

МОДЕЛИРОВАНИЕ ФАКТОРОВ РИСКА ЗАБОЛЕВАЕМОСТИ В МЕДИЦИНСКИХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

*Дацун Н.Н., к.ф.-м.н., Суворова И.П., асс., Харченко Е.А., студ.,
Коломойцева И.А. студ., (к-ра ПМиИ)*

ВВЕДЕНИЕ

Медицинская кибернетика - это направление, в котором на основе единых для кибернетики научных методов и идей изучаются системы управления в медицине. Назначение подобных систем - определение состояния больного (диагностика) и выбор наиболее эффективного лечебного воздействия.

В рамках этого направления быстрое распространение получают медицинские компьютерные системы (МКС). В зависимости от сложности выполняемых операций МКС могут быть разделены на четыре типа: системы обработки данных; информационные системы (содержащие информацию, полученную в результате обработки первичных данных); системы обработки знаний; интеллектуальные системы (показывающие интеллектуальное поведение, развитие болезни, формирующие новые знания).

Проблема сбора, накопления, обработки и анализа данных в медицинских исследованиях стоит довольно остро и, в первую очередь, это связано со следующими причинами:

- отсутствием средств автоматизированного сбора, накопления и ведения данных;
- трудоемкостью создания и компьютерной поддержки специализированных форм сбора данных для конкретного класса задач;
- требованиями к комплексному наблюдению за различными системами организма (большое количество различных измеряемых показателей);
- сложностью полноценной интерпретации в силу узкой специализации врачебных знаний.

1. ТЕХНОЛОГИЧЕСКИЙ КОМПЛЕКС "РЕПРОДУКТИВНОЕ ЗДОРОВЬЕ ДОНБАССА"

В рамках договора о научном сотрудничестве между кафедрой прикладной математики и информатики факультета ВТИ и Донецким региональным центром охраны материнства и детства (ДРЦ ОМД) был разработан проект технологического комплекса "Репродуктивное здоровье Донецкого региона".

Технологический комплекс "Репродуктивное здоровье Донецкого региона" как объект компьютеризации является сложной системой. В нем можно выделить следующие слои стратификации:

- структурный;
- функциональный;
- возрастной.

Структурно комплекс охватывает учреждения здравоохранения различного уровня:

- на уровне учреждений (поликлиники, больницы, женские консультации, родильные дома);
- на уровне района, города (районные гинекологи, Дома санитарного просвещения, службы семьи);
- на уровне региона (диагностические центры, ДРЦ ОМД и его филиалы, средние медицинские учебные заведения, медицинский университет и его факультет усовершенствования врачей).

Функционально комплекс включает медицинские компьютерные системы различного назначения:

- системы массовых профилактических осмотров населения (скрининга);
- системы лечебно-клинических учреждений;
- медицинские обучающие системы для различных категорий пользователей.

В технологическом комплексе предусмотрен дифференцированный возрастной подход к различным категориям населения (акушерство и гинекология детей, подростков, взрослых) в связи с ранним половым созреванием детей и ранним началом половой жизни.

Структурно-функциональная схема комплекса "Репродуктивное здоровье Донецкого региона" имеет следующий вид:

РЕПРОДУКТИВНОЕ ЗДОРОВЬЕ ДОНЕЦКОГО РЕГИОНА

- Системы массовых профилактических осмотров
- Системы больниц и поликлиник (отделения и подразделения):
 - Лаборатория
 - Клинические испытания лекарств
 - Скрининг данных
 - Диагностика
 - Мониторинг пациента
 - Лечение
 - Уход за больными
 - Реабилитация
 - АРМы врача
 - Ведение историй болезни
- Обучающие системы

В данном технологическом комплексе все его структурные единицы представляют собой медицинские компьютерные системы. Далее в статье рассмотрены те компоненты этого комплекса, которые связаны с такой областью медицины как детская и подростковая гинекология. Назовем причины, которые определили наивысший приоритет научных исследований и практических реализаций именно в указанном направлении. Главная из них - демографическая ситуация Донецкого региона. По данным ДРЦ ОМД за последнее десятилетие (1984-1994 гг.) количество девочек-подростков - будущих матерей сократилось вдвое. По данным медицинских профилактических осмотров четверть девочек и девушек нашего региона имеют заболевания гинекологического профиля. Вторая причина - отсутствие МКС, которые имели бы указанную нами возрастную направленность - детская и подростковая гинекология. Нами был выполнен аналитический обзор публикаций по такому разделу медицины как акушерство и гинекология (раздел 2 данной статьи), который показал отсутствие систем-аналогов в мировом потоке публикаций.

Структура технологического комплекса подтверждает особенность медицинских систем, которые оперируют с большими объемами фактографических данных. Современный уровень информационных технологий позволяет в каждой структурной единице разрабатываемого технологического комплекса использовать интеллектуальные компоненты. Мы описываем наш подход к решению задачи анализа факторов риска гинекологической заболеваемости детей и подростков на таких уровнях функционального слоя стратификации как системы массовых профилактических осмотров и система отделения детской и подростковой гинекологии ДРЦ ОМД. На этапе скрининга формируются базы данных о пациентках, принадлежащих определенной группе риска того и иного вида заболеваний. Такие пациентки направляются для дальнейшего наблюдения в отделение детской и подростковой гинекологии, где производится более дифференцированная диагностика заболевания и проводится терапевтическое лечение или хирургическая операция. Затем эти пациентки находятся определенное время в поле зрения отделения детской и подростковой гинекологии, которое выполняет мониторинг

состояния их здоровья до полного излечения. Традиционно для проведения диагностики и назначения терапевтического лечения в современных МКС используют экспертные системы (ЭС). Менее традиционным применением названного вида информационных систем является их использование для скрининга и мониторинга.

Нами предложена комплексная модель анализа риска гинекологической заболеваемости для этапов скрининга, диагностики и терапии, а также мониторинга состояния пациентки.

2. АНАЛИЗ ПОТОКА ПУБЛИКАЦИЙ ПО ЭКСПЕРТНЫМ СИСТЕМАМ В МЕДИЦИНЕ

Анализ публикаций проводился на основе вторичного документального потока (рефераты публикаций в компьютерной базе данных MEDLINE) с глубиной поиска 8 лет (1988-1995 гг.). Полнотекстовая гипертекстовая база данных (БД) MEDLINE содержит рефераты журнальных статей по биологии и медицине. Запрос к этой БД был выполнен по ключевым словам "EXPERT" и "SYSTEMS" классификатора MESH. Полученная информация затем далее анализировалась по следующим параметрам:

- распределение по разделам медицины согласно рубрикатора, используемого Медицинским Реферативным Журналом (МРЖ);
- распределение по странам и языкам публикаций;
- распределение по наименованиям источников публикаций;
- распределение по количественному составу авторским коллективов;
- распределение по функциональному назначению (диагностика, терапия, прогноз, анализ и т.д.).

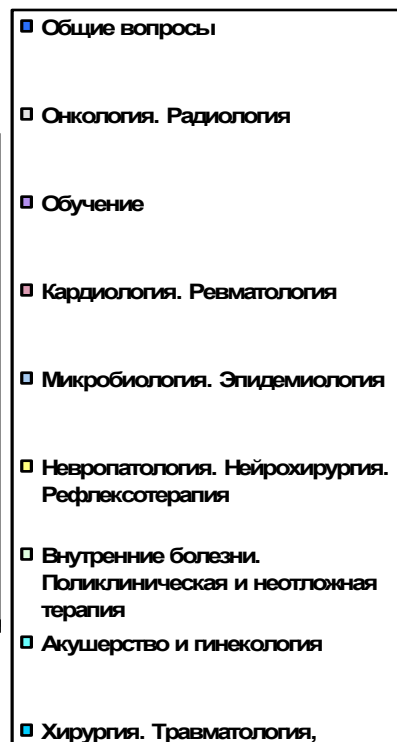
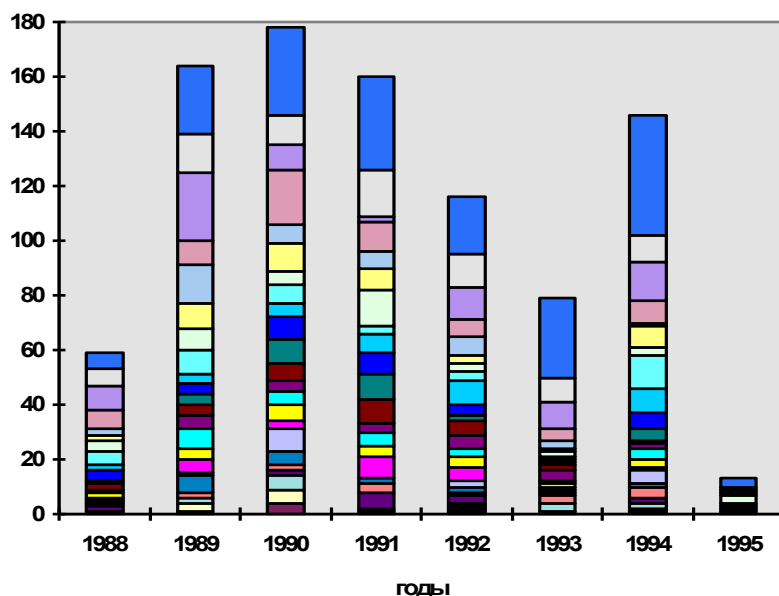
Кроме этого, были проанализированы медицинские экспертные системы (МЭС) по таким характеристикам как:

- способы представления знаний;
- способность к проведению неточных рассуждений;
- интеграция с другими медицинскими информационными системами;
- инструментальные средства, использованные при реализации.

Всего в БД MEDLINE общее количество публикаций по тематике "Экспертные системы" составило 964.

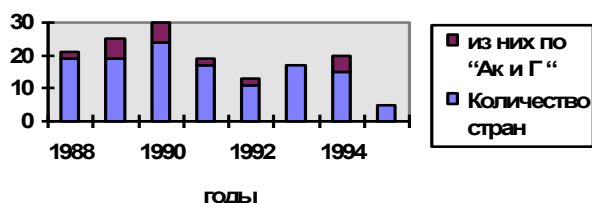
Из них опубликовано в 1988г. 92 статьи, в 1989г. - 155, в 1990г. - 185, в 1991г. - 171, в 1992г. - 57, в 1993г. - 125, в 1994г. - 164, в 1995г. (к настоящему времени) - 15. Из всего потока публикаций нами были исследованы публикации только по медицинским экспертным системам. Распределение публикаций по разделам медицины представлено на следующей диаграмме. Интересующий нас раздел "Акушерство и гинекология" (Ак и Г) по интенсивности публикаций среди 22 разделов медицины занимает 6 место (на диаграмме учтены ЭС междисциплинарного плана, которые посвящены общим вопросам и вопросам обучения). Общее количество публикаций об экспертных системах по этому разделу медицины составило по годам: 1988г. - 5 (5,4%), 1989г. - 9 (5,8%), 1990г. - 7 (3,8%), 1991г. - 3 (1,75%), 1992г. - 3 (5,3%), 1994г. - 12 (7,3%). В 1993г. и 1995г. публикаций по ЭС в акушерстве и гинекологии нет.

Распределение МЭС по разделам медицины

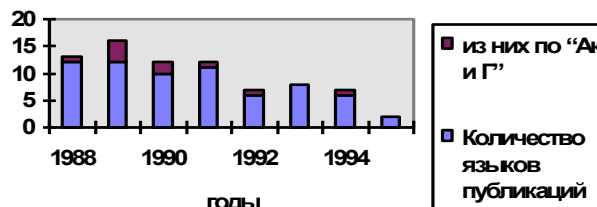


Далее мы будем проводить анализ потока публикаций именно для этого раздела медицины в сравнении с основным потоком.

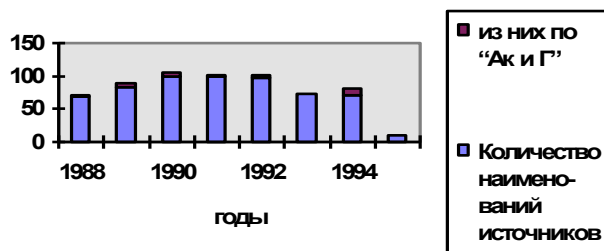
Распределение МЭС по странам публикаций



Распределение МЭС по языкам публикаций



Распределение МЭС по источникам публикаций

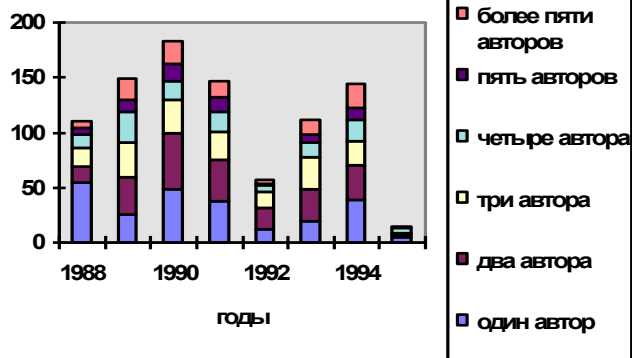


Большинство публикаций по медицинским экспертным системам выполнено в изданиях следующих стран: США - 41.3%, Нидерланды - 12.4%, Германия - 12.2%, Англия - 11.6%. Однако по разделу "Ак и Г" распределение выглядит иначе: Англия - 30.8%, США - 20.5%, Германия - 12.6%, Нидерланды - 5.1%. Абсолютное большинство публикаций выполнено на английском языке как во всем

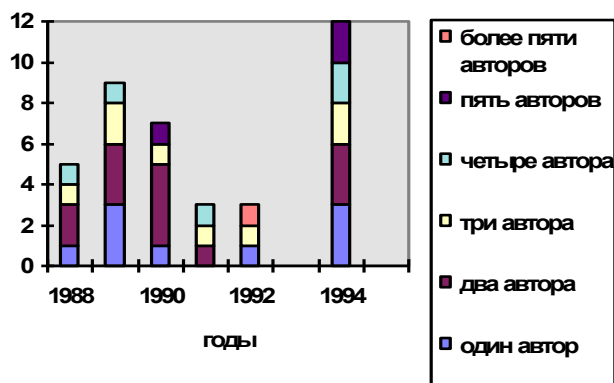
исследованном нами потоке (82.7%), так и по разделу "Ак и Г" (82.1%). Наибольшее количество публикаций по ЭС в медицине сосредоточено в следующих изданиях: "Comput. Method. Program. Biomed." (Нидерланды - Ирландия) - 15.6%, "Proc. Annu. Symp. Comput. Appl. Med. Care" (США) - 14.5%, "Method. Inf. Med." (Германия) - 13.7%, "Med.

Inf. Lond.” (Англия) - 10.4%, “Int.J.Biomed.Comput.” (Англия) - 7.1%. Статьи же по ЭС в “Ак и Г” опубликованы в “Int. J. Biomed. Comp.” (Англия) - 12.8%, “Baillieres. Clin. Obstet. Gynecol.” (Англия) - 10.2%, “Arch. Gynecol. Obstetr.” (Англия) - 7.7%, в остальных же изданиях опубликовано по одной статье.

Распределение публикаций о МЭС по количеству авторов



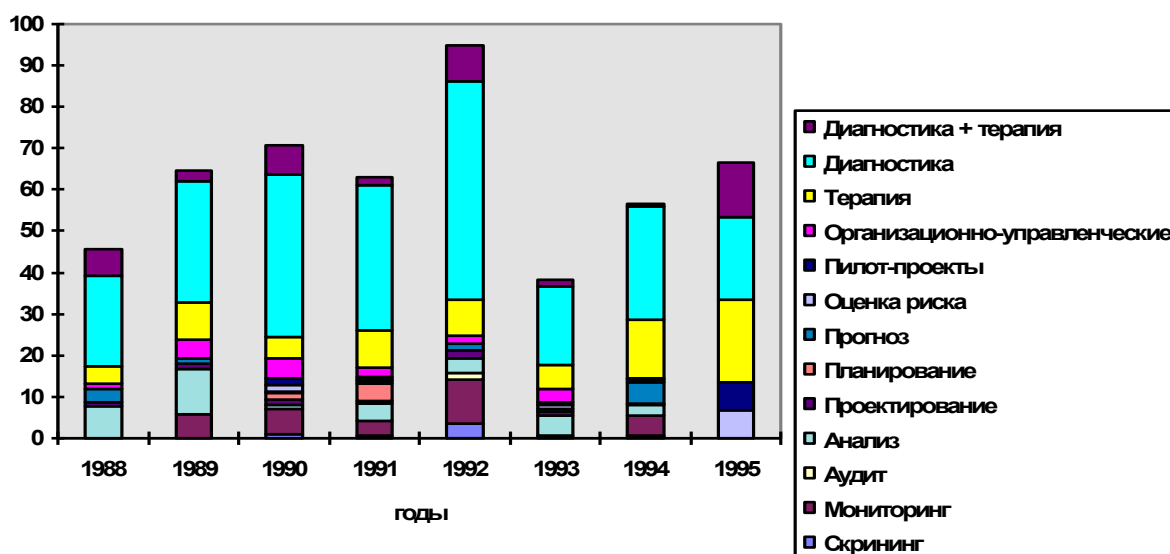
Распределение публикаций об ЭС раздела "Ак и Г" по количеству авторов



К сожалению, рефераты БД MEDLINE не позволяют оценить сложность описанных в них экспертных систем хотя бы по такому критерию, как объем баз знаний или затраты на создание ЭС (временные или в единицах измерения “человеко-лет”), чтобы найти зависимость между сложностью ЭС и численностью авторского коллектива публикации. Однако выполненное исследование подтверждает тезис о том, что экспертные системы являются сложными информационными системами, особенно применительно к медицине.

Распределение медицинских экспертных систем, описанных в публикациях, по функциональному назначению выглядит следующим образом:

Распределение МЭС по функциональному признаку (в %)

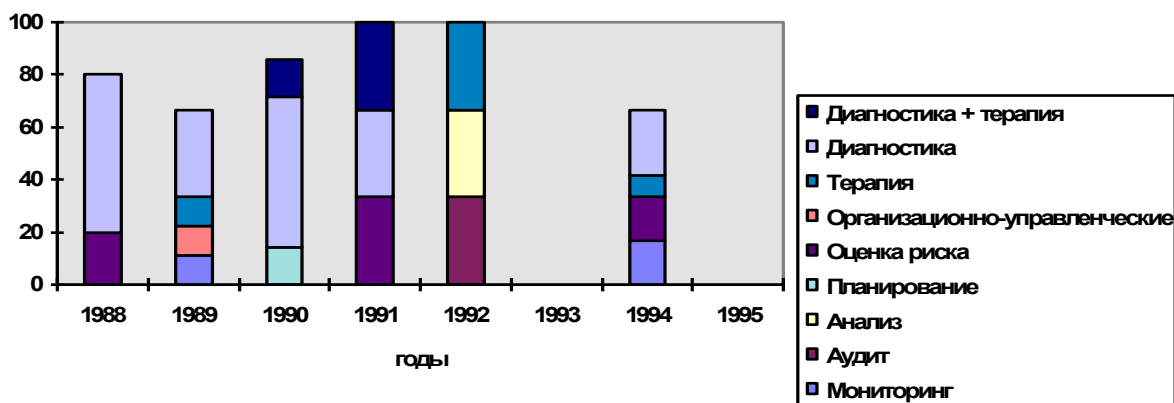


Абсолютное большинство МЭС являются диагностическими и терапевтическими (или их сочетанием), так как именно в диагностике и терапии накоплен определенный опыт формализации и отторжения знаний от экспертов. Интересующая же нас группа

МЭС, посвященная анализу риска, невелика: количество публикаций в 1990г. - 3 (1,6%), в 1992г. - 4(2,1%), в 1994г. - 5 (3%).

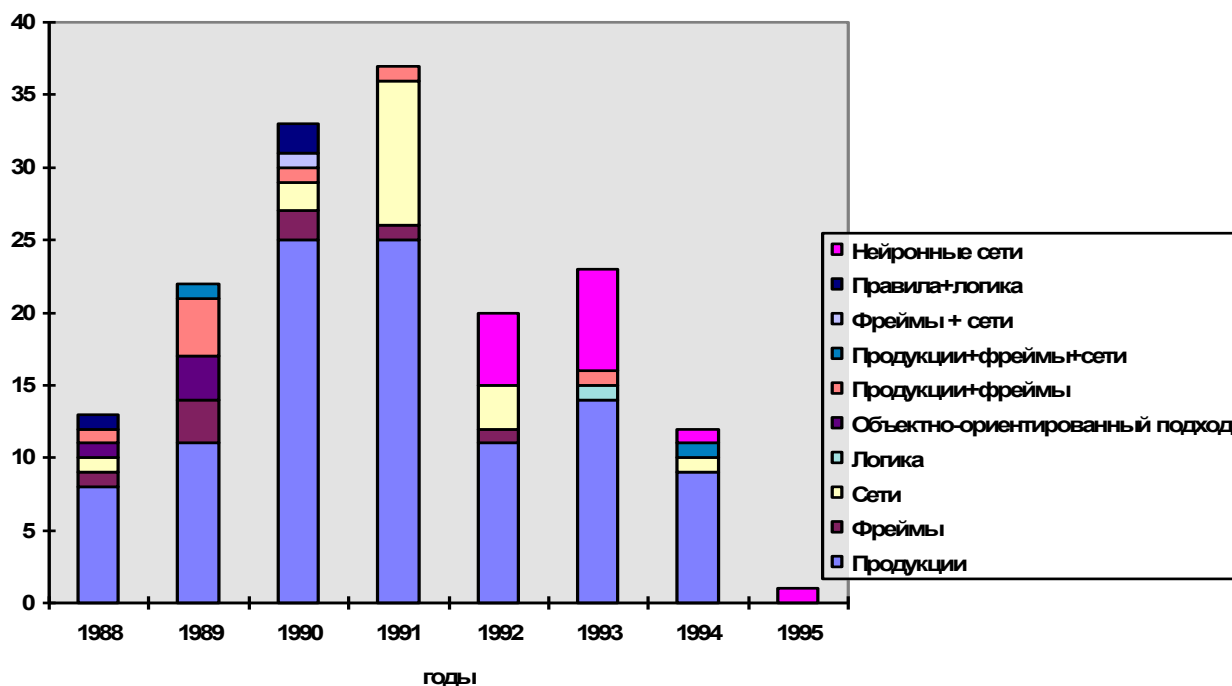
По разделу “Ак и Г” это же распределение таково: в этом разделе медицины отсутствуют ЭС проектирования, скрининга, прогноза и пилот-проекты, а оценке риска посвящено следующее количество публикаций: 1990г. - 1 (33,3 %), 1994г. - 2 (16,7 %).

Распределение ЭС по разделу "Ак и Г" по функциональному назначению (в %)



Анализ способов представления знаний в МЭС показал, что в них используются все известные формализмы (продукционные правила, фреймы, сети, логика, объектно-ориентированный подход, нейронные сети), а также комбинации различных способов представления знаний.

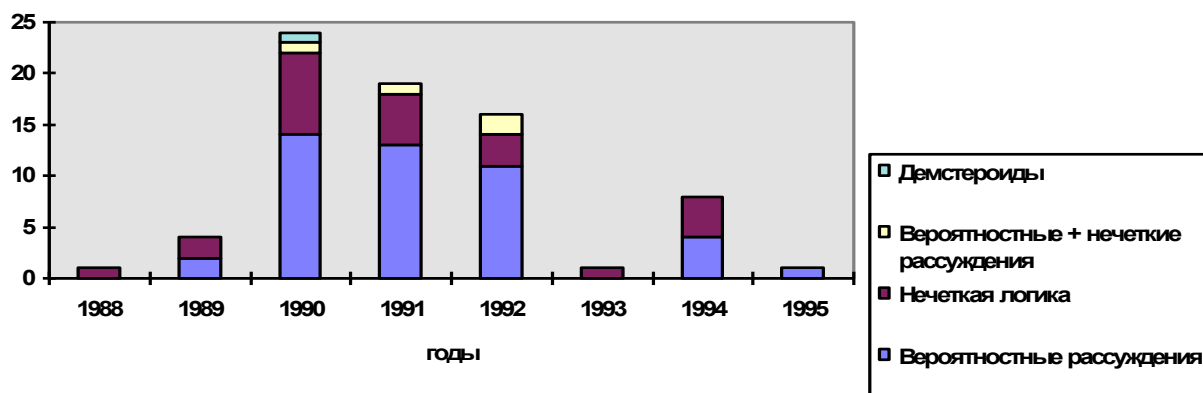
Распределение МЭС по способам представления знаний



В ЭС по “Ак и Г” для представления знаний использованы только продукции (1988 г. - 40 %, 1989 г. - 11,1%, 1994г. - 25%) и нейронные сети (1994 г. - 16,7 %).

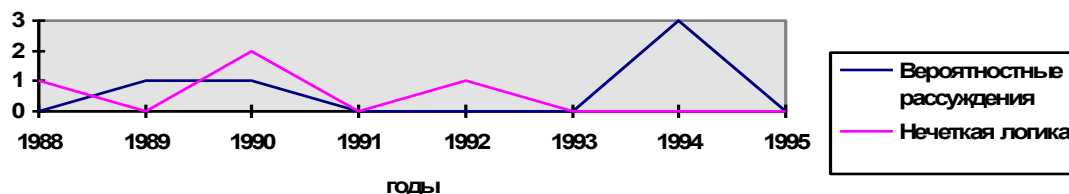
Анализ способности МЭС к неточным рассуждениям показал, что часть из них использует либо вероятностные рассуждения, либо нечеткую логику:

Распределение МЭС по способности к неточным рассуждениям



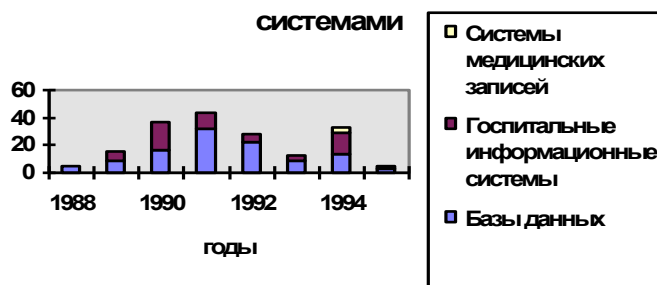
Распределение же способности к неточным рассуждениям ЭС по разделу “Ак и Г” выглядит так:

Распределение ЭС по разделу "Ак и Г" по способности к неточным рассуждениям



Часть МЭС интегрирована с базами данных и другими медицинскими информационными системами:

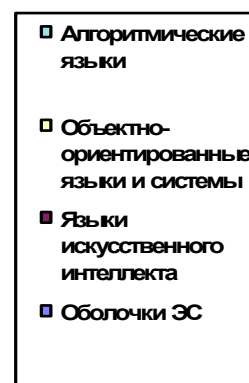
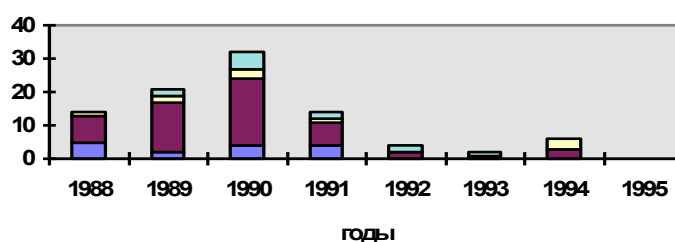
Интеграция МЭС с другими медицинскими информационными системами



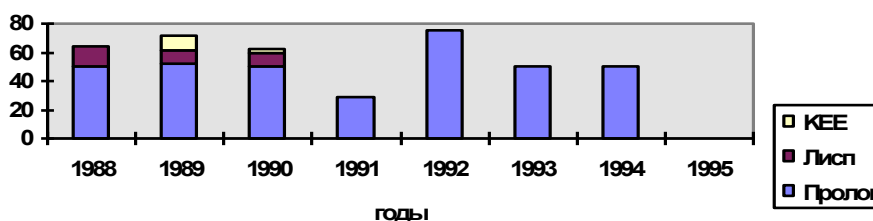
Процентное отношение МЭС, интегрированных с другими информационными системами составляет в 1988г. 5.4%, в 1989г. - 9.8 %, в 1990г. - 20 %, в 1991г. - 25.7 %, в 1992г. - 49.1 %, в 1993г. - 10.4%, в 1994г. - 20.1 %, в 1995г. - 33.3 %.

Анализ публикаций исследованного нами потока показал, что при разработке МЭС использованы инструментальные средства различного уровня:

Использование инструментальных средств для реализации МЭС



Использование языков искусственного интеллекта в МЭС (в % к общему количеству инструментальных средств)



Среди языков искусственного интеллекта как средства реализации МЭС абсолютное лидерство принадлежит Прологу.

По разделу “Ак и Г” распределение инструментальных средств таково: оболочки ЭС (1988г. - 60%, 1990г. - 14,3%), языки искусственного интеллекта (1990г. - 14,3%), алгоритмические языки (1990г. - 14,3%).

3. МОДЕЛЬ ФАКТОРОВ РИСКА ОТКЛОНЕНИЙ В РЕПРОДУКТИВНОЙ ФУНКЦИИ ДЕВОЧЕК-ПОДРОСТКОВ

3.1. Модель факторов риска в информационных системах скрининга девочек-подростков и отделения гинекологии.

Некоторые данные профилактического осмотра позволяют отнести девочек-подростков к группам риска. Кроме тех случаев, когда медработник, проводящий осмотр, ставит диагноз (НАРУШЕНИЕ МЕНСТУАЛЬНОГО ЦИКЛА: аменорея, опсоменорея и т.д.), имеется ряд показателей, которые необязательно являются признаком заболевания, но могут ему сопутствовать.

Нами предложено выделять следующую иерархию данных в модели факторов риска отклонений в репродуктивной функции девочек-подростков:

- группы наблюдений;
- названия наблюдений;
- характеристики наблюдений;
- список значений характеристики.

Модель факторов риска отклонений в репродуктивной функции девочек-подростков опишем в терминах теории множеств и реляционной алгебры.

Пусть:

G - множество допустимых групп наблюдений:

$$G = \{ g[i], i = 1, KG \}, \quad (3.1.1)$$

где KG - множество допустимых групп наблюдений;

N - множество допустимых наблюдений в i-й группе:

$$N = \{ n[i,j], j = 1, KN \}, \quad (3.1.2)$$

где KN - количество допустимых наблюдений в i-й группе;

X - множество допустимых характеристик j-го наблюдения в i-й группе;

$$X = \{ x[i,j,L], L = 1, KX \}, \quad (3.1.3)$$

где KX - количество допустимых характеристик j-го наблюдения в i-й группе;

V - множество допустимых значений L-й характеристики j-го наблюдения в i-й группе:

$$V = \{ v[i,j,L,k], k = 1, KV \}, \quad (3.1.4)$$

где KV - количество допустимых значений L-й характеристики j-го наблюдения в i-й группе.

Тогда декартового произведения указанных множеств

$$B = G * N * X * V \quad (3.1.5)$$

есть множество всех возможных комбинаций характеристик наблюдений по всем группам наблюдений.

Результат S-го единичного опроса p-й пациентки можно представить в виде:

$$O[s,p] = \text{projected_to } v[i1,j1,L1,k1], v[i2,j2,L2,k2], \dots, v[in,jn,Ln,kn]$$

$$(B(\text{selected_on } G = \langle \text{название группы} \rangle \wedge N = \langle \text{название наблюдения} \rangle \wedge X = \langle \text{название характеристики} \rangle)) \quad (3.1.6)$$

Модель риска p-й пациентки есть объединение таких единичных опросов:

$$m[p] = \{ O[s,p], s = 1, KS \}, \quad (3.1.7)$$

где KS - количество единичных опросов пациентки.

В случае модели для скрининга это результат профилактического осмотра пациентки. В случае модели для отделения гинекологии сюда включаются как первичный опрос, так и все повторные при мониторинге.

Тогда модель риска есть

$$M = \text{union } m[p], p = 1, KP; \quad (3.1.8)$$

где KP - количество пациенток.

3.1.1. Оценка модели факторов риска в информационной системе скрининга девочек-подростков.

Количественные показатели этого уровня модели таковы.

Количество групп наблюдений - 9: наличие гирсутизма; степень гирсутизма; отношение роста к весу (меньше или больше нормы); явно выраженное ожирение или, наоборот, слишком низкая масса тела; отсутствие вторичных половых признаков; антропометрические данные; высокая степень гипертрихоза; наличие трофических изменений кожи; тяжелые или хронические заболевания, не связанные с половой системой, перенесенные девочкой-подростком.

В каждой группе от 2 до 5 наблюдений. Всего 23 наблюдения, которые содержат от 1 до 4 характеристик, каждая из которых содержит от 1 до 5 значений.

Например, в группе "ПОЛОВАЯ ФОРМУЛА" имеется 4 вида наблюдений. Вид наблюдения "РАЗВИТИЕ ВОЛОС НА ЛОБКЕ" имеет 2 характеристики. Характеристика "ВОЛОСЫ ЕСТЬ" имеет 3 значения, характеризующие степень развития волос.

3.1.2. Оценка модели факторов риска в информационной системе отделения гинекологии.

На уровне отделения детской и подростковой гинекологии информационная система (банк данных) об отклонениях в репродуктивной функции девочек-подростков включает факторы риска, учитываемые на этапе скрининга. Однако здесь для наиболее полной дифференциации заболевания и его стадии в модель включаются дополнительные факторы.

Приведем количественные показатели названного уровня модели, которые позволяют оценить **степень сложности модели**.

Количество групп наблюдений - 23: антропометрические данные; течение беременности матери; течение родов матери; анамнез пациентки; характеристики менархе; характеристики менструального цикла; общее состояние организма; состояние щитовидной железы; гирсутизм; вторичные половые признаки; наружные гениталии; вагиноскопия; диагноз основной, сопутствующий; анализ крови; анализ мочи; анализ выделений; консультации; кольпоцитология; базальная температура; генетические обследования; УЗИ - обследования; гормональные обследования; экскреция гормонов с мочой.

В каждой группе от 2 до 39 наблюдений. Всего 170 наблюдений, которые содержат от 1 до 9 характеристик, каждая из которых содержит от 1 до 10 значений.

Например, в группе "ТЕЧЕНИЕ БЕРЕМЕННОСТИ МАТЕРИ" имеется 4 вида наблюдений. Вид наблюдения "ХАРАКТЕРИСТИКИ БЕРЕМЕННОСТИ" имеет 2 характеристики. Характеристика беременности "ОСЛОЖНЕННАЯ" имеет 9 осложнений. Осложнение "ЭКСТРАГЕНИТАЛЬНЫЕ ЗАБОЛЕВАНИЯ" имеет 10 значений этого осложнения.

3.2. Модель факторов риска в экспертной системе диагностики и назначения лечения.

Модель факторов риска в экспертной системе принятия решений по назначению консервативного курса лечения (ККЛ) заболеваний синдрома склеро-кисто́за яичников (ССКЯ) девочки-подростка формально может быть представлена в виде следующего отображения:

$$S : D \rightarrow L, \quad (3.3.1)$$

где S - система назначения лечения;

D - множество входов (диагнозы и наблюдения);

L - множество выходов (назначаемые лечения).

Отображение имеет концептуальный характер, поэтому модель представляется в виде зависимостей причинно-следственного характера (здесь D - множество причин, а L - следствия). Множество выходных назначаемых лечений описывается иерархической структурой. На верхнем уровне находятся обобщенные назначения лечения. В процессе движения вниз по дереву решения назначаемые лечения конкретизируются и уточняются (с помощью проверки наличия/отсутствия дополнительных наблюдений).

Среди заболеваний ССКЯ диагностируются 4 группы заболеваний по возрастанию тяжести заболевания (классификация предоставлена Матыщиной Л.А.):

- группа O1 - нестойкая опсоменорея;

- группа A - аменорея;

- группа O2 - стойкая опсоменорея;

- группа Ю - ювенильные маточные кровотечения.

В каждой из групп заболеваний возможно назначение трех курсов лечения (I, II, III), причем в результате выполнения терапевтического курса исходный диагноз пациентки может измениться. Граф переходов диагнозов после выполнения курсов терапии:

Ю --> O2 (I)

Ю --> A (I, II)

O2 --> A (I)

O2 --> O1 (I, II)

A --> O1 (I)

O1 --> выздоровление (II, III)

(3.3.2)

Здесь наименование заболевания слева от стрелки --> соответствует узлу-источнику графа, наименование заболевания справа от стрелки - узлу-приемнику графа, а

списки номеров курсов лечения (в скобках) указывают применяемый вид консервативного лечения. При назначении курса лечения используются факторы риска из модели информационной системы отделения гинекологии (гирсутизм, ожирение), а также дополнительные факторы (нарушения центральной нервной системы, организм ослаблен и т.д.).

БЛАГОДАРНОСТИ

Авторы выражают свою благодарность проф. Слепцову А.И. (ДонГТУ) и проф. Чайке В.К. (ДРЦ ОМД) за поддержку выполнения этой работы, зав. отделением детской и подростковой гинекологии ДРЦ ОМД Матыциной Л.А. за постановку задач и предоставление материалов предметной области. Также мы выражаем благодарность студентам факультета вычислительной техники и информатики ДонГТУ, принимавших непосредственное участие в программной реализации информационной системы MEDLINE PROFILER по статистической обработке данных БД MEDLINE, с помощью которой были проведены исследования: Шичинову Р., Земскому П., Горбуновой О., Ивановой Ю., а также Поль И. и Кобзиной Ю.