

ПРОГНОЗ ЛЕЧЕНИЯ ГЕПАТИТА

И.А. Дашенко, И.С. Репенко, А.С. Патес

На сегодняшний день разработка интеллектуальных систем анализа данных приобретает все большие масштабы. Подобные инновационные проекты реализуются и в медицине, в частности, в области терапии гепатита С. Интерферонотерапия вирусного гепатита является достаточно дорогостоящим методом лечения и, более того, не гарантирует положительного результата. Оценка возможных результатов терапии имеет важное значение, как для больных, так и для врачей. Прогнозирование ответа на лечение предполагает оценку его вероятности.

Факторный анализ (выявление влияния факторов или группы факторов на вероятность успешного исхода лечения) имеет большие перспективы в медицине, однако до сих пор отсутствуют данные о значимости конкретных признаков. Прогностическое значение исходных характеристик и факторов, которые оценивают перед началом терапии, важно подтвердить с помощью методов математической статистики. В этой связи имеет большое значение создание компьютерной программы, которая позволит комплексно оценить у каждого конкретного больного вероятность ответа на терапию на основании исходных данных.

Сибирско-Американский факультет менеджмента принял активное участие в разработке программы для Иркутского государственного медицинского университета, с целью создания системы аналитической базы данных прогнозирования вылечиваемости гепатита.

В 2007 году группой разработчиков А.Волковым и Е.Любочкиным была создана база данных, включающая одну таблицу (Med_2007), в которой хранятся данные обо всех пациентах, больных гепатитом, которые прошли лечение. Около 60 файлов, написанных на языке программирования РНР, составили основу для программной оболочки, включающей в себя пользовательский интерфейс и скрипты для изменения информации в базе данных. Также группа разработчиков занялась проведением первичного однофакторного и двухфакторного анализа по факторам, подошедшим по репрезентативности выборки. Результаты, полученные группой разработчиков А.Волковым и Е.Любочкиным, вы можете найти в материалах научно-практической конференции 2007 года.

В 2008 году новой группой разработчиков в составе И.Дашенко, И.Репенко и А.Патеса проведена следующая работа:

- Уточнены результаты отбора факторов. Количество факторов для анализа расширено. Добавлены факторы, подходящие по репрезентативности. Таковыми явились

показатели крови – эритроциты, эпителй, лейкоциты и СОЭ. Также добавлены факторы, непосредственно, по данным литературы, влияющие на процесс вылечиваемости – генотип вируса, вирусная нагрузка, фиброз, пол, возраст и вес;

- Усовершенствован пользовательский интерфейс;
- Добавлены страницы с полезной информацией для клиента;
- Прогнозирование на основе имеющихся данных одно- и двухфакторного анализа.

Результаты однофакторного, двухфакторного и смешанного анализов представлены здесь: <http://stud.buk.irk.ru:8100/2007/proj1/SAS/prognosis1.php>; <http://stud.buk.irk.ru:8100/2007/proj1/SAS/prognosis.php>

С подробным описанием методик и результатов исследований можно ознакомиться в сборнике статей научно-практической конференции 2008 года.

За 2009 год в базу было внесено некоторое количество новых данных, а именно данные по 109 новым пациентам. К сожалению, данные были внесены не по всем факторам, что усложнило задачу анализа данных. Был произведен пересчет старых коэффициентов формул прогноза с учетом новых данных.

В 2009 году был реализован прогноз по ключевым факторам, которые были определены специалистами медицинского университета. Согласно экспертной оценке, были особо выделены генотип вируса, вирусная нагрузка, масса тела больного, индекс массы тела, группа крови и резус- фактор, длительность заболевания, возраст больного и пол. В новом прогнозе учитываются 4 фактора из вышеприведенных: пол, возраст, длительность заболевания и вирусная нагрузка. Другие факторы были предварительно отброшены в силу не репрезентативности, т.к. содержали много полей с некорректными данными.

Таким образом, на данный момент доступны 3 прогноза:

1. По факторам крови:

- Лейкоциты
- Эритроциты
- СОЭ
- Эпителй

2. По факторам физиологии:

- Генотип
- Вирусная нагрузка
- Пол
- Фиброз

3. По факторам особо значимым для прогноза:

- Пол
- Возраст
- Длительность заболевания
- Вирусная нагрузка

Ниже приведены коэффициенты (указан процент вылечиваемости по категориям), которые были использованы для подсчета результатов прогноза.

	вирусная нагрузка	макс	67.03%
		сред	69.81%
		мин	66.67%
пол	жен		66.67%
	муж		69.39%
возраст	16-37		77.92%
	38-58		60.66%
	59-78		50.00%
длительность заболевания	0-3		83.33%
	4-8		79.31%
	9-18		50.00%
	19-30		46.67%

Также в прогнозе реализован подход подборки весов. Т.е. специалист, зная коэффициенты, может с большей точностью определить результат. На данном этапе этот подход реализован на уровне однофакторного анализа.

Ссылка на этот прогноз http://www.cgc-hse.buk.irk.ru/prognosis/SAS/prognosis/_analysis-_NEW.php.

Для окончательного анализа различий во влиянии факторов на вылечиваемость разработчиками был предложен метод Саати – процедура получения оценки проблемы на основе группового мнения экспертов. Данный метод позволяет путем попарного сравнения факторов выявить общий вес каждого фактора в данной группе. Интерфейс данной методики представлен в Microsoft Excel в следующем виде:

Критерий	Возраст	Пол	ИМТ	Длит.	Генотип	Вир.наг.	П _i	Корень бст. (П _i)	P _i
Возраст	1,00	3,00	3,00	1,00	0,33	0,33	1	1,000	0,154
Пол	0,33	1,00	5,00	3,00	1,00	0,33	2	1,089	0,168
ИМТ	0,33	0,20	1,00	0,33	0,20	1,00	0	0,405	0,063
Длит.	1,00	0,33	3,00	1,00	0,33	5,00	2	1,089	0,168
Генотип	3,00	1,00	5,00	3,00	1,00	0,33	15	1,571	0,242
Вир.наг.	3,00	3,00	1,00	0,20	3,00	1,00	5	1,325	0,204

В каждой из ячеек выставляется оценка, отражающая превалирование критерия в строке над критерием в столбце (или наоборот). Путем ряда вычислений в правой колонке таблицы были получены соответствующие весовые коэффициенты для каждого из рассматриваемых критериев.

Для следующих этапов продолжения проекта мы видим:

1. Сбор мнений экспертов по проблеме выставления весовых коэффициентов по методу Саати;
2. Разработка методов корректировки (научения) системы при появлении новых данных;
3. Разработка программной оболочки, реализующей приведенные выше действия;
4. Возможно проведение трех- и четырехфакторного анализа данных.