

УДК 004.021

**Д.В. Юрьев** (студент), **Н.А. Ярошенко** (канд. техн. наук, доц.)

Донецкий национальный технический университет

E-mail: [dimayriev@email.ua](mailto:dimayriev@email.ua)

## **ИСПОЛЬЗОВАНИЕ СОВРЕМЕННЫХ МЕТОДОВ ПОСТАНОВКИ ДИАГНОЗА ПО БИОХИМИЧЕСКОМУ АНАЛИЗУ КРОВИ**

Рассмотрена общая постановка проблемы, основные положения, которые позволяют рассматривать заболевание с биохимической позиции, и принципы лабораторного исследования, а также описание методов, которые позволяют автоматизировать процесс постановки предварительного диагноза.

**Ключевые слова:** показатель, нормы, математическая модель, база знаний, коэффициент уверенности.

### ***Введение***

Биохимический анализ отличается высокой степенью достоверности, поэтому он является одним из основных способов для определения стадии заболевания и постановки диагноза. Биохимический анализ крови — это лабораторный метод исследования, использующийся в медицине, он отражает функциональное состояние отдельных органов и систем организма человека. По результатам анализов оценивают работу многих внутренних органов в частности печени, почек, а так же воспалительные процессы, нарушение водно-солевого обмена и дисбаланс микроэлементов. Однако в настоящее время обработка результатов анализа и постановка диагноза осуществляется врачами вручную. Поэтому, автоматизация постановки предварительного диагноза заболеваний по биохимическому анализу крови и мочи, является актуальной.

### ***Общая постановка проблемы***

Биохимический анализ крови выдает двадцать пять показателей [1], по которым можно оценить состояние человека. Показатели анализа следующие: общий белок; общий билирубин; прямой билирубин; аспаратаминотрансфераза (АсАТ); аланинаминотрансфераза (АлАТ); щелочная фосфатаза; амилаза; гамма-глутамилтрансфераза (Гамма-ГТ); креатинфосфокиназа; холестерин; триглицериды; глюкоза; мочевины; креатинин; мочевиная кислота; железо; магний; фосфор; кальций; хлориды;

лактатдегідрогеназа (ЛДГ); калий; натрий; С-реактивный белок; ревматоидный фактор; антистрептолизин-О.

Существуют нормы биохимического анализа крови по которым можно определить наличие или отсутствие заболевания. Это статистически установленные показатели здоровых людей, которые принимаются как эталонные. Отклонения показателей от нормы являются признаками наличия заболевания, и могут быть положены в основу биохимического анализа крови.

Новизной работы является разработка математической модели и программного обеспечения, для автоматизации процесса постановки диагноза. Практическая ценность работы состоит в том, что система сможет автоматически обрабатывать данные, и по разработанной модели ставить предварительный диагноз, что ускоряет и оптимизирует работу медперсонала.

Проведенный анализ существующих методов показал, что приемлемым и для достижения сформулированной цели являются система поддержки принятия решения и генетический алгоритм.

### ***Система поддержки принятия решения для постановки диагноза***

Основой любой системы поддержки принятия решений (СППР) является совокупность знаний, структурированная в целях упрощения процесса принятия решения.[3] Для специалистов в области искусственного интеллекта термин знания означает информацию, которая необходима программе, чтобы она вела себя «интеллектуально». Эта информация принимает форму фактов и правил. Факты и правила в СППР не всегда либо истинны, либо ложные. Иногда существует некоторая степень неуверенности в достоверности факта или точности правила. Если это сомнение выражено явно, то оно называется «коэффициентом уверенности».

Коэффициент уверенности – это число, которое означает вероятность или степень уверенности, с которой можно считать данный факт или правило достоверным или справедливым. Данный коэффициент является оценкой степени доверия к решению, выдаваемому системой. Теория уверенности предлагает понятия доверия и недоверия. Эти понятия являются независимыми друг от друга и таким образом не могут быть объединены тем же путем, как вероятности, но они могут быть объединены и представлены в математической модели постановки диагноза, представленного ниже.

Принципы работы системы поддержки принятия решений, основанной на знаниях, иллюстрируются на рис.1. Пользователь передает в систему факты или, в нашем случае, симптомы и получает в качестве

результата экспертные знания или диагноз. По своей структуре экспертная система подразделяется на два основных компонента - базу знаний и машину логического вывода. База знаний содержит знания (показания), на основании которых машина логического вывода формирует заключения (диагноз). Эти заключения представляют собой ответы системы поддержки принятия решений на запросы пользователя, желающего получить экспертные знания.

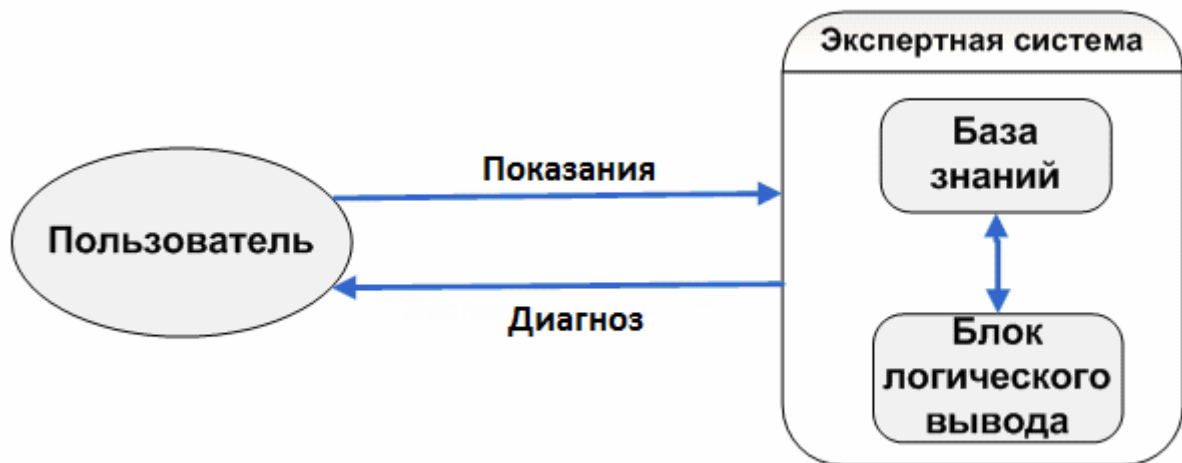


Рисунок 1 — Основные принципы функционирования системы поддержки принятия решений

***К математической модели постановки диагноза предъявляются следующие требования***

- обеспечение адекватности и точности получаемых результатов;
- обеспечение быстродействия при обработке данных.

Обеспечение адекватности и точности получаемых результатов заключается в СППР. Из выше изложенного следует, что теория уверенности предлагает понятия доверия и недоверия. Эти понятия являются независимыми друг от друга и таким образом не могут быть объединены тем же путем, как вероятности, но они могут быть объединены в соответствии со следующей формулой:  $KУ(V,C) = МД(V,C) - МН(V,C)$ , где КУ - коэффициент уверенности, МД - мера доверия, МН - мера недоверия, V – вероятность, C - свидетельство или событие. Возможны несколько способов достижения цели, каждый с различными КУ для данного множества фактов. Когда мы имеем систему, основанную на знаниях с несколькими взаимосвязанными правилами, каждое из которых делает то же заключение, но с различным коэффициентом уверенности, тогда каждое правило может рассматриваться как часть свидетельства, которое поддерживает совместное заключение. Результат объединения вычисляется как:  $KУ(П1,П2) = KУ(П1) + KУ(П2)[1 - KУ(П1)]$ .

Обеспечение быстрого действия при обработке данных заключается от сложности построения диагностирующих производственных экспертных систем (ЭС), и аппаратных возможностях СКС.

Рассмотрим один из подходов к созданию диагностирующих производственных ЭС. Соответствующие интерпретации могут быть разными, и потому область применимости обсуждаемых конструкций оказывается достаточно широкой. Заданное огромное количество фактов

$$A = \{a_{ij}\} \cup \{q_i\} = (a_1, a_2, \dots, a_n),$$

которые состоят из элементов двух типов. Элементы  $a_{ij}$  определяют обычные декларативные знания из конкретной предметной области. Элементы  $q_i$  определяют вид взаимодействия с внешней средой и в данном случае являются вопросами пользователю в виде альтернативного меню:

$$q_i = (a_{i1}, \dots, a_{ik})$$

Некоторые из  $q_i$ , имеют другой смысл - результирующих выводов или диагнозов, оформленных в виде соответствующих сообщений пользователю. Продукции в данной системе имеют вид

$$a_{ij} \rightarrow q_m = \{a_{m1}, \dots, a_{mk}\}$$

Все огромное количество фактов и продукций организованы в некоторую систему, представленную в виде графа "ИЛИ". Фрагмент такого диагностирующего графа с вершинами-диагнозами  $q_9, q_{10}, q_{11}, q_{12}$  (терминальными вершинами) представлен на рис.2 .

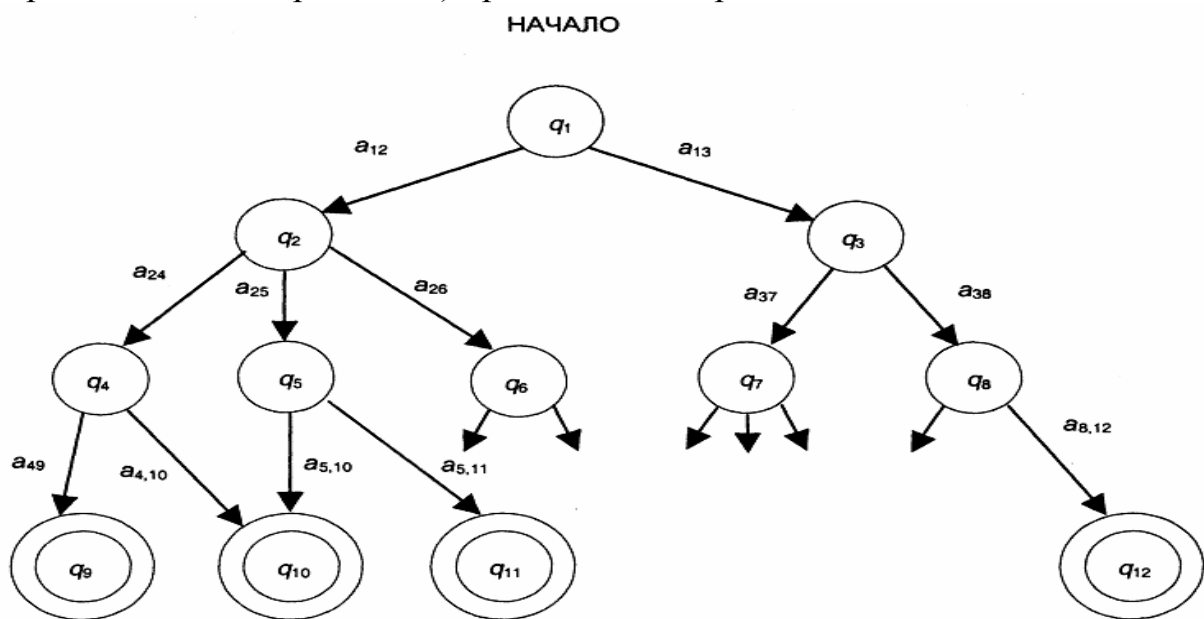


Рисунок 2 — Диагностирующий граф

**Недостатки системы поддержки принятия решения сводятся к следующему:**

- Единственный источник данных, потенциально сужающий круг вопросов, на которые может ответить система

- Большая нагрузка на операционную систему. Сложные запросы могут привести к остановке работы операционной системы, что весьма нежелательно

### ***Возможность использования генетического алгоритма для постановки диагноза***

Генетический алгоритм (ГА) представляет собой метод оптимизации, основанный на концепциях естественного отбора и генетики. В этом подходе переменные, характеризующие решение, представлены в виде ген в хромосоме. ГА оперирует конечным множеством решений (популяцией) - генерирует новые решения как различные комбинации частей решений популяции, используя такие операторы, как отбор, рекомбинация (кроссинговер) и мутация. Новые решения позиционируются в популяции в соответствии с их положением на поверхности исследуемой функции.

Идею ГА подсказала сама природа и работы Дарвина. Делается предположение, что если взять 2 вполне хороших решения задачи и каким-либо образом получить из них новое решение, в нашем случае наиболее вероятный диагноз, то будет высокая вероятность того, что новое решение получится хорошим или даже более лучшим, то есть вероятный диагноз с высоким коэффициентом уверенности .

Для реализации этого используют моделирование эволюции (естественного отбора) или если проще - борьбы за выживание. В природе, по упрощенной схеме, каждое животное стремится выжить, чтобы оставить после себя как можно больше потомства. Выжить в таких условиях могут лишь сильнейшие. Тогда нам остается организовать некоторую среду - популяцию, населить её решениями - особями, и устроить им борьбу. Для этого нужно определить функцию, по которой будет определяться сила особи - качество предложенного ею решения. Основываясь на этом параметре можно определить каждой особи, количество оставляемых ею потомков, или вероятность того, что эта особь оставит потомка. Причем, не исключен вариант, когда особь со слишком низким значением этого параметра умрёт.

Допустим нам нужно оптимизировать некоторую функцию  $F(X_1, X_2, \dots, X_n)$ . Пусть мы ищем её глобальный максимум. Тогда, для реализации ГА нам нужно придумать, как мы будем хранить решения. По сути, нам нужно поместить все  $X_1$ - $X_n$  в некоторый вектор, который будет играть роль хромосомы. Один из наиболее распространенных способов - кодировать значения переменных в битовом векторе. Например, выделим каждому иксу по 8 бит. Тогда наш вектор будет длины  $L=8*n$ . Для простоты будем считать, что биты лежат в массиве  $X[0..L-1]$ . Пусть каждая особь состоит из массива  $X$  и значения функции  $F$  из переменных, извлеченных из этого массива.

Тогда ГА будет состоять из следующих шагов:

Генерация начальной популяции - заполнение популяции особями, в которых элементы массива  $X$  (биты) заполнены случайным образом.

При выборе родительской пары всегда использую элитный отбор, то есть берут  $K$  особей с максимальными значениями функции  $F$  и составляют из них все возможные пары  $(K*(K-1)/2)$ .

Кроссинговер - берем случайную точку  $t$  на массиве  $X$  ( $0..L-1$ ). Теперь, все элементы массива с индексами  $0-t$  новой особи (потомка) заполняем элементами с теми же индексами, но из массива  $X$  первой родительской особи. Остальные элементы заполняются из массива второй родительской особи. Для второго потомка делается наоборот - элементы  $0-t$  берут от второго потомка, а остальные - от первого.

Новые особи с некоторой вероятностью мутируют - инвертируется случайный бит массива  $X$  этой особи. Вероятность мутации обычно полагают порядка 1%. Полученные особи-потомки добавляются в популяцию после переоценки. Обычно новую особь добавляют взамен самой плохой старой особи при условии, что значение функции на новой особи выше значения функции на старой-плохой особи.

Если самое лучшее решение в популяции нас не удовлетворяет, то переход на шаг 2. Хотя, чаще всего этого условия нет, а итерации ГА выполняются бесконечно.

Вообще, если строго придерживаться правил, то ГА должен содержать еще такие шаги как отбор особей для размножения и генерация пар из отобранных особей. При этом каждая особь может быть задействована в одной и более паре, в зависимости от используемого алгоритма.

### ***Недостатки генетических алгоритмов сводятся к следующему:***

- Время исполнения функции оценки велико;
- Конфигурация является не простой (кодирование решения).

В результате проведенного анализа для постановки диагноза принят метод системы поддержки принятия решения. Для автоматизации процесса постановки диагноза разработано соответствующее созданию программное обеспечение и создана база данных.

На рисунке 4 приведён пример результата работы программного обеспечения в виде экранной формы.

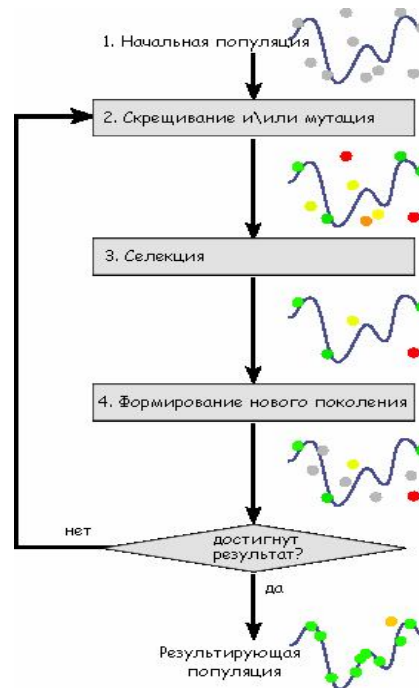


Рисунок – 3 Схема работы генетического алгоритма

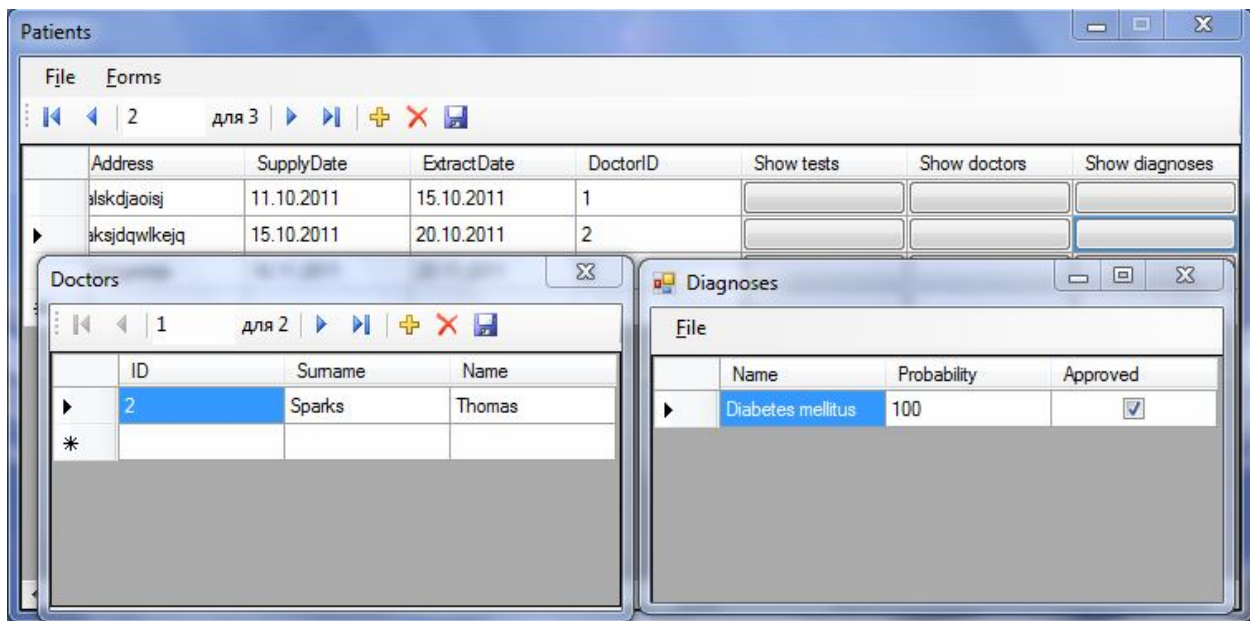


Рисунок – 4 Результат работы программного обеспечения

Из экранной формы следует, что автоматически поставленный системой предварительный диагноз подтверждён доктором.

### **Вывод**

Благодаря предложениям, выдвигаемым СППР, открываются новые подходы к решению повседневных и нестандартных задач. В дальнейших исследованиях предполагается использование генетического алгоритма.

## **Список литературы**

1. Цыгаренко А.Я. Клиническая биохимия: учеб. пособ / А.Я. Цыгаренко, В.И. Жуков, В.В. Мясоедов. - М.: "Триада-Х", - 2002. - 504с.
2. Данилова Л.А. Анализ крови и мочи; 4-е изд., исправ / Л.А. Данилова. – СПб.: "Салит-Медкнига", 2003. - 128с.
3. Черноруцкий И.Г. Методы принятия решений / И.Г. Черноруцкий. – СПб.: "БХВ-Петербург", 2005. – 416 с.

*Надійшла до редакції 01.11.2011.*

*Рецензент: канд. техн. наук, доц. Коренев В.Д.*

**Д.В. Юрьев, Н.А. Ярошенко**

Донецький національний технічний університет

**Використання сучасних методів постановки діагнозу по біохімічному аналізу крові.** Розглянуто загальну постановку проблеми, основні положення, які дозволяють розглядати захворювання з біохімічної позиції, і принципи лабораторного дослідження, а також опис методів, які дозволяють автоматизувати процес постановки попереднього діагнозу.

**Ключові слова:** показник, норми, математична модель, база знань, коефіцієнт впевненості.

**D.V. Yur'ev, N.A. Yaroshenko**

Donetsk National Technical University

**Using Modern Methods of Diagnosis by Biochemical Analysis of Blood.** We consider a general formulation of the problem, the basic provisions that allow the disease to examine the biochemical positions and principles of laboratory research, as well as a description of methods that allow you to automate the process of setting a preliminary diagnosis.

**Keywords:** the index, norms, mathematical model, the knowledge base, the coefficient of confidence.