

Summary. Thus a laboratory, being a subject of the market relations, should solve a problem of quality management, as a significant proof element, maintenance of the procedure for conformity assessment.

According to the requirements of the standard ISO 9001:2000 Scientific and Research Laboratory for Drugs Quality Control, State Institution "a.m. Marzeyev Institute for Hygiene and Medical Ecology, Academy of Medical Sciences of Ukraine", has introduced the Quality Management Systems in its activity, realising a process approach.

ВИКОРИСТАННЯ МАТЕМАТИЧНОГО АПАРАТУ РОЗПІЗНАВАННЯ ОБРАЗІВ ДЛЯ ВИРІШЕННЯ ЗАДАЧ КІЛЬКІСНОЇ ОЦІНКИ КЛІНІКО-ГЕНЕТИЧНОГО СТАТУСУ ДІТЕЙ, ЯКІ НАРОДИЛИСЯ ВІД БАТЬКІВ-УЧАСНИКІВ ЛІКВІДАЦІЇ НАСЛІДКІВ АВАРІЇ НА ЧАЕС

Скварська О.О.

ДУ «Науковий Центр радіаційної медицини АМН України», м. Київ

Для оцінки ймовірних негативних наслідків опромінення батька для першого покоління нащадків використано сімейний підхід. Використання математичного апарату розпізнавання образів дозволило виділити комплекс фенотипічних ознак, які вирізняють дітей, що народились у опроміненіх батьків, від їх сибсів, народжених до аварії. Отримані дані можуть бути використані для створення моніторингових систем для оцінки екологічного впливу на базі комплексного обліку малих аномалій розвитку і клінічних ознак.

Серед медичних наслідків Чорнобильської аварії важливе місце посідає оцінка ймовірних генетичних ефектів в першому поколінні нащадків опроміненіх батьків. Проте, отримані результати мають суперечливий характер [1,2], що потребує подальших наукових досліджень та уточнень. Нами запропоновано методику кількісної оцінки клініко-генетичного статусу дітей з використанням математичного апарату розпізнавання образів.

Мета. Оцінити ефективність математичного апарату розпізнавання образів на базі завданого набору фенотипічних ознак з метою вирішення задач кількісної оцінки клініко-генетичного статусу дітей, які народилися від учасників ліквідації наслідків аварії на ЧАЕС.

Для оцінки ймовірних негативних наслідків опромінення батька для I покоління нащадків використано сімейний підхід. Обстежено 297 сімей учасників ліквідації наслідків аварії (ЛНА) на ЧАЕС та 97 сімей групи порівняння, до складу яких увійшли: мати, батько, дві дитини, одна з яких була народжена до (сібс), друга – після аварії на ЧАЕС (пробанд). Головним критерієм добору родин в основну групу був факт впливу іонізуючого випромінювання на батька в результаті його участі в ліквідації наслідків аварії на Чорнобильській АЕС у 1986-1987 рр. До групи порівняння відбиралися родини, в яких жоден з батьків не піддавався впливу радіаційного чинника.

Комплексне клінічне обстеження дітей та їх батьків проводилося на базі педіатричного та терапевтичного відділень поліклініки клініки НЦРМ АМН України. Усі члени родин були оглянуті фахівцями наступного профілю: генетиком, педіатром (терапевтом), ендокринологом, невропатологом, психіатром, гематологом, кардіологом, стоматологом, оториноларингологом, окулістом.

Інструментальне обстеження включало проведення УЗД органів черевної порожнини, щитоподібної залози, зачеревного простору і, за потреби – органів малого таза, ЕЕГ, ЕКГ, аудіометрії. На підставі отриманих даних оцінювався стан здоров'я батьків

та їх дітей. Проводився розподіл дітей за групами здоров'я.

Для фіксації малих аномалій розвитку (МАР) членів обстежуваних родин був розроблений уніфікований протокол клініко-генетичного обстеження, в якому усі МАР було згруповано за принципом анатомічної локалізації [3]. Також він включав дані генеалогічного анамнезу з відомостями про родичів про банду, стан їх здоров'я, наявності вродженої і спадкової патології в родині, дані про можливі контакти з потенційними мутагенами навколишнього середовища, дозу загального опромінення батька, дані акушерського анамнезу, відомості про перебіг вагітності та пологів, неонатального періоду і розвитку дитини в наступні роки життя, результати інструментальних і лабораторних методів дослідження.

Нами була зроблена спроба визначити чи формуються які-небудь фенотипічні особливості у пробандів, що відрізняють їх від сібсів, і якщо вони формуються, то яка сукупність цих ознак. Одним з методів, який дозволяє враховувати безліч ознак у їх взаємозв'язку є метод групового обліку аргументів

(МГОА). Його суть полягає в класифікації двох і більше генеральних сукупностей і виробленню вирішального правила, що дозволяє віднести новий елемент за комплексом його ознак до одного з класів даної безлічі [4].

Для вирішення задачі був використаний програмний комплекс, до складу якого входили наступні програми: ідентифікації характеристик складних систем ("IDENTA"), моделювання й оптимізації складних багатокритеріальних систем ("Grand-96"), розпізнавання класів об'єктів на основі ймовірного алгоритму методу групового обліку аргументів (МГОА) VERSION, адаптованих до обробки клініко-генетичних характеристик, а також програмного пакета для обробки статистичної інформації "Superplus-2000" [4,5].

Для вирішення задачі розпізнавання вихідна інформація була трансформована до вигляду, зручного для введення в програму розпізнавання. Сутність трансформації полягала в наступному: для кожної із стигм у відношенні кожного з дітей (А – народженого до аварії на ЧАЕС і В – народженого після участі батька в ліквідації наслідків аварії на ЧАЕС) вирішувалися наступні рівняння:

$$A_i = \text{if}(X_i \neq V_i, 1, \text{if}(X_i \neq U_i, 1, 0)), \quad (1)$$

$$B_i = \text{if}(Z_i \neq V_i, 1, \text{if}(Z_i \neq U_i, 1, 0)), \quad (2)$$

де: X_i – наявність (відсутність) і стигми у дитини, що народилася до аварії;

Z_i – наявність (відсутність) і стигми у дитини, що народились після аварії;

V_i і U_i – наявність чи відсутність даної стигми в батька та матері, відповідно.

У програмному продукті вигляд дитини може бути описаний як кількісними, так і якісними ознаками. Для вирішення задачі у програмному продукті реалізовано заснований на принципах самоорганізації ймовірносний алгоритм МГОА. Оцінка якості синтезованого вирішального правила проводилася за допомогою відомих положень теорії ймовірностей, заснованих на теоремі Бернуллі.

Реалізований алгоритм, заснований на принципі самонавчання, був призначений

для генерації на етапі навчання алгоритмічного вирішального правила (індикаторної функції), яке дозволяло переводити безліч багатомірних крапок простору ознак X , що описують образ (ситуацію) у безліч класів R , де $R = \{1, 2\}$.

При вирішенні задачі розпізнавання для визначення апостеріорної ймовірності належності образу до того чи іншого класу, найбільш оптимальним при роботі з генеральною сукупністю є відома формула Байеса:

$$P(R_i/X) = \frac{P(R_i/x_1 x_2 \dots x_n)}{\sum_{i=1}^2 P(R_i/x_1 x_2 \dots x_n)} = \frac{P(R_i)P(x_1/R_i)P(x_2/x_1 R_i) \dots P(x_n/x_1 x_2 \dots x_{n-1} R_i)}{\sum_{i=1}^2 P(R_i)P(x_1/R_i) \dots P(x_n/x_1 x_2 \dots x_{n-1} R_i)} \quad (3)$$

де: $P(R_i/X)$ – апостеріорна ймовірність належності образу X до класу R_i ;

$P(R_i)$ – апіорна ймовірність існування класу;

$P(x_j/x_{j-1} \dots x_1 R_i)$ – умовна ймовірність присутності ознаки x_j в класі R_i за умови наявності ознак $x_1 \dots x_{j-1}$;

повна ймовірність наявності ознак $x_1 \dots x_n$ во всіх класах;

$$\sum_{i=1}^2 P(R_i)P(x_1/R_i) \dots P(x_n/x_1 x_2 \dots x_{n-1} R_i)$$

n – кількість ознак.

Проте при вирішенні реальних задач розпізнавання образів на обсяг даних накладається ряд обмежень, які описують класи. З цієї причини застосування формули Байєса в класичному вигляді стає практично неможливим. Виникає необхідність застосування методів, що дозволили б вирішити поставлене завдання в умовах обмеженої інформації. До таких методів належить заснований на принципах самоорганізації і математичної селекції метод групового обліку аргументів. Ймовірнісні алгоритми МГОА працюють у припущенні про залежність вхідних ознак. Крім того, замість ймовірностей, відповідно до теореми Бернуллі, використовуються відповідні частки, розраховані для наявних даних і враховується тільки чисельник формули Байєса.

При залежних ознаках і малому обсязі експериментальних даних, коли безпосереднє застосування формули Байєса неможливе, для відновлення багатомірної щільності розподілу «повний розподіл» $P(R_i/X)$ замінюється набором "приватних описів" двовимірних щільностей розподілу у вигляді:

$$P(R_i/x_k x_l) = P(x_k/R_i)P(x_l/x_k R_i), \quad (4)$$

де: k, l – менше чи дорівнює кількості крапок у вибірці;

$P(x_k/R_i)$ – умовна ймовірність того, що ознака x_k присутня в образі R_i ;

$P(x_l/x_k R_i)$ – умовна ймовірність того, що ознака x_l присутня в образі R_i за умови наявності x_k .

Отримана формула (4) є чисельником формули Байєса. Для одержання стійких рішень при визначенні апостеріорної ймовірності приналежності образу X к класу R_i вибірка розподіляється на дві частини – навчальну і перевірочну. Перша її частина слугує для побудови вирішального правила у вигляді приватного опису, друга – для оцінки його якості.

Після цього відбувається добір безлічі M найбільш точних «приватних описів», що здійснюється по одному чи декількох критеріях. Потім формується масив даних для наступного ряду селекції – значення апостеріорних ймовірностей, отриманих по M відібраним «приватним описам» у всіх крапках навчальної послідовності. Ця процедура завершує поточний ряд селекції вирішального правила за методом групового обліку аргументів. На наступних рядах селекції процес навчання і добору повторюється.

Алгоритм класифікації МГОА являє собою багаторядну структуру, у який реалізується підхід «навчання з вчителем». У

зв'язку з цим, у вихідних даних, що надходять на вхід алгоритму класифікації, можна виділити дві групи. Перша визначає значення керованих параметрів алгоритму і дозволяє скерувати на рішення конкретної задачі, друга група – послідовності представників багатомірних образів визначеного класу, до якого відноситься конкретний образ (послідовність багатомірних крапок простору ознак, що описують об'єкти, які класифікуються, заданих двох класів). Другу групу вихідних даних для стислості називають вихідною послідовністю. Алгоритм класифікації складається з трьох послідовних етапів – підготовчого, селекції вирішальних правил, відновлення і добору вирішальних правил.

На підготовчому етапі вихідні дані перетворюються відповідно до ідеології МГОА: вихідні послідовності класів поділяються на дві частини – навчальну і перевіряючу. Перед тим, як зробити такий розподіл, вихідні послідовності класів упорядковуються за зменшенням дисперсії. В алгоритмі реалізовано спосіб розподілу вихідної

послідовності на навчальну і перевіряючу "через крапку": непарні представники багатомірних образів об'єктів (сібсів чи пробандів) у кожному класі (крапки багатомірного простору ознак) відносяться до навчальної послідовності, парні – до перевіряючої. Підготовлені таким чином навчальні і перевірочні послідовності двох класів дискретизуються на завдану користувачем кількість рівнів дискретизації.

На етапі селекції вирішальних правил відбувається багаторазове послідовне виконання двох операцій: генерація множини елементарних вирішальних правил, добір заданої кількості кращих з них.

Спільне виконання цих двох операцій визначає поняття ряду селекції. При цьому кожне таке правило формується на основі фіксованої пари ознак з обліком тільки тієї інформації, що міститься в навчальних послідовностях. Кількість елементарних вирішальних правил відповідає числу сполучень пар ознак з наданого користувачем списку.

Оцінка вирішальних правил здійснюється шляхом обчислення точності рішення задачі класифікації за даними перевіряючих послідовностей, наданих на вхід алгоритму. У такий спосіб реалізується принцип зовнішнього доповнення. Пари ознак ранжуються за значенням обраного критерію селекції. З ранжованого ряду пар ознак відбирається безліч M кращих для обраного критерію селекції. Число M визначає волю вибору рішень. Відібрані пари виступають у наступному ряді селекції як аргументи, а рішення, отримані по кожній парі аргументів, приймаються як їх значення. Таким чином, при збільшенні кількості рядів селекції підвищується складність вирішальних правил.

Аргумент поточного ряду селекції являє собою пари аргументів попереднього ряду, кожний з яких, у свою чергу, також складається з пари аргументів попереднього ряду і т.п. до першого ряду селекції, де використовуються первинні аргументи (параметри), за допомогою яких описуються представники класів навчальних і перевіряючих вибірок. Розрахунок припиняється по закінченні виконання завданної кількості рядів селекції.

На етапі відновлення і добору вирішальних правил формуються «дерева рі-

шень» для всіх рядів селекції. Для кожного вирішального правила з множини вирішальних правил, утворених при відновленні на кожному ряді селекції, визначаються наступні оцінки: точності прийняття рішень для кожного класу, середньої точності по класах, дисперсії точності.

Оцінки було проведено з використанням вихідних послідовностей. За даними оцінок вибиралося одне вирішальне правило, у якого була максимальна середня точність по класах, мінімальна дисперсія точності та в яке входила максимальна кількість класоутворюючих ознак.

Як показав кореляційний аналіз вихідної вибірки, між окремими ознаками існують вірогідні статистичні залежності. Це дало можливість редукувати простір ознак, залишаючи для розгляду лише найбільш інформативні з них без істотних втрат інформативності вибірки в цілому.

Рішення задачі класифікації здійснювалося за допомогою програмного продукту VERSION, призначеного для класифікації складних систем за допомогою синтезованого за методом групового обліку аргументів вирішального правила. Раціональна складність моделі була досягнута на 5-му ряду селекції (подальше її ускладнення не призвело до збільшення ймовірності правильного розпізнавання класів).

Результати математичного аналізу показали, що найбільш характерним морфогенетичним варіантом пробандів була наявність множинних стигм дизембріогенезу, причому в 54,2% випадків до них входили не менше 5 MAP з наступного комплексу ознак: підвищена розтяжність шкіри, телеангіектазії, підвищена рухливість суглобів, кіфосколиотична постава, плоскоступість, дисплазії сполучнотканинних структур серця, шлунково-кишкового тракту та аномалії розвитку чашково-мискового комплексу нирок, тобто ознак наявності у цих дітей СНДСТ (синдрому недиференційованої дисплазії сполучної тканини). Ймовірність правильного розпізнавання склала 0,80303..

Таким чином, вирішення задачі розпізнавання дозволило виділити комплекс фенотипічних ознак, які вирізняють дітей, що народились у опромінених батьків, від їх сібсів, народжених до аварії.

Отримані дані можуть бути використані для створення моніторингових систем для оцінки екологічного впливу на базі комплексного обліку малих аномалій розвитку і клінічних ознак. З методичної та економічної точки зору створення такої системи є ефективним, доступним, а результати можна отримати за короткий час.

ЛІТЕРАТУРА

1. Фенотипические и цитогенетические особенности детей, родившихся в семьях участников ликвидации Чернобыльской аварии /Е.И. Степанова, Е.А. Скварская, В.Ю. Вдовенко //Генетические последствия чрезвычайных радиационных ситуаций: Тез.докл. III междунард. конф., Дубна, 4-7октября 2005 г. - М., - 2005. - С. 108-109.
2. Bard D.et al.Chernobyl, 10 Years after //Epid.Rev. - 1997. - Vol.19. - N2. - P.187-197.
3. Уніфікований протокол оцінки фенотипу членів родин учасників ліквідації наслідків аварії на ЧАЕС /Є.І. Степанова, О.О. Скварська, Ж.А. Мішаріна, О.С. Леонович, М.В. Гуртавцева //Інф. лист. - К.: "Укрмедпатентінформ", - 2005.
4. Лефтер А.В., Бакуменко А.Ф. Программный продукт «Version». //Руководство пользователя. - Киев. - Тридента, - 1992. - 24 с.
5. Кокуев В.И., Герасименко О.В. Программный продукт «Optimus», версия 3.27 //Анализ, прогноз, оптимизация сложный многокритериальных систем. - Киев, - 1994. - 41 с.

**ИСПОЛЬЗОВАНИЕ МАТЕМАТИЧЕСКОГО АППАРАТА РАСПОЗНАВАНИЯ ОБРАЗОВ
ДЛЯ РЕШЕНИЯ ЗАДАЧ ЧИСЛОВОЙ ОЦЕНКИ КЛИНИКО-ГЕНЕТИЧЕСКОГО
СТАТУСА ДЕТЕЙ, РОЖДЕННЫХ У УЧАСТНИКОВ ЛИКВИДАЦИИ
ПОСЛЕДСТВИЙ АВАРИИ НА ЧАЭС**

Скварская Е.А.

Для оценки вероятных негативных последствий облучения отца для первого поколения детей использован семейный подход. Использование математического аппарата распознавания образов позволило выделить комплекс фенотипических признаков, отличающих детей, рожденных после аварии, от их старших братьев и сестер, рожденных в доаварийный период. Полученные данные могут быть использованы для создания мониторинговых систем для оценки экологического воздействия на базе комплексного учета малых аномалий развития и клинических признаков.

**THE USE OF MATHEMATICAL TOOLS FOR PATTERN RECOGNITION FOR SOLVING
THE ISSUES OF NUMERICAL EVALUATION OF CLINICAL AND GENETIC STATUS
OF CHILDREN BORN OF LIQUIDATORS OF THE CHERNOBYL ACCIDENT**

E.A. Skvarkaya

To assess the potential adverse effects of irradiation of a father for the first generation of children a family approach was used. Using mathematical tools for pattern recognition allowed to mark out the complex of phenotypic traits that distinguish children born after the accident, from their older brothers and sisters born in the pre-accident period. The data obtained can be used to create monitoring systems for environmental impact assessment on the basis of a comprehensive accounting of small maldevelopments and clinical features.

Куратор розділу – д. мед. наук Нікітіна Н.Г.