
45. Меньшиков В.В. От достижений фундаментальной науки – через лабораторию- к эффективной диагностике и лечению.//Клин. Лаб. Диагн.-2003.-№ 6.-с.53-54.

46. Медицинская лабораторная диагностика. Программы и алгоритмы: Справочник. /Под ред. А.И. Карпищенко.-Санкт-Петербург:Интермедика,1997.- 296с.
47. Мирвалиев Мирсултан. Исследования по обобщенным статистикам типа Хи-квадрат.: автореферат дисс. На соиск. Учен. Степ. Докт.Физ-мат наук.Ташкент.-2001.
48. Мэйндоналд Дж.Вычислительные алгоритмы в прикладной статистике.-М.:Финансы и статистика.-1988.-349с.
49. Москаленко О.А., Поддубная И.В., Османов Д.Ш., Пробатова Н.А.,
50. Нейронные сети. STATISYICA Neural Networks: Пер. с англ.- М.:Горячая линия-Телеком.2001.-182с.
51. Новицкий В.В., Степовая Е.А., Гольдберг В.Е., Колосова М.В., Рыжаков В.М., Губерт Э.А., Баженова Н.Г., Часовских Н.Ю., Соколова И.Б., Корешкова К.Г. Функциональная характеристика эритроцитов периферической крови у онкологических больных.// Клин. лаб.диагн.-1997.-№ 12.-с.34-36.
52. Номограммы для определения некоторых интегральных показателей крови человека / Борисов С. Н., Карпов В. Н., Лаптева Д. Г.,Тихончук В.С., Ушаков И.Б., Хованский Г.С./Под ред.Прудникова А.П.—М.:Изд. ВЧ АН СССР, 1989.-30 с.
53. Плескун В.И.. Воронина Е.Д. Теоретические основы организации и анализа выборочных данных в эксперименте.:Ленинград.-Изд-во ленингр. ун-та.-1979.-23с.
54. Понкратова Т.С., Дубинина И.Г., Старостина Е.Г. Комплексный подход к ведению заболевания: Клинический диагност-лабораторныеосновы/ Под ред. В.В.Меньшикова.-М.:Лабинформ,1997.-301 с.
55. Савченко А.А., Шакина Н.А., Россиев Д.А. Нейросетевая классификация лиц с хроническими неспецифическими заболеваниями легких по величине иммунологических параметров крови и активности метаболических ферментов лимфоцитов // Вопросы медицинской химии.- 1999.- № 5.- с. 65-69.

56. Савченко Р.П., Комаров В.Т., Петренко А.Г., Татарченко И.П. Системный подход к клинико-лабораторной диагностике инфекционного эндокардита. // Клиническая лабораторная диагностика.-2001.-№6.-с.44-46.
57. Славин М.Б. Методы системного анализа в медицинских исследованиях.-М.: Медицина, 1989.-304с.
58. Сусликов В.И. Метод вычисления риска заболевания пациента по количественным данным его истории болезни.- М.: Ин-т. прикл. мат. им. М.В. Келдыша. Препринт №47.-1999.-30 с.
59. Тепляков В.В., Лебеденко И.М., Ковалевский Е.Е. Использование метода АКС для клинической оценки "ответа на лечение" онкологических больных. // Современная онкология.- 2002.-Т.4.-№4.- С.202-206.
60. Трапезников Н.Н., Аксель Е.М. Заболеваемость злокачественными новообразованиями и смертность от них населения стран СНГ в 1998 г. .-М.: РОНЦ им.Н.Н.Блохина РАМН.-200.-270с.
61. Флетчер Р., Флетчер С., Вагнер Э.- Клиническая эпидемиология. Основы доказательной медицины.-М.- Медиасфера.-1998.-с.347.
62. Харченко В.П., Кузьмин И.В. Рак легкого. Руководство для врачей.- М.: Медицина.-1994.- 480с.
63. Харченко В.П., Панышин Г.А., Сотников В.М., Даценко П.В. Современная лучевая терапия лимфогранулематоза // Доклад на VI ежег. Росс. онкол. конф.-Тез. докл.- М.- 2002.-с.148.
64. Чехонадский В.Н., Беневский А.И., Давыдов М.И. Формализация выбора метода лечения в клинической онкологии // Вопросы онкологии.-2002.-№6. -С. 691-694
65. Чазов Е.И., Царегородцев Г.И., Кротков В.И. Опыт философско-методологического анализа врачебной диагностики. -Вопросы философии.-1986.-№9.-с.86-87.
66. Эмануэль В.Л., Генкин А.А. Компьютеризация лабораторных исследований – путь развития лабораторной медицины // Клин. лаб. диагн.- 1997.-№5.-с.34.

67. Adamson J.W., Ludwig h. Predicting the hematopoietic response to recombinant human erythropoietin (epoetin alfa) in the treatment of the anemia of cancer// *Oncology* 1999.-v. 56 (1).- p.46-53.
68. Aitchison T.C., Sirel J.M., Watt D.C., MacKie R.M. Prognostic trees to aid prognosis in patients with cutaneous malignant melanoma. Scottish Melanoma Group.// *Brit. Med. J.*-1995.-v.311(7019).-p.1536-9.
69. Armitage P. Theory and practice in medical statistics / *Statistics in Medicine.*- 2001.- v. 20.- Issue 17-18.-p. 2537-2548
70. Axdorph U., Sjoberg J., Grimfors G. et al. Biological markers may add to prediction of outcome achieved by the international prognostic score in Hodgkin's disease // *Ann. Oncol.*-2000.-Vol.11.-P.1405-1411.
71. Batchelor R., Dua P. Forecaster diversity and the benefits of combining forecasts. // *Management Science.*-1995.- v. 41.-p.68–75.
72. Bates D.W., Cohen M., Leape L.L., overhage J.M., Shabot M.M., Sheridane T. Reducing the frequency of errors in medicine using information technology.// *J. Amer. Med. Infor. Assoc.*-2001.-v.8.-p.299-308.
73. Berruti A, Osella G, Raucci CA, Roncari A, Dogliotti L. Transient increase in total serum alkaline phosphatase predicts radiological response to systemic therapy in breast cancer patients with osteolytic and mixed bone metastases.// *Oncology* 1993.- v.50(4).- pp.218-21
74. Beguin Y. Predicting responses to rhEPO: Choosing the best strategy.// *Anaemia in cancer.*-2002.-v.3.№1.-p.8-10.
75. Bierman P.J., Lynch J.C., Bociek R.G., Whalen V.L., Kessinger A., Vose J.M., Armitage J.M. The International prognostic factors project score for advanced Hodgkin's disease is useful for predicting outcome of autologous hematopoietic stem cell transplantation.// *Ann. Onc.*-2002.-v.13.-p.1370-1377.
76. Blough D.K., Anderson K.K. A comparison of artificial neural networks and statistical analyses.// *Technical Reports. Pacific North.N.Lab.Richmond.Washington.*-1994.-rep 94.
77. Boracchi P, Biganzoli E, Marubini E. Modelling cause-specific hazards with radial basis function artificial neural networks: application to 2233 breast cancer patients. // *Stat. Med.*- 2001.-Vol. 20.-№ 24.-P.3677-94.

78. Bottini A, Berruti A, Brizzi MP, Bersiga A, Generali D, Allevi G, Aguggini S, Bolsi G, Bonardi S, Bertoli G, Alquati P, Dogliotti L. Pretreatment haemoglobin levels significantly predict the tumour response to primary chemotherapy in human breast cancer // *Br. J. Cancer.*- 2003.- Vol.89.-№6.-P.977-982.
79. te Brake GM, Karssemeijer N, Hendriks JH. An automatic method to discriminate malignant masses from normal tissue in digital mammograms // *Phys. Med. Biol.*- 2000.- Vol.45.-№10.-P.2843-57.
80. de Bruijne M.H.J., le Cessie S., Kluine-Nelemas H.C., van Houwelingen H.C. On the use of Cox regression in the presence of an irregularly observed time-dependent covariate // *Statistic in medicine.*- 2001.-v.20.-№ 24.-p.3817-3829.
81. Buccheri GF, Ferrigno D. Prognostic factors in lung cancer: tables and comments // *Eur Respir J.*- 1994.- v.7.- pp.1350-1364.
82. Buccheri GF, Ferrigno D, Ginardi C, Zuliani C. Haemostatic abnormalities in lung cancer: prognostic implications. // *Eur. J. Cancer* 1997; 33: 50-57.
83. Bustin S., Dorudi S. Molecular assessment of tumour stage and disease recurrence using PCR-based assays // *Mol.Med.Today.*-1998.- September.- p.389-395.
84. Christakis N.A., Lamont E.B. Extent and determinants of error in doctors' prognoses in terminally ill patients: prospective cohort study // *British Medical J.*- 2000.-v.320.- pp.469-473.
85. Clare A., King R.D. Machine learning of functional class from phenotype data // *Bioinformatics* .- 2002.-Vol18.-№1.-P.160-166.
86. Clayton D. Fitting a general family of failure time distributions using GLIM. *Applied Stat.*- 1983.-Vol.32.-P.268-275.
87. Cockburn F., Cooke R.W.J., Gamsu H.R. The CRIB (clinical risk index for babies) score: a tool for assessing initial neonatal risk and comparing performance of neonatal intensive care units // *Lancet.*- 1993.-Vol. 342.- P.193-198.
88. Coiffier B. Non-Hodgkin's Lymphomas. Clinical presentation, treatment and outcome.-Basel.:F.Hoffmann-La Roche Ltd, 2001.- 74 p.
89. Cox, D. R. (1972). Regression models and life tables // *Journal of the Royal Statistical Society*, 34, 187-220.

90. Cox, D. R., Oakes D. Analysis of survival data / New York: Chapman & Hall. -1984.
91. Crivellari D., Price K.N., Hagen M., Goldhirsh A., Gelber R.D., Castiglione M., Coates A.S., Rudenstam C.M., Collin J., Lindter J. Routine tests during follow-up of patients after primary treatment for operable breast cancer. International (Ludwig) Breast Cancer Study Group. // Ann. Oncol.-1995.- v.6(8).-p.769-76.
92. Cullen D.J., Civetta J.M., Briggs B.A. Therapeutic intervention scoring system: a method for quantitative comparison of of patient care // Crit. Care Med.- 1974.-Vol2.-№1.-P. 57-62.
93. Dalluge K.H., Ziegenbein R. Enzymatic tests in the prognostic and course in bronchial carcinoma patients.// Arch. Geschwulstforsch.-1982.-v.52(3).-p.199-202.
94. Denne S.J. Sample size recalculation using conditional power//Statistics in Medicine.- 2001. - V. 20.-Issue 17-18.- P.- 2645-2660
95. Dhodapkar M.V., Jacobson J.L., Gertz M.A., Rivkin S.E. Prognostic factors and response to flurabine therapy in patients with Waldenstrommavroglobulinemia: results of United States intergroup trial (Southwest Oncology Group S9003).// Blood.-2001.-v.98.-№ 1.-p.41-48.
96. Erlichman C., Warde P., Gadalla T., Ciampi A., Baskerville T., RECAP analysis of prognostic factors in patients with stage III breast cancer. // Breast Cancer Res. Treat. 1990.- v.16.- pp. 231.
97. Engelbrecht A.P., Cloete I., Zurada J.M. Determining the significance input parameters using sensitivity analysis: Proceeding of International Workshop on Artificial Neural Network, Malaga-Terramolinos, 2001.- (pp.382-388)
98. Ferrigno D, Buccheri G, Camilla T. Prognosis and lung cancer: the contribution of plasma proteins. // Oncol. Rep. 1995.- v.2.- pp.637-641.
99. Forster C.M., Lynn J. Can doctors accurately predict the life expectancy of patients with terminal cancers? // Arch. Int. Med.- 1988.- Vol.148.-P.2540-43.
100. Freindlin B., Korn E.L. A testing procedure for survival data with few responders. // Statistics in Medicine.-2002.-v.21.-№ 1.- p.65-78.

101. Fumagalli LA, Vinke J, Hoff W, Ypma E, Brivio F, Nespoli A. Lymphocyte Counts Independently Predict Overall Survival in Advanced Cancer Patients: A Biomarker for IL-2 Immunotherapy. // J Immunother.-2003.-Vol. 26.-№5.- P.394-402.
102. Giles F.J., Lim S.W., Chaimogkol B., et al. Clinical characteristics at diagnosis of asian non-Hodgkin's lymphoma patients on the international oncology study group (IOSG) NHL1 study. //Proceeding of ASCO.-1996.-Vol.15.-P.506.
103. Golub T.R. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring.// Science.2001-v.286.-p.531-537.
104. Gould A.L. Sample size re-estimation: recent developments and practical considerations. // Statistics in Medicine.-2001.-v. 20.-p.2625-2643.
105. Jiang Y., Nishikawa R.M., Papaioannou J. Dependence of computer classification of clustered microcalcifications on the correct detection of microcalcifications // Med. Phys.- 2001.- Vol.28.-№9.-P.1949-57.
106. Hasenclever D., Diehl V. A numerical index to predict tumor control in advanced Hodgkin's disease // Blood.- 1996.- Vol. 88.-P. 673a, (suppl 1).
107. Haykin, S. (1994). Neural Networks: A Comprehensive Foundation / New York: Macmillan College Publishing.
108. Justice A.C., Feinstein A.R., Wells C.R. A new prognostic staging system for the acquired immunodeficiency syndrome // N Eng J Med.- 1989.- 320.-1388-1393.
109. Hannisdal E., Tveit K.M., Theodorsen L., Host H. Host markers and prognosis in recurrent rectal carcinomas treated with radiotherapy.//ActaOncol.-1994.-Vol.33.-415-21.
110. Health service guidelines 95(8).Department of health guidelines.-1995.: HMSO
111. Hess K.R. Assessing time-by-covariate interactions in Cox proportional hazards regression models using cubic spline functions.// Stat.Med.-1994.-v.13.-p.1045-1063.
112. Heyes-Moore L.H., Jonson-Bell V.E. Can doctors accurately predict the life expectancy of patients with terminal cancer. // Palliative Medicine.-1987.-v.1.-pp.165-166.

113. Henderson R. Analysis of multivariate survival data.-NY.:Springer, 2000.-542 p.
114. Henderson R., Jones M., Stare J. Accuracy of point predictions in survival analysis // *Statistics in Medicine*.-2001.-v.20.-Is.20.-p.3083-3096.
115. Holford T.R. The analysis of rates and of survivorship using log-linear models. *Biometrics*.- 1980.-Vol 36.-P.299-305.
116. Huijgen H.J., Sanders G.T., Koster R.W., reeken J., Bossuyt P.M. The clinical value of lactate dehydrogenase in serum: a quantitative review.//*Eur. J.Clin. Chem. Clin. Biochem*.-1997.- Vol. 35.-P.569-79.
117. Ikeda M, Natsugoe S, Ueno S, Baba M, Aikou T. Significant host- and tumor-related factors for predicting prognosis in patients with esophageal carcinoma // *Ann Surg*.- 2003.- Vol.238.-№ 2.-197-202.
118. Kimura E., Murae M., Kobayashi S., Aoki M, Isonishi S., Yasuda M., Tanaka T. Clinical evaluation of computer aided multivariate pattern analysis system II (CAMPAS OV II) for discriminating malignant from benign ovarian masses (BOM).// *ASCO, 32 Ann.Meet.*:1996.-v.15.-p.90.
119. Kloess M., Wunderlich A., Truemper L., Pfreunschuch M., Loeffler M. Prognostic factors for hematotoxicity in aggressive NHL.-5th International Conference on Malignant Lymphoma.-Lugano 1999,p.186.
120. Knaus W.A., Wagner D.P., Draper E.A. The APACHE-III prognostic system // *Chest*.-1991.- Vol100.-№6.-P.1619-1636.
121. Lagerwaard F.J., Levendag P.C., Nowak P.J., Eijkenboom W.M., Hanssens P.E., Schmitz P.I. Identification of prognostic factors in patients with brain metastase: a review of 1292 patients.// *Int.J. Radiat. Oncol. Biol. Phys*.-1999.-Vol.43.- 795-803.
122. Laird N.M., Olivier D. Covariance analysis of censored survival data using log-linear analysis techniques.//*J.Amer.Sta.Assoc*.-1981.-Vol.76.- P.231-240.
123. De Laurentiis M, De Placido S, Bianco AR, Clark GM, Ravdin PM. A prognostic model that makes quantitative estimates of probability of relapse for breast cancer patients // *Clin. Cancer Res*.- 1999.- Vol.5.-№12.-4133-9
124. Lee Y.H., Haymond H.R., Feder B. Biochemical evaluation of patients with breast cancer.// *J. Surg. Oncol*.-1982.-v.19.-№4.-3Ю 197-200.

125. Lee Y.H., van Peenen H.J., Waston F.R. Chemical and hematological screening in patients with malignant and non-malignant conditions.// *J.Surg. Oncol.*-1980.-v.15(1).-p.43-51.
126. Lefkopoulou M., Ryan L., Global tests for multiple binary outcomes.// *Biometrics.*-1993.-Vol.49.-P.975-988.
127. Legler J.M., Lefkopoulou M., Ryan L. Efficiency and power of tests for multiple binary outcomes. *J. Amer. Stat. Assoc.*-1995.-Vol.90.-P.680-693.
128. Lo J.Y., Markey M.K., Baker J.A., Floyd C.E. Jr. Cross-institutional evaluation of BI-RADS predictive model for mammographic diagnosis of breast cancer // *Am. J. Roentgenol.*- 2002.- Vol. 178.-№ 2.- P. 457-63.
129. Lundin M, Lundin J, Burke HB, Toikkanen S, Pylkkanen L, Joensuu H. Artificial neural networks applied to survival prediction in breast cancer // *Oncology.*- 1999.-Vol.57.- №4.-P.281-6.
130. MacKillop W.J., Quirt C.F. Measuring the accuracy of prognostic judgments in oncology.// *J. Clin. Oncol.*-1997.-v.50.- 21-29.
131. Marshall G., Warner B., MaWhinney S., Hammermeister K. Prospective prediction in the presence of missing data.// *Statistics in Medicine.*-2002.-v.21.-№4.-p.561-570.
132. Martinow A.J., Yuen K., Matthews J.P., Juneja S., Wolf M., Januszewicz H., Prince H.M. Prognostic markers of disease activity in Hodgkin's disease.// *Leuk. Lymphoma.*- 1998.- Vol.29.-P. 282-9.
133. McCullagh P. Nelder J.A. *Generalized Linear Models*, 2nd ed. Chapman and Hall: London.-1989.- 378 p.
134. Mendenhall N.P., Cantor A.B., barre D.M., Lynch J.W., Millon R.R. The role of prognostic factors in treatment selection for early-stage Hodgkin's diseases.// *Am.J.Clin.Oncol.*-1994.-v.17.-№3.-p.189-195.
135. Midis GP, Shen Y, Owen-Schaub LB. Elevated soluble Fas (sFas) levels in nonhematopoietic human malignancy.// *Cancer Res.*- 1996.- v.56(17).- p.3870-4.
136. Montoto S., Lopez-Guillermo A., Ferrer A., Camos M.- Survival after progression in patients with follicular lymphoma: analysis of prognostic factors.- *Annals of Oncology.*- 2002.-v.13.-p.523-530

137. Mori M., K. Kitamura, M. Masuda, T. Hotta, T. Miyazaki, R. Miura, H. Mizoguti, A. Shibata, H. Saitoh, T. Matuda, T. Masaoka, M. Harada, Y. Niho, F. Takaku. Prognostic Factors in Elderly with Non-Hodgkin's Lymphoma: Development of a Model to Predict Survival. /The Japanese Clinical Study Group of THP Lymphomas in the Elderly (JGTLE).- Tokyo, Japan, 1999
138. Muers M.F., Shevlin P., Brown J. Prognosis in lung cancer: physicians' opinions compared with outcome and a predictive model.// Thorax.- 1996.- v.51.-pp 894-902.
139. Naguib R.N.G., Hamdy F.C.. Prognostic neuroclassification of prostate cancer patients, Proc. IEEE Internat. Conf. Eng. Med. Biol. Soc., Chicago, 1997, pp. 1003-1006.
140. Naguib R.N.G., Hamdy F.C. A general regression neural network analysis of prognostic markers in prostate cancer. // Neurocomputing.- 1998.- v.19.- pp. 145-150.
141. Naguib R.N., Sakim H.A., Lakshmi M.S., Wadehra V., Lennard T.W., Bhatavdekar J., Sherbet G.V. DNA ploidy and cell cycle distribution of breast cancer aspirate cells measured by image cytometry and analyzed by artificial neural networks for their prognostic significance // IEEE Trans. Inf. Technol. Biomed.- 1999.- Vol.3.-№1.-P.61-9.
142. Naguib R.N.G, Robinson M.C., Apakama I., Neal D.E., Hamdy F.C.. Neural networks analysis of prognostic markers in prostate cancer.// Br. J.Urol.- 1996.- v. 77 (1) pp. 50.
143. Naguib R.N.G, Robinson M.C., Neal D.E., Hamdy F.C. Neural networks: a new tool to predict outcome in prostate cancer. // Eur. Urol.- 1996.- v. 30 (2).- pp. 73.
144. Nexo E. Рациональное применение измерений и наблюдений в клинической лабораторной диагностике. (Докл на XVI межд.конг. по клин. химии(Лондон, июль 1996))//Клин.лаб.диагн.-1998.-№1.-с.46-47.
145. Ng E.Y., Ung L.N., Ng F.C., Sim L.S. Statistical analysis of healthy and malignant breast thermography. J. Med. Eng. Technol. 2001.- Vol.25.- №6.-P.253-63.
146. Nguyen D.V., Rocke D.M. Tumor classification by partial least squares using microarray gene expression data. // Bioinformatics.- 2002.- Vol. 18.-

pp. 39-50.

147. Normand S-L.T., Zou K.H. Sample size considerations in observational health care quality studies.// *Stat in Med.*-2002.-Vol21.-P.331-345.
148. Olshen A.B., Jain A.N. Deriving quantitative conclusions from microarray expression data. // *Bioinformatics.*-2002.-Vol.18.-No.7.-Pp. 961-970
149. Park P.J., Tian Lu, Kohane I.S. Linking gene expression data with patient survival times using partial least squares.// *Bioinformatics.*-2002.-v.18.-no 90001.-p.s120-s127.
150. Papatestas A.E., Lesnick G.J., Jenkins G., Aufses A.H.Jr. The prognostic significance of peripheral lymphocyte counts in patients with breast carcinoma.// *Cancer.*-1976.-v.37.-№.1.-p.164-8.
151. Parkes C.M. Accuracy of predictions of survival in later stages of cancer.// *British Medical J.*- 1972.-v.2.-pp29-31.
152. Pastore G, Viscomi S, Gerov GL, Terracini B, Madon E, Magnani C. Population-based survival after childhood lymphoblastic leukaemia in time periods corresponding to specific clinical trials from 1979 to 1998--a report from the Childhood Cancer Registry of Piedmont (Italy) // *Eur J Cancer.*-2003.- Vol. 39.-№ 7.-P.952-60.
153. Pauler K.D., Laird N.M. Non-linear hierarchical models for monitoring compliance // *Statistics in Medicine.*- 2002.- V. 21.- Iss. 2.- P. 219-229.
154. Piccini L., Croci E., Sacchi S., Curci G.- Hematologic changes in renal neoplasms.// *Minerva Med.* 1984,-N 10,- 75 (43),- pp 2597-602.
155. Prognostic factors in Cancer (second edition)/ Ed. M.K. Gospodarowicz et al.- USA: Publish offisen-Wiley-Liss, 2001.-2001.- 809 p.
156. Raychaudhuri, Sutphin P.D., Chang J.T., Altman R.B. Basic microarray analysis: grouping and feature reduction // *Trends in Biotech.*-2001.-Vol.19.-№5.-P.189-192.
157. Ripley, B. D. Pattern recognition and neural networks. /Cambridge: Cambridge University Press.- 1996.- p.57.

158. Rochon J., Gillespie B.W. A methodology for analysing a repeated measures and survival outcome simultaneously. // *Statistics in Medicine*.-2001.- V. 20.- P. 1173-1184.
159. Rotstein S., Blomberg H., Petrini B., Wasserman J., Nillson B., Baral E. Blood lymphocyte counts with subset analysis in operable breast cancer. Relation to the extent of tumor disease and prognosis.// *Cancer*.-1985.-v.56 (6).- p.1413-9.
160. Salles G. Towards new prognostic factors in diffuse large cell non-Hodgkin's lymphoma. *Ann. Oncol.*-1996.- v. 7.-p.993-996.
161. Salomon J-C. If paraneoplastic syndromes facilitate tumor growth!// *Cancer J.*- 1998.-v. 11.-№ 5.-p. 5-7.
162. Salomon J-C. The structure of the disease "cancer".//*The Cancer Journal*,-1997.-v.10.-N3.-P. 3-4.
163. Schauenstein K., Schauenstein E. Diagnostic relevance of non-specific tumor associated immune dysfunctions.// *Cancer J.*- 1998.-v.11.-№3.-p.158-163.
164. Schneider, G., Wrede, P., 1998. Artificial neural networks for computer-based molecular design.// *Prog. Biophys. Mol. Biol.* 70, 175-222.
165. Shipp M.A., Harrington D.P., Anderson J.R., et. al. A predictive model for aggressive Non-Hodgkin's lymphoma.// *N.Engl. J. Med.* – 1993.-v.329.- p.978-994.
166. Shyn C.K., Park M.S., Memory and neural network based expert system. // *Expert systems with applications*. 1999.- v.16.- pp. 145-155.
167. Shia S.M., Xub D.L., Bao Liua. Improving the accuracy of nonlinear combined forecasting using neural networks. //*Expert Systems with Applications* .-1999.- v. 16.- p.49–54.
168. Sim I., Gorman P., Greenes A., Haynes B., Kaplan B., Lehmann H., Tang P.C. Clinical decision support systems for the practice of evidence-based medicine // *J. Amer. Med. Inform. Assoc.*-2001.-Vol.8.-P.527-534.
169. Snow P.B., Smith D.S., Catalona W.J., Artificial neural networks in the diagnosis and prognosis of prostate cancer: a pilot study.// *J. Urol.*- 1994.- v. 152.- pp. 1923.
170. Sung A.H., Ranking importance of input parameters of neural networks,- *Expert systems with applications*. 1998.- v.15.- pp. 405-411.

171. Special Paper Methodological and Statistical Issues of Quality of Life (QoL) and Economic Evaluation in Cancer Clinical Trials: Report of a Workshop.// *European J. of Cancer.*- 1998.- Vol. 34.- No. 9.-P. 1317-33.
172. Sudbø J., Bankfalvi A., Bryne M., Marcelpoil R, Boysen M., Piffko J., Hemmer J., Kraft J., Reith A. Prognostic Value of Graph Theory-Based Tissue Architecture Analysis in Carcinomas of the Tongue.// *Lab. Investigation.*- 2000.-Vol. 80.-P.1881-1889.
173. Sposto R. Cure model analysis in cancer: an application to data from the Children's Cancer Group.// *Statistics in Medicine.*-2002.-v.21.-№ 2.-293-312.
174. Tourassi GD, Markey MK, Lo JY, Floyd CE Jr. A neural network approach to breast cancer diagnosis as a constraint satisfaction problem // *Med. Phys.*- 2001.-Vol.28.-№5.-P.804-11.
175. Unnebrink K., Windeler J. Intention-to-treat: methods for dealing with missing values in clinical trials of progressively deteriorating diseases.// *Statistics in Medicine.*-2001.-v.20.-№ 24.-p.3931-3946.
176. Valet G., L. Russman, R.Wirsching. Automated flow-cytometric identification of colorectal tumor cells by simultaneous DNA, CEA-antibody and cell volume measurements // *J.Clin.Chem.Clin.Biochem.*-1984,-22,-pp.935-942.
177. Valet G., H.G.Hoffkes.- Automated classification of patients of patients with chronic lymphatic leukemia and immunocytoma from flowcytometric three colour immunophenotypes // *Cytometry (CCC).*- 1997.-30,-pp.275-288.
178. Valet G., G.Roth, W.Kellermann. Risk assessment for intensive care patients by automated classification of flow cytometric oxidative burst, serine and cysteine proteinase activity measurements using CLASSIF1 triple matrix analysis // *Cytometric Cellular Analysis*, Eds: J.P.Robinson,G.Bobcock, Wiley-Liss, New York 1998, pp.289-306.
179. Wagner W.,Hermann R.,Hartlapp J. Prognostic value of hemoglobin concentrations in patients with advanced head and neck cancer with combined radio-chemotherapy and surgery// *Starhlenter Onkol.*-2000.-v176.- p.73-80.

180. Wallach N., Gur Y. Leukocyte alkaline phosphatase and serum alkaline phosphatase in patients with metastatic breast and colon cancer.// Neoplasma.-1996.-Vol.43.-P.297-300.
181. Weitzman J.B. Chip-ping away at cancer classification // Mol. Med. Today.-2000.-Vol6.-P.50.
182. Welstead, S. T. (1994). Neural network и fuzzy logic applications in C/C++. New York: Wiley.
183. Winston P.W., Bates D.H. Saturated solutions for the control of humidity in biological research // Ecology.-1960.-Vol.41.-P.232-237
184. Wulfsohn M.S., Tsiatis A.A. A joint model for survival and longitudinal data measured with error. //Biometrics.-1997.-Vol53.-P.330-9.
185. Yang B.S., Chen K.H., Lin R.S. Clustering scheme for heterogeneity grouping of healthy adolescents based on a profile of serum biochemical variables.// Proc. Natl. Counc. Repub. China B.-1994.-v.18.-№ 2.-p.76-84.
186. Yee D., Prior M.G., Florence L.Z. Development of predictive models of laboratory animal growth using artificial neural networks.// Bioinformatics.-1993.-v. 9.-p.517-522.
187. Yen GG, Meesad P. Constructing a Fuzzy Rule-Based System Using the ILFN Network and Genetic Algorithm // Int. J. Neural. Syst.- 2001.- Vol.11.-№5.-P.427-43