

УДК 681.518

А.И. Поворознюк, Н.А. Чикина, И.В. Антонова

Национальный технический университет «ХПИ», Харьков

## МОДЕЛЬ НЕЧЕТКОЙ ЭКСПЕРТНОЙ СИСТЕМЫ ПРОГНОЗА РИСКА РАЗВИТИЯ ПРОФЕССИОНАЛЬНО ОБУСЛОВЛЕННЫХ ЗАБОЛЕВАНИЙ

*Предложен метод идентификации состояния здоровья на основе выявленных внешних и внутренних факторов риска в условиях вредного производства. Разработана модель нечеткой экспертной системы по прогнозу уровня риска развития профессионально обусловленных аллергодерматозов и определены функциональные характеристики ее модулей. Проведено компьютерное моделирование в среде MATLAB.*

**Ключевые слова:** идентификация состояния, нечеткая экспертная система, прогноз риска развития, компьютерное моделирование.

### Введение

#### Постановка проблемы и анализ литературы.

За последние 20 – 30 лет существенно изменилась социальная и производственная среда, что обусловило увеличение числа рисков возникновения профессионально обусловленных заболеваний (ПОЗ). На предприятиях химико-фармацевтической промышленности к профессиональным рискам относится воздействие активных химических веществ, вызывающих различные заболевания аллергического генеза. Поэтому актуальной является разработка системы прогноза риска развития ПОЗ у рабочих и служащих предприятия с вредными условиями производства.

В ходе исследований, проводимых в рамках научно-исследовательских работ НИИ Дерматологии и Венерологии АМН Украины совместно с НТУ "ХПИ" на предприятиях химико-фармацевтической промышленности кроме профессиональных рисков авторами были выявлены групповые и внутренние факторы риска развития ПОЗ [1 – 3]. Наличие таких факторов риска дает возможность построить экспертную систему (ЭС) прогноза риска развития ПОЗ.

В соответствии с современными представлениями о факторах риска возникновения и развития различных заболеваний всё многообразие факторов, влияющих на здоровье, можно разделить на две основные группы: внутренние (биологические) и внешние (природные и социальные).

К первой группе относятся пол, возраст, наследственность, тип высшей нервной деятельности, конституциональные особенности. Вторая группа факторов, касающихся окружающей среды обитания и образа жизни, более обширна и включает природные (географические, метеорологические, климатические, экологические и др.) и социальные (производственно-трудовые, общественно-политические, социально-культурные, связанные с бытом и медицинской активностью) факторы.

Прогнозирование риска возникновения заболевания и разработка рекомендаций по его первичной профилактике имеет большое значение для службы здравоохранения, где приоритетность профилактической медицины должна в ближайшее время стать безусловной, так как она при значительно более высокой эффективности требует меньших экономических затрат, чем вторичная профилактика. Первичная профилактика является тем инструментом, который позволяет сохранить здоровье населения страны на основе таких направлений своей деятельности, как профилактика причин, факторов риска, формирование условий и факторов здорового образа жизни, здорового быта, развития личности.

В ходе исследований с помощью метода корреляционной адаптометрии авторами были выявлены закономерности развития адаптационных реакций в зависимости от некоторых факторов риска развития аллергодерматозов у рабочих химико-фармацевтических предприятий, а также стабильный диапазон с наибольшим по величине влияния на состояние здоровье рабочих кумулятивным эффектом [3].

Появление и совершенствование автоматизированных систем для массовых обследований населения нельзя рассматривать в отрыве от процессов, связанных с проникновением компьютера во врачебное мышление. Если в начале развития медицинской кибернетики речь шла о создании ЭС для дифференциальной диагностики заболеваний с похожей симптоматикой, то в последние годы ставится задача создания автоматизированных информационных систем – советчиков врача [4].

Общепринятая классификация ЭС отсутствует, однако наиболее часто ЭС различают по назначению, предметной области, методам представления знаний, динамичности и сложности [5].

Человек способен принимать практически полезные решения в условиях неполной и неопределенной (нечеткой) информации. Поэтому построение моделей, использующих рассуждения человека

и применение их в компьютерных системах представляет собой одну из важнейших научно-технических проблем. При количественном описании и построении моделей биологических объектов и систем решения конкретных прикладных задач целесообразно, а в ряде случаев – и необходимо, использовать указанную способность человеческого интеллекта с тем, чтобы адекватно учесть специфику биообъектов. Мощным инструментом совместного решения этих проблем является математический аппарат нечеткой логики, основы которого были предложены Л. Заде еще в 1965 г. [6]. Это подтолкнуло исследователей к созданию нечетких ЭС.

Логико-лингвистические методы описания нечетких систем основаны на том, что поведение исследуемой системы описывается в естественном (или близком к естественному) языке в терминах лингвистических переменных [7]. Входные и выходные параметры системы являются лингвистическими переменными, а качественное описание процесса задается совокупностью простейших правил следующего вида:

$$L_j: \text{если } \tilde{A}_j, \text{ то } \tilde{B}_j, (j = \overline{1, k_0}),$$

где  $\tilde{A}_j$  и  $\tilde{B}_j$  – нечеткие подмножества, заданные на декартовом произведении универсальных множеств входных и выходных лингвистических переменных соответственно. Кроме того, в основе нечетких ЭС лежит принцип вычисления суперпозиции многих влияний на окончательный результат (композиционное правило Заде), что позволяет учесть совместное влияния целого ряда факторов в решении задачи прогноза риска развития ПОЗ.

**Целью статьи** является описание структуры и принципов работы нечеткой ЭС прогноза уровня риска развития профессионально обусловленных аллергодерматозов у рабочих предприятий химико-фармацевтической промышленности.

### Основная часть

Рассматриваемая нечеткая ЭС проектируется на основе выявленных в ходе предыдущих исследований факторов риска [1 – 3]. В качестве экспертов при создании нечеткой базы знаний были привлечены руководители и ответственные исполнители НИР [8] НИИДиВ АМНУ и НТУ "ХПИ".

Пусть для описания множества  $\Omega$ , состоящего из  $m$  объектов, используется  $n$  признаков (выявленных факторов риска):  $\{X_1, X_2, \dots, X_n\}$ . Каждому объекту из множества  $\Omega$  соответствует некоторое значение целевого признака  $X_0$ . Признаки  $X_i$  ( $i = \overline{0, n}$ ) измерены в порядковой или номинальной шкале. Целевой признак  $X_0$  имеет  $k_0$  градаций, т. е. принимает  $k_0$  различных значений.

Для всех градаций признаков  $X_i$  ( $i = \overline{1, n}$ ) по-

строены функции принадлежности  $\mu^{jp}(X_i)$  входного признака  $X_i$  нечеткому терму  $X_i^{jp}$ .

На основе результатов предыдущих исследований был синтезирован набор  $k_0$  правил вида:

$$L_j: \text{ЕСЛИ } (X_1 = X_1^{j1}) \text{ И } (X_2 = X_2^{j1}) \text{ И } \dots \text{ И } (X_n = X_n^{j1}) \text{ (с весом } \omega_{j1})$$

$$\text{ИЛИ } (X_1 = X_1^{j2}) \text{ И } (X_2 = X_2^{j2}) \text{ И } \dots \text{ И } (X_n = X_n^{j2}) \text{ (с весом } \omega_{j2})$$

.....

$$\text{ИЛИ } (X_1 = X_1^{jqj}) \text{ И } (X_2 = X_2^{jqj}) \text{ И } \dots \text{ И } (X_n = X_n^{jqj}) \text{ (с весом } \omega_{jqj}), \text{ ТО } X_0 = X_0^j,$$

где  $L_j$  – решающее правило с номером  $j$  ( $j = \overline{1, k_0}$ );

$X_i^{jp}$  – градация входной переменной  $X_i$  в правиле с номером  $jp$  ( $j = \overline{1, k_0}, p = \overline{1, q_j}$ );  $q_j$  – количество правил, в которых выход  $X_0$  оценивается термом  $X_0^j$ .

На основе базы нечетких правил  $L_j$  ( $j = \overline{1, k_0}$ )

была получена модель Мамдани нечеткого логического вывода для прогноза риска развития ПОЗ.

Согласно [9] лингвистическая переменная описывается набором  $\{X, T(X), U, G, M\}$ , в котором  $X$  – название этой переменной;  $T(X)$  – терм-множество  $X$ , т.е. совокупность ее лингвистических значений;  $U$  – универсальное множество;  $G$  – синтаксическое правило, порождающее термы множества  $T(X)$ ;  $M$  – семантические правила, задающие функции принадлежности нечетких термов, порожденных синтаксическими правилами  $G$ . В процессе структурной идентификации модели для каждой переменной  $X_i$  ( $i = \overline{0, n}$ ) задавалось ее имя, терм-множества (градации)  $T_i$ , соответствующие им универсальные множества  $U_i$ .

Приведем в качестве примера некоторые терм-множества входных переменных (факторов риска) и их множества значений:

$$X_1 = \text{"Уровень адаптации"}: T_1 = \{\text{"низкий"}, \text{"средний"}, \text{"высокий"}\}, U_1 \in [0; 1];$$

$$X_2 = \text{"Стаж работы на предприятии"}: T_2 = \{\text{"небольшой"}, \text{"средний"}, \text{"большой"}\}, U_2 \in [0; 40];$$

$$X_3 = \text{"Возраст"}: T_3 = \{\text{"молодой"}, \text{"средний"}, \text{"старший"}\}, U_3 \in [20; 60];$$

$$X_4 = \text{"Наличие контакта с лекарственными препаратами по месту работы"}: T_4 = \{\text{"нет"}, \text{"отчасти"}, \text{"есть"}\}, U_4 \in [0; 1];$$

$X_5$ ="Обобщенный внутренний фактор риска":  
 $T_5$ ={"слабый", "повышенный", "высокий"},  
 $U_5 \in [0;1]$ .

Для целевого признака:  $X_0$ ="Уровень риска развития ПОЗ":  $T_0$ ={"есть риск", "повышенный риск", "высокий риск"},  $U_0 \in [0;1]$ .

Признак, по которому авторы определяли индивидуальный уровень риска развития ПОЗ, был назван обобщенным фактором риска и описан в [10]. В ходе дискриминантного анализа было получено информативное подмножество факторов риска развития профессионально обусловленных аллергодерматозов, необходимое для разделения всей совокупности на три состояния, соответствующие различным уровням риска развития профзаболевания, а именно  $S_1$  ("есть риск развития ПОЗ"),  $S_2$  ("повышенный риск развития ПОЗ") и  $S_3$  ("высокий риск развития ПОЗ") (рис. 1).

На основе этих данных строились функции принадлежности  $\mu^{jp}(X_i)$  для всех выявленных факторов риска. Графики функций принадлежности, соответствующие факторам риска  $X_2$ ="Стаж работы на предприятии" и  $X_3$ ="Возраст" представлен на рис. 2, 3.

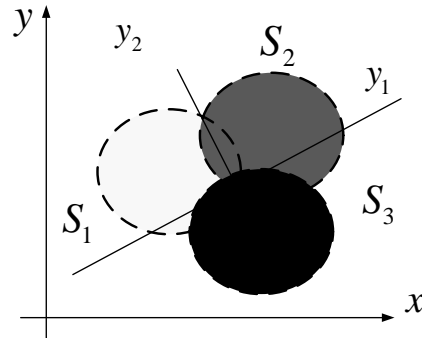


Рис. 1. Разделение обследованных по уровням риска на состояния с помощью дискриминантных функций

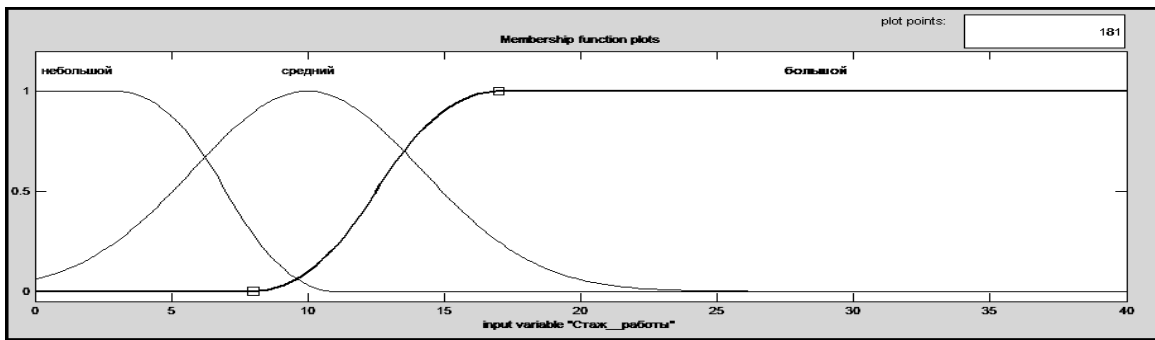


Рис. 2. Функция принадлежности фактора риска  $X_2$  = "Стаж работы на предприятии"

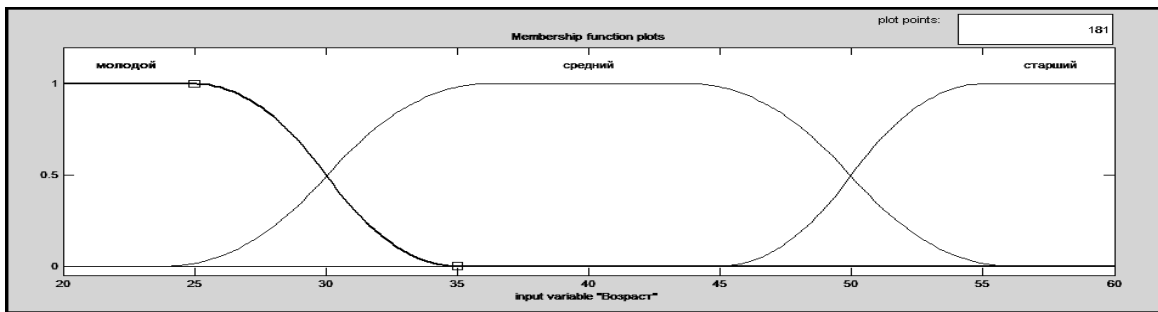


Рис. 3. Функция принадлежности фактора риска  $X_3$  = "Возраст"

Правило нечеткого вывода, соответствующее, например, повышенному риску развития ПОЗ имеет вид:

ЕСЛИ ( $X_1$  = "низкий") (с весом 0,6) ИЛИ ( $X_2$  = "средний") (с весом 1)

ИЛИ ( $X_3$  = "средний") (с весом 0,7) ИЛИ ( $X_4$  = "есть") (с весом 0,8)

ИЛИ ( $X_5$  = "высокий") (с весом 0,6), ТО  $X_0$  = "высокий".

Согласно механизму Мамдани (Mamdani) нечеткий вывод в ЭС осуществляется следующим образом:

1. На этапе фаззификации определяются степени истинности, т.е. значения функций принадлежности  $\mu^{jp}(X_i)$  ( $i = \overline{1, n}, j = \overline{1, k_0}, p = \overline{1, q_j}$ ) для левой части правила  $L_j$  (состоящей из  $j_p$  подправил).

2. Для текущих входных значений  $\vec{X}_0 = (X_1^*, X_2^*, \dots, X_n^*)$  минимаксным методом определяется значение функции принадлежности, соответствующее степени истинности целевого признака, т.е. число

$$\mu_0^j(\vec{X}^*) = \bigvee_{p=1, q_j} (\omega_{jp} \wedge \mu^{jp}(X_i^*)), \quad j = \overline{1, k_0},$$

где  $\vee(\wedge)$  – операция s-нормы (t-нормы).

3. На этапе дефаззификации определяется четкое значение целевого признака  $X_0$ , соответствующее входному вектору  $\bar{X}_0 = (X_1^*, X_2^*, \dots, X_n^*)$ . В настоящих исследованиях это значение определяется методом центра тяжести.

На основе построенных функций принадлежности, соответствующих факторам риска, и сформированных на основании выявленных закономерностей правил нечеткого логического вывода в пакете Fuzzy Logic среды MATLAB была реализована модель нечеткой экспертной системы прогноза риска развития ПОЗ.

На рис. 4 представлено окно нечеткого логического вывода разработанной модели в среде MATLAB.

Характер влияния факторов риска "Возраст" и "Стаж работы" на значение целевого признака  $X_0$  = "Уровень риска развития ПОЗ" отражен на графике поверхности значения целевого признака в зависимости от указанных факторов риска (рис. 5). На графике, в частности, видно, что наибольшее значение целевой признак достигает при стаже работы на предприятии химико-фармацевтической отрасли 10 – 12 лет в возрасте 38 – 42 года, что соответствует ранее выявленным закономерностям.

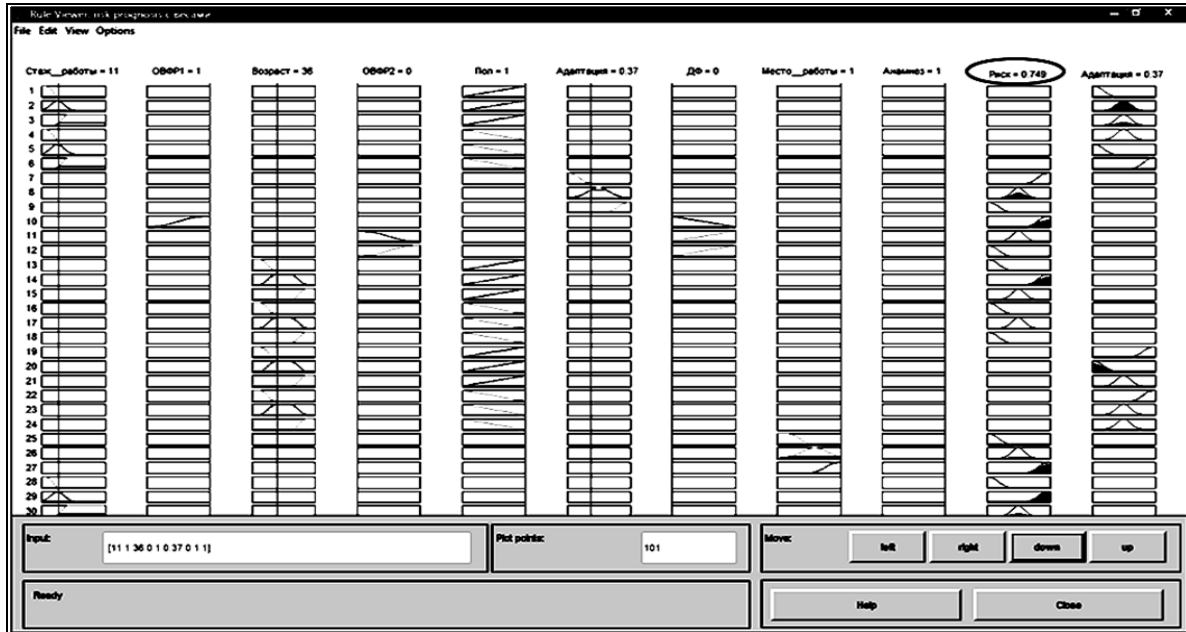


Рис. 4. Окно вывода нечеткой экспертной системы в среде MATLAB

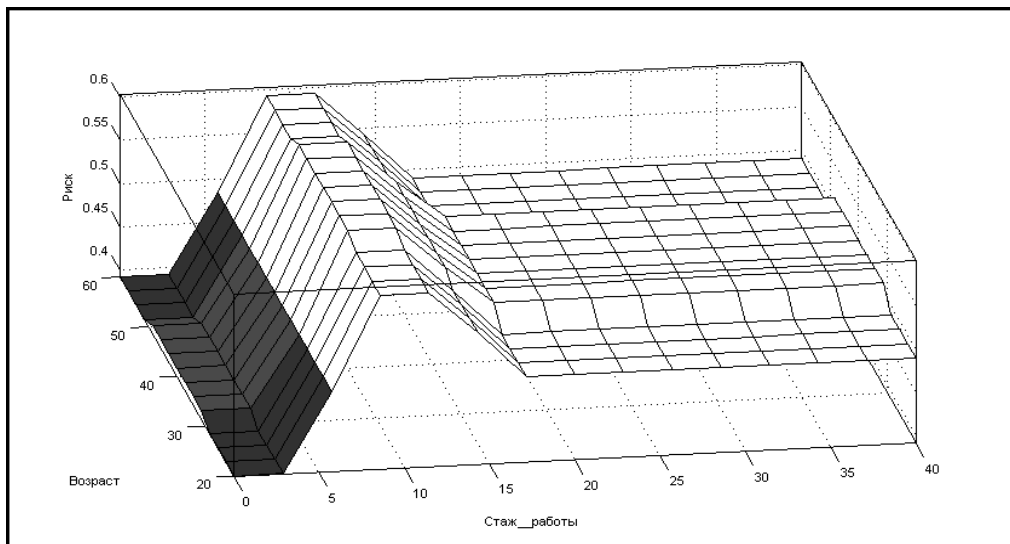


Рис. 5. Поверхность "Входы-выход" зависимости уровня риска развития ПОЗ от факторов "Возраст" и "Стаж работы"

Проверка адекватности модели нечеткой ЭС проводилась средствами пакета Fuzzy Logic среды MATLAB на случайно сформированной выборке из имеющейся базы, содержащей информацию о 430 об-

следованных рабочих профильных предприятий. В выборку вошли представители каждой из групп риска в одинаковых пропорциях (всего 150 обследованных). Результаты проверки представлены на рис. 6.

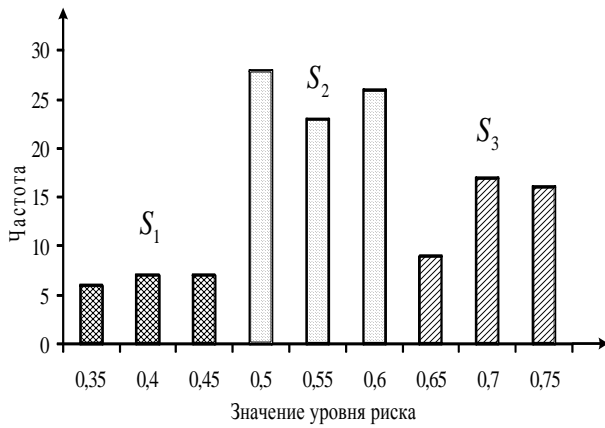


Рис. 6. Распределение объектов тестовой выборки по признаку "Риск развития ПОЗ"

Из рисунка видно, что проведенная апробация подтверждает наличие трёх нечетких состояний обследованных в зависимости от наличия (отсутствия) тех или иных факторов риска. Для повышения качества распознавания ЭС предполагается в дальнейшем провести ее оптимизацию.

**Выводы**

Разработана модель нечеткой ЭС по прогнозу уровня риска развития профессионально обусловленных аллергодерматозов и определены функциональные характеристики ее модулей. Проведенное компьютерное моделирование в среде MATLAB показало, что нечеткость в исходных данных не является препятствием для идентификации состояния обследуемых. Оптимизацию разработанной модели ЭС предполагается осуществить путем ее параметрической идентификации по обучающей выборке. Результаты работы данной ЭС могут быть использованы на предприятии химико-фармацевтической промышленности при проведении профилактических медицинских осмотров в качестве рекомендаций по мерам профилактики профзаболевания.

**Список литературы**

1. Чикина Н.А. Изучение влияния внутренних факторов риска на развитие аллергодерматозов у рабочих химико-фармацевтических предприятий / Н.А. Чикина, И.В. Антонова // *MicroCAD – 2007: тр. Международной науч.-техн. конф.* – Х.: НТУ "ХПИ", 2007. – Ч. 8. – С. 195-200.
2. Чикина Н.А. Идентификация состояния здоровья на основе анализа типов реакции адаптации у рабочих предприятий химико-фармацевтической промышленности / Н.А. Чикина, И.В. Антонова // *Вестник НТУ "ХПИ". Тематический выпуск: Информатика и моделирование.* – Х.: НТУ "ХПИ", 2008. – № 24. – С. 178-184.
3. Чикина Н.А. Математические модели адаптации к вредным условиям труда на основе метода корреляционной адаптометрии / Н.А. Чикина, И.В. Антонова // *Вестник НТУ "ХПИ". Тематический выпуск: Информатика и моделирование.* – Х.: НТУ "ХПИ", 2008. – № 49. – С. 184-189.
4. Баевский Р.М. Оценка адаптационных возможностей организма и риска развития заболеваний / Р.М. Баевский, А.П. Берсенева. – М.: Медицина, 1997. – 60 с.
5. Джарратано Д. Экспертные системы: принципы разработки и программирование / Джарратано Д., Г. Райли. – М.: Издательский дом "Вильямс", 2006. – 1152 с.
6. Zadeh L.A. Fuzzy sets / L.A. Zadeh // *Information and Control.* – 1965. – № 8 (3). – P. 338-353.
7. Штовба С.Д. Проектирование нечетких систем средствами MATLAB / С.Д. Штовба. – М.: Горячая линия-Телеком, 2007. – 288 с.
8. Отчет о научно-исследовательской работе "Разработка комплекса мероприятий с автоматизированной системой управления по ранней диагностике, лечению и профилактике больных аллергодерматозами среди рабочих с вредными условиями труда". – № 01.9.10 050715, 1993 г.
9. Заде Л.А. Понятие лингвистической переменной и его применение к принятию приближенных решений / Л.А. Заде. – М.: Мир, 1976. – 165 с.
10. Прогноз развития профессионально обусловленных заболеваний с помощью дискриминантного анализа / А.И. Поворознюк, Н.А. Чикина, Ю.Л. Геворкян, И.В. Антонова // *Системи обробки інформації: зб. наук. пр.* – Х.: ХУ ПС, 2010. – Вип. 1 (82). – С. 200-203.

Поступила в редколлегию 21.04.2010

Рецензент: д-р техн. наук, проф. Е.Л. Пиротти, Национальный технический университет "ХПИ", Харьков.

**МОДЕЛЬ НЕЧІТКОЇ ЕКСПЕРТНОЇ СИСТЕМИ ОЦІНКИ РИЗИКУ РОЗВИТКУ ПРОФЕСІЙНО ОБУМОВЛЕНИХ ЗАХВОРЮВАНЬ**

А.І. Поворознюк, Н.О. Чікіна, І.В. Антонова

Запропоновано метод ідентифікації стану здоров'я на основі виявлених зовнішніх і внутрішніх чинників ризику в умовах шкідливого виробництва. Розроблено модель нечіткої експертної системи прогнозу рівня ризику розвитку професійно обумовлених алергодерматозів і визначені функціональні характеристики її модулів. Проведено комп'ютерне моделювання в середовищі MATLAB.

**Ключові слова:** ідентифікація стану, нечітка експертна система, прогноз ризику розвитку, комп'ютерне моделювання.

**FUZZY EXPERT SYSTEM MODEL OF RISK PROGNOSIS OF DEVELOPMENT OF THE PROFESSIONALLY CONDITIONED DISEASES**

A.I. Povoroznuk, N.A. Chikina, I.V. Antonova

The method of identification of the health state is offered on the basis of the indicated external and internal risk factors in the conditions of harmful production. The model of fuzzy expert system of risk prognosis of development of the professionally conditioned allergodermatoses is worked out and functional descriptions of its modules are determined. A computer modeling is conducted in the MATLAB environment.

**Keywords:** state identification, fuzzy expert system, prognosis of risk development, computer modeling.