

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ РАДІОЕЛЕКТРОНІКИ

Максюта Наталія Валеріївна

УДК 61:004.9:519.254

**СИСТЕМА АНАЛІЗУ РЕЗУЛЬТАТІВ ГЕМОРЕОЛОГІЧНИХ ДОСЛІДЖЕНЬ НА ОСНОВІ  
ІЄРАРХІЧНОЇ КЛАСТЕРИЗАЦІЇ ДІАГНОСТИЧНИХ ОЗНАК**

Спеціальність 05.11.17 – біологічні та медичні прилади і системи

Автореферат  
дисертації на здобуття наукового ступеня  
кандидата технічних наук

Харків – 2007



Дисертацією є рукопис

Робота виконана у Національному технічному університеті „Харківський політехнічний інститут” Міністерства освіти і науки України

Науковий керівник – кандидат технічних наук, доцент

**Поворознюк Анатолій Іванович,**

Національний технічний університет „Харківський політехнічний інститут”  
Міністерства освіти і науки України, професор кафедри обчислювальної  
техніки та програмування

Офіційні опоненти: доктор технічних наук, професор

**Злепко Сергій Макарович,**

Вінницький національний технічний університет Міністерства освіти і  
науки України, завідувач кафедри проектування медико-біологічної  
апаратури

кандидат технічних наук

**Дацок Олег Михайлович,**

Харківський національний університет радіоелектроніки Міністерства  
освіти і науки України, доцент кафедри біомедичних електронних пристроїв  
та систем

Провідна установа:

Національний технічний університет України “Київський політехнічний інститут” Міністерства  
освіти і науки України, кафедра фізичної та біомедичної електроніки, м. Київ

Захист відбудеться „ 5 ” червня 2007 р. о 13.00 годині на засіданні спеціалізованої вченої ради  
К64.052.05 у Харківському національному університеті радіоелектроніки за адресою: 61166, м.  
Харків, пр. Ленина, 14.

З дисертацією можна ознайомитись у науковій бібліотеці Харківського національного  
університету радіоелектроніки (61166, м. Харків, пр. Ленина, 14).

Автореферат розісланий „ 13 ” квітня 2007 р.

Вчений секретар

спеціалізованої вченої ради

\_\_\_\_\_

М.П. Мустецов

## ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

**Актуальність теми.** Підвищення імовірності діагностики захворювань є актуальною проблемою при проектуванні інтелектуальних комп'ютерних систем медичної діагностики (ІКСМД). При цьому якість діагнозу, який синтезує ІКСМД, в значній мірі визначається характеристиками вхідних даних - показниками організму людини, які знаходяться у складній взаємодії один з одним. Відомо, що в лікувальній практиці при визначенні діагнозу присутній ефект істотної багатомірності та різнотипності вихідних показників, а основна задача діагностики полягає в розпізнаванні патологічних змін і реакцій організму, які складають сутність захворювання. При цьому розглядаються не тільки причина та наслідок змін в організмі, але і їх складна взаємодія. У зв'язку з цим при проектуванні ІКСМД виникає необхідність у ієрархічній кластеризації діагностичних ознак (ДО), під якою розуміється формування груп тісно зв'язаних між собою показників і побудова ієрархічної структури ДО з метою відбору діагностично цінних з них відносно заданої системи діагнозів. У свою чергу дослідження реологічних властивостей крові (гемореології) на сьогоднішній день представляють собою особливий інтерес і є актуальними, тому що порушення гемореології виступають у якості первинних прогностичних ознак захворювань кровоносної та серцево-судинної систем і своєчасна їхня корекція дозволяє запобігати розвитку патології. Однак, кровоносна система є зв'язуючим елементом всіх систем організму людини, тому кров реагує на будь-які зміни в організмі, а її показники сильно взаємозалежні. У зв'язку з цим при ранній діагностиці захворювань крові та серцево-судинної системи крім вимірювань гемореологічних показників (ГП) необхідно проводити дослідження структури взаємозв'язків між ними та відбір найцінніших з них. У цьому випадку для лікаря-дослідника ІКСМД служать як потужний помічник.

**Зв'язок роботи з науковими програмами, планами, темами.** Робота виконана відповідно до плану науково-дослідних робіт кафедри „Обчислювальна техніка та програмування” Національного технічного університету „Харківський політехнічний інститут” за темою № М6016 „Розробка теорії та методів проектування інтелектуальних комп'ютерних систем медичної діагностики на основі структурної ідентифікації об'єктів діагностики та перетворення простору діагностичних ознак” (№ держреєстрації 0104U003361).

**Мета та задачі дослідження.** Метою роботи є розробка системи аналізу результатів гемореологічних досліджень (ГД) на основі ієрархічної кластеризації діагностичних ознак з використанням ІКСМД з позиції теорії потокових моделей. Для досягнення даної мети необхідно вирішити наступні задачі:

- провести аналіз існуючих методів, які використовуються на різних етапах автоматизованої комп'ютерної обробки ГП, з метою обґрунтування можливості використання потокових моделей для формування мінімально необхідного набору ДО;

- провести аналіз існуючих приладів і методів дослідження ГП з метою розробки узагальненої структурної схеми системи аналізу результатів ГД з урахуванням трудомісткості їхніх вимірювань;

- розробити метод ієрархічної кластеризації ДО на основі потокових моделей з метою формування мінімально необхідного набору діагностичних гемореологічних ознак (ГО) для забезпечення якісного діагнозу;

- розробити методику побудови ієрархічної структури ДО мінімально необхідного обсягу з урахуванням взаємозв'язків між ними;

- розробити принципи побудови і програмне забезпечення (ПЗ) підсистеми формування мінімально необхідного набору діагностичних ГО на основі методу ієрархічної кластеризації ДО;

- провести адаптацію одного з перспективних алгоритмів розв'язання потокових задач з обмеженням - алгоритму „дефекту” до задачі ієрархічної кластеризації ДО;

- експериментально обґрунтувати та виконати перевірку адекватності застосування методу ієрархічної кластеризації ДО на основі потокових моделей для синтезу діагнозу в ІКСМД.

*Об'єктом дослідження є процес діагностування захворювань на основі гемореологічних показників.*

*Предметом дослідження виступає ієрархічна кластеризація діагностичних ознак на основі потокових моделей в інтелектуальних комп'ютерних системах медичної діагностики.*

*Методи дослідження.* На підставі системного підходу на основі потокових моделей сформульована задача ієрархічної кластеризації ДО. Відбір найцінніших ГО відносно заданої системи захворювань здійснено за допомогою теоретико-інформаційного підходу. За допомогою кореляційного аналізу і розробленого методу представлена структура зв'язків між ГП. Перевірку адекватності застосування методу ієрархічної кластеризації ДО на основі потокових моделей для синтезу діагнозу в ІКСМД виконано за допомогою дискримінантного та кластерного аналізу.

***Наукова новизна отриманих результатів.*** Основні наукові результати виконаної роботи полягають в тому, що:

- запропоновано метод ієрархічної кластеризації діагностичних ознак на основі потокових моделей, який дозволяє отримати оптимальне рішення та звести задачу кластеризації до пошуку мінімального розрізу на кожному її етапі;

- отримав подальший розвиток метод рішення потокових задач – алгоритм „дефекту” для ієрархічної кластеризації діагностичних ознак на основі потокових моделей, що дозволило зняти обмеження на розмірність задачі кластеризації;

- адаптовано параметри потокової моделі до задачі ієрархічної кластеризації діагностичних ознак, що дозволило сформувати ієрархічну структуру зв'язків між показниками;

- розроблено методику відбору діагностично цінних гемореологічних ознак на основі ієрархічної кластеризації, яка дозволяє враховувати їх статистичну залежність, діагностичну цінність відносно заданої системи діагнозів та особливості їх вимірювань з метою забезпечення якісного діагностування захворювань.

***Практичне значення отриманих результатів*** полягає в наступному:

- розроблено структуру системи аналізу результатів ГД на основі ієрархічної кластеризації ДО, що дозволяє виконувати ГД з урахуванням наявних приладів вимірювань;

- розроблено ПЗ реалізації підсистеми формування мінімально необхідного набору ДО на основі ієрархічної кластеризації;

– отримані дані ГД пацієнтів з різними кардіологічними патологіями, які підтверджують діагностичну чутливість запропонованого методу ієрархічної кластеризації ДО на основі потокових моделей, що свідчить про перспективність його використання в практичній медицині;

- за результатами ГД побудована ієрархічна структура ГО мінімально-необхідного обсягу і на нижньому її рівні виконаний відбір найцінніших ДО відносно заданої системи діагнозів.

*Результати роботи впроваджені та використовуються:*

– у Державному підприємстві (ДП) „Харківський НДІ гігієни праці та профзахворювань” для обробки ГП (акт впровадження);

– в Інституті дерматології і венерології Академії медичних наук України при визначенні комплексу діагностичних показників здоров'я у хворих поширеними дерматозами (акт впровадження);

– на кафедрі „Обчислювальна техніка та програмування” Національного технічного університету „Харківський політехнічний інститут” у науково-дослідній роботі та навчальному процесі (акт впровадження).

**Особистий внесок здобувача.** Всі основні результати, які складають сутність дисертаційної роботи і знайшли відображення в пунктах наукової новизни та практичного значення отримані автором особисто. У роботах [1–8], виконаних у співавторстві автором належать основні ідеї, пов'язані з вибором методів досліджень, а також способів рішення поставленої задачі: у роботах [1, 6] автором запропоновано метод ієрархічної кластеризації ДО на основі потокових моделей та виконано адаптацію алгоритму „дефекту” до ієрархічної кластеризації; у роботі [2] – запропоновано та виконано побудову ієрархічної структури діагностичних ГО; у роботі [3] – проведено порівняльний аналіз існуючих методів зниження розмірності простору вихідних показників та розглянуті їх недоліки; у роботі [4] – запропоновано методику відбору діагностично цінних ГО та виконано перевірку адекватності використання розробленого методу ієрархічної кластеризації ДО на основі потокових моделей для синтезу діагнозу; у роботі [5] – розроблено структурну схему системи аналізу результатів ГД на основі їх ієрархічної кластеризації з метою забезпечення якісної діагностики захворювань з урахуванням діагностичної цінності і трудомісткості вимірювань ГП; у роботі [7] – запропоновано процедуру синтезу діагнозу з використанням ієрархічної кластеризації ДО на основі потокових моделей; у роботі [8] – показано призначення розробленого методу ієрархічної кластеризації ДО в ІКСМД.

**Апробація результатів дисертації.** Основні положення дисертації були представлені на наступних конференціях: Міжнародні науково-практичні конференції „Інформаційні технології: наука, техніка, технологія, освіта, здоров'я” (Харків, НТУ „ХПІ”, 16-17 травня 2002 р., 15-16 травня 2003 р., 20-21 травня 2004 р., 19-20 травня 2005 р., 18-19 травня 2006 р.); Міжнародні науково-технічні конференції „Проблеми інформатики і моделювання” (Харків, НТУ „ХПІ”,

25-27 листопада 2004 р., 24-26 листопада 2005 р., 23-25 листопада 2006 р.); конференція „Комп’ютерна Медицина-2004” (Харків, ХМАПО, травень 2004 р.).

**Публікації.** Основні положення і результати дисертації опубліковано у 9 роботах: 5 робіт у фахових видання ВАН, 1 стаття в українському науковому журналі та 3 тези доповідей в українських наукових виданнях.

**Структура та обсяг роботи.** Дисертація складається із вступу, п’яти розділів, висновків, списку використаних джерел і додатків. Повний обсяг дисертації становить 247 сторінок, серед яких 2 ілюстрації на 2 сторінках, 3 таблиці на 3 сторінках, 12 таблиць і 25 ілюстрації по тексту, 6 додатків на 62 сторінках, список використаних джерел з 152 найменувань на 13 сторінках.

## ОСНОВНИЙ ЗМІСТ

У **вступі** обґрунтовано актуальність теми дисертації, її наукова та практична цінність, сформульовані мета та задачі роботи, наведена її загальна характеристика.

У **першому розділі** розглянуті принципи постановки діагнозу та методи обстеження пацієнтів, які використовуються в сучасній медицині. Показано, що постановка діагнозу і відбір діагностично цінних відносно нього ознак для лікаря є складним багаторівневим процесом, що включає аналіз великого обсягу різнорідних показників, які перебувають у складній взаємодії один з одним і характеризують функціонування організму людини. Відзначено, що ця проблема особливо актуальна при ГД, тому що по взаємній зміні ГП здійснюється рання діагностика захворювань крові та серцево-судинної системи. У зв’язку з цим обґрунтовано застосування ІКСМД, які дозволяють детально враховувати всілякі комбінації вихідних показників, з метою полегшення лікарського процесу ГД. Для забезпечення даної мети формалізовані етапи обробки медико-біологічної інформації в ІКСМД: збір медичної інформації; формалізація опису ДО; первинна обробка вихідних даних; зниження розмірності простору ДО; відбір діагностично цінних ознак відносно заданої системи діагнозів; синтез діагностичного вирішального правила. Наведено перелік задач, які вирішуються на кожному етапі, та математико-статистичних методів, які використовуються при цьому. Показано, що етапи зниження розмірності простору ДО та відбору найцінніших з них відносно заданої системи діагнозів особливо важливі з погляду підвищення здатності ІКСМД розрізняти об’єкти одного і декількох класів.

Проведений аналіз методів зниження розмірності простору ДО, який показав, що не існує універсального методу, який одночасно знімав би обмеження на розмірність даної задачі та давав би оптимальне рішення з урахуванням структури взаємозв’язків між всіма показниками і їхньої діагностичної цінності відносно заданої системи діагнозів, не здійснюючи повного перебору. Для рішення даних питань запропоновано розглядати задачу зниження розмірності простору ДО як задачу їхньої ієрархічної кластеризації з позиції потокових моделей. Визначено мету дисертації та сформульовані задачі, які необхідно вирішити в ній.

У **другому розділі** проведений аналіз існуючих приладів і методів вимірювань ГП, який показав, що це дослідження є дуже трудомістким процесом, який вимагає значних витрат часу та

наявності відповідного персоналу. Розроблено структурну схему системи аналізу результатів ГД, яка враховує наявний набір приладів вимірювань, класи захворювань і діагностичну цінність ГО відносно заданої системи діагнозів з метою формування мінімально необхідного набору ДО та забезпечення на його основі якісної діагностики (рис.1).

Визначено функціональні призначення кожної окремої підсистеми даної системи та задачі, які вирішуються в підсистемі формування мінімально необхідного набору ДО, розробка якої є предметом дослідження: аналіз взаємозв'язків між ГП за допомогою коефіцієнтів кореляції, розрахунок яких виконується залежно від типів показників; визначення значимих кореляційних зв'язків за відповідними критеріями; коректне зниження розмірності вихідного простору ГО – формування груп тісно зв'язаних між собою показників та відбір найцінніших з них відносно заданої системи діагнозів і наявного набору приладів вимірювань. Результатом роботи даної підсистеми є мінімально необхідний набір діагностичних ГО, який може коректуватися відповідно до поставленої мети – одержання якісного діагнозу при наявному наборі приладів вимірювання. У випадку, якщо зазначена мета не може бути досягнута, система формує рекомендації: скорочення станів, що діагностуються, тобто менш поглиблена діагностика, або проведення додаткових досліджень, що вимагають наявності нових приладів. Реалізацію підсистеми формування мінімально необхідного набору ДО запропоновано здійснювати за допомогою методу ієрархічної кластеризації ДО на основі потокових моделей, розробка якого наведена у третьому розділі.

**У третьому розділі** на основі потокових моделей запропоновано та сформульовано задачу формування мінімально необхідного набору ДО відносно заданої системи діагнозів з урахуванням наявних приладів вимірювань ГП, що дозволяє одержувати оптимальне рішення, враховувати структуру зв'язків між показниками та зняти обмеження на розмірність задачі. Суть розробленого методу ієрархічної кластеризації ДО полягає в наступному.

1. Структура об'єкта представляється у вигляді графа, вершинами якого є ГП, а дуги характеризують взаємозв'язок між ними. ГП, які розглядаються в роботі, є кількісними ознаками, тому в якості міри зв'язку між двома ГП  $i$  та  $j$  прийнятий коефіцієнт кореляції Пірсона ( $r_{ij}$ ), значення якого є вагою дуги між відповідними вершинами графа. При цьому дугами з'єднуються тільки ті ГП (вершини графа), для яких, відповідно до  $t$ -критерію Стьюдента,  $r_{ij}$  є значимим, що зменшує число дуг графа.

2. Формується потокова задача, рішенням якої є мінімальний розріз графа по дугах, які мають мінімальну пропускну здатність, у якості якої виступає  $r_{ij}$ , що ідентично формуванню груп показників, усередині яких показники мають тісний кореляційний зв'язок, а між групами – слабкий, тобто вирішується задача кластеризації показників. Для потокової задачі з обмеженнями на пропускну здатність дуг виконання мінімального розрізу пов'язане з пошуком оптимальних потоків по всіх дугах ( $f_{ij}$ ), тому в роботі запропоновано формулювання задачі ієрархічної кластеризації ДО у термінах потокової моделі: „Знайти оптимальні потоки дуг графа, значення



яких можуть характеризувати максимально можливий ступінь зв'язку між показниками в даному графі, та виконати ітераційну процедуру мінімального розрізу графа на підграфи за допомогою теореми про максимальний потік і мінімальний розріз". Для рішення даної задачі на основі отриманого графа будується потокова модель шляхом введення фіктивних джерела  $s$  і стоку  $t$ , приєднання до них вершин графа та замикання їх зворотною дугою (див. рис. 2).

3. Пошук  $f_{ij}$  у поточної моделі запропоновано виконувати за допомогою алгоритму „дефекту”, який дає оптимальне рішення шляхом одночасного рішення прямої і двоїстої задач лінійного програмування (ЛП) та вирішує задачу знаходження циркуляції, що мінімізує сумарну вартість потоків по всіх дугах, Пряма задача ЛП формулюється в такий спосіб:

$$\text{мінімізувати } \sum_{(i,j) \in S} c_{ij} f_{ij} \quad (1)$$

при умові, що:

$$\sum_{j \in N} f_{ij} - \sum_{j \in N} f_{ji} = 0, \text{ для всіх } i \in N, i \neq s, i \neq t; \quad (2)$$

$$f_{ij} \leq H_{ij}, (i, j) \in S; \quad (3)$$

$$f_{ij} \geq D_{ij}, (i, j) \in S; \quad (4)$$

$$f_{ij} \geq 0, \quad (5)$$

де  $S$  – множина дуг графа  $G$ ;

$N$  – число вершин (показників) графа;

$H_{ij}$  – верхня границя потоку  $f_{ij}$  по дузі  $(i, j)$ ;

$D_{ij}$  – нижня границя потоку  $f_{ij}$  по дузі  $(i, j)$ .

При цьому обмеження (2–5) мають наступне значення: (2) – зберігання потоку; (3) – обмеження на пропускну здатність дуги  $(i, j)$  зверху; (4) – обмеження на пропускну здатність дуги  $(i, j)$  знизу; (5) – умова позитивності потоку.

Двоїста задача ЛП формується наступним чином:

$$\text{мінімізувати } \sum_{(i,j) \in S} H_{ij} \cdot \lambda_{ij} - D_{ij} \cdot \mu_{ij} \quad (6)$$

при умові, що:

$$\psi_i - \psi_j + \lambda_{ij} - \mu_{ij} \geq -c_{ij} \text{ для всіх } (i, j) \in S; \quad (7)$$

$$\lambda_{ij} \geq 0 \text{ для всіх } (i, j) \in S; \quad (8)$$

$$\mu_{ij} \geq 0, \text{ для всіх } (i, j) \in S, \quad (9)$$

де  $\lambda$  – змінні, що відповідають обмеженням на потоки по дугах зверху у прямій задачі;

$\mu$  – змінні, що відповідають обмеженням на потоки по дугах знизу;

$\psi$  – змінні, які відповідають обмеженням, що описують умову збереження потоку для прямої задачі, можуть приймати довільні значення та не мають обмежень за знаком для всіх  $i \in N$ .

Таким чином, кожної змінної  $f_{ij}$  у прямій задачі відповідає деяке обмеження у двоїстій задачі. При цьому алгоритм "дефекту" дає оптимальне рішення при виконанні наступних умов оптимальності:

$$m1: \text{якщо } \bar{c}_{ij} < 0, \text{ то } f_{ij} = H_{ij};$$

$$m3: \text{якщо } \bar{c}_{ij} = 0, \text{ то } H_{ij} \geq f_{ij} \geq D_{ij};$$

$$m2: \text{якщо } \bar{c}_{ij} > 0, \text{ то } f_{ij} = D_{ij};$$

$$m4: \sum_{j \in N} f_{ij} - \sum_{j \in N} f_{ji} = 0, \text{ для всіх } i \in N \text{ (умова збереження потоку)}.$$

Тут  $\bar{c}_{ij} = c_{ij} + \psi_i - \psi_j$  - скоректована вартість проходження одиниці потоку по дузі  $(i, j)$ .

Якщо одна з умов  $m1-m3$  порушена, то дуга є "дефектною". Рішення є оптимальним, якщо усунуті дефекти всіх дуг і виконано умову збереження потоку. При роботі алгоритму "дефекту" визначаються значення  $\psi_i$  і  $f_{ij}$ , для яких виконані умови оптимальності  $m1-m4$ , за допомогою процедури розміщення позначок. Якщо таких потоків по дугах не існує, то задача не має розв'язання.

4. Для рішення сформульованої задачі ієрархічної кластеризації діагностичних ГО виконано адаптацію параметрів потокової моделі.

4.1. Визначені параметри дуг графа.

*Потоки по дугах*  $f_{ij}$  характеризують оптимальний ступінь зв'язку між ГП, що виступають у якості вершин графа, та обчислюються за алгоритмом „дефекту”. Первісні значення потоків дорівнюють 0, що не порушує умови (2).

*Верхня границя потоку по дузі*  $(i, j)$   $H_{ij}$  характеризує наявний ступінь зв'язку між ГП. У якості неї виступає значення коефіцієнта кореляції  $|r_{ij}|$ .

*Нижня границя потоку по дузі*  $(i, j)$   $D_{ij}$  характеризує відсутність зв'язку між ГП, тобто дорівнює 0. Негативні значення коефіцієнта кореляції не враховуються, тому що в роботі не виконується аналіз напрямку зв'язку між ГП.

*Вартість проходження одиниці потоку по дузі*  $(i, j)$   $c_{ij}$  характеризує те, що кожний показник може в рівнозначному ступені виступати як найцінніший представник групи тісно зв'язаних ознак, яку отримано в результаті кластеризації вихідного простору показників. Її значення приймається рівним одиниці.

4.2. Визначені параметри дуг, які з'єднують фіктивні джерело і стік з вершинами графа, що приєднані до них:  $H_{xy} = \infty$ ,  $D_{xy} = 0$ ,  $c_{xy} = 1$ ,  $f_{xy} = 0$ .

4.3. Визначені параметри зворотної дуги:  $H_{ts} = \infty$ ,  $D_{ts} = 0$ ,  $c_{ts} = -\infty$ ,  $f_{ts} = 0$ .

4.4. З метою побудови потокової моделі розроблений критерій вибору вершин графа, що приєднуються до фіктивного джерела  $s$  і стоку  $t$ :

$$K = \frac{\sum_{i \in N_s} \sum_{j \in N_t} r_{ij}}{\sum_{i, j \in N_s} r_{ij} + \sum_{i, j \in N_t} r_{ij}} \rightarrow \min, \quad (10)$$

де  $N_s$  – число вершин, що приєднуються до  $s$ ;

$N_t$  – число вершин, що приєднуються до  $t$ , при цьому  $N_s = N_t = N_{st}$ ;

$\sum_{i \in N_s} \sum_{j \in N_t} r_{ij} \rightarrow \min$  – характеризує ступінь зв'язку між ГП, що приєднуються до  $s$ , з

показниками, що приєднуються до  $t$ , який має бути мінімальним;

$\sum_{i, j \in N_s} r_{ij} \rightarrow \max$  – характеризує ступінь зв'язку ГП, що приєднуються до  $s$ , між собою, який

має бути максимальним;

$\sum_{i, j \in N_t} r_{ij} \rightarrow \max$  – характеризує ступінь зв'язку показників, що приєднуються до  $t$ , між

собою, який також має бути максимальним.

Таким чином, вершини графа, що приєднуються до фіктивного джерела, повинні мати мінімальний кореляційний зв'язок (або його відсутність) з вершинами, що приєднуються до фіктивного стоку, а між собою вершини, що приєднуються до фіктивного джерела або стоку, повинні мати максимальну кореляцію. Слід зазначити, що забезпечення мінімуму критерію (10), можливо тільки шляхом повного перебору та аналізу значень кореляцій між всіма показниками. Однак, це не припустимо у нашому випадку, тому що більш глобальна задача пошуку мінімального розрізу (кластеризація) виконуються без повного перебору. Тому запропоновано алгоритм вибору вершин, що приєднуються до фіктивного джерела і стоку, який прагне до виконання критерію (10) без повного перебору.

5. По знайденим за алгоритмом „дефекту” оптимальним значенням потоків по дугах виконується ітераційна процедура розрізу графа на підграфи за допомогою теореми про максимальний потік і мінімальний розріз: спочатку вихідний граф розбивається на 2 підграфи, потім кожний з підграфів розбивається ще на два для одержання підграфів другого рівня і т.д. доти, поки виконується розроблений критерій виходу з процедури кластеризації, який дозволить визначити, чи потрібно виконувати подальшу розбивку підграфів  $k$ -го рівня:

$$E_{G_i G_j} = \frac{R_{G_i G_j}}{R_{G_i} + R_{G_j}} \leq T, \quad (11)$$

де  $R_{G_i G_j}$  – кореляція ознак, які належать групі  $G_i$ , с ознаками, які належать групі  $G_j$  (міжгрупова кореляція ознак);

$R_{G_i}$  – кореляція ознак, які належать групі  $G_i$ , друг с другом (внутрішньогрупова кореляція ознак);

$T$  – деяке граничне значення критерію (11).

$T$  прийнято рівним 0,45 на підставі того, що при значенні коефіцієнта кореляції (за Дворецьким) менш 0,3 зв'язок між показниками оцінюється як слабкий, від 0,31 до 0,5 - помірний, від 0,51 до 0,7 - значний, від 0,71 до 0,9 - тісний, 0,91 і вище - дуже тісний. Обране значення  $T$  підтвердилось проведеним експериментом (див. розділ 5).

Міжгрупову кореляцію ознак пропонується обчислювати за формулою:

$$R_{G_i G_j} = \frac{\sum_{i=1}^{N_{G_i}} \sum_{j=1}^{N_{G_j}} r_{ij}}{A_{G_i G_j}}, \quad i \in G_i, \quad j \in G_j, \quad i \neq j,$$

де  $N_{G_i}$  – кількість показників в групі  $G_i$ ;

$A_{G_i G_j}$  – кількість кореляційних зв'язків між показниками групи  $G_i$  і  $G_j$ .

Внутрішньогрупову кореляцію ознак – наступним чином:

$$R_{G_i} = \frac{\sum_{i=1}^{N_{G_i}} \sum_{j=1}^{N_{G_i}} r_{ij}}{A_{G_i}}, \quad i, j \in G_i, \quad i \neq j,$$

де  $A_{G_i}$  – число кореляційних зв'язків між показниками усередині групи  $G_i$ .

Згідно з теоремою про максимальний потік і мінімальний розріз деякий припустимий потік з вузла  $s$  в вузол  $t$  величиною  $F$  є максимальним, а розріз мінімальним, якщо для цього потоку та для деякого розрізу  $R$  величиною  $W_{st}$  виконана рівність  $F=W_{st}$  і справедливі умови:

$$f_{ij} = H_{ij}, \quad \text{для всіх } (i, j) \in R \quad (12)$$

$$f_{ji} = 0, \quad \text{для всіх } (j, i) \in R. \quad (13)$$

Тобто для дуг, які належать мініальному розрізу, характерно наступне: прямі дуги мають максимальне насичення потоків – значення потоків рівні верхній пропускну здатності (12); зворотні дуги мають мінімальне насичення потоків – значення потоків дорівнюють нулю (13).

Таким чином, після розбивки вихідного графа на два підграфи, показники, які належать різним підграфам, мають мінімальний кореляційний зв'язок, а показники усередині підграфів – максимальну кореляцію, що забезпечується виконанням умов (12–13), з яких випливає виконання наступного критерію:

$$J = \frac{\sum_{i, j \in N} f_{ij}}{\sum_{i, j \in N_s} f_{ij} + \sum_{i, j \in N_t} f_{ij}} \rightarrow \min \quad (14)$$

Вираз (14) справедливий завдяки тому, що

$$\sum_{i, j \in N} f_{ij} = \sum_{i, j \in N_s} f_{ij} + \sum_{i, j \in N_t} f_{ij} + \sum_{i \in N_s} \sum_{j \in N_t} (f_{ij} + f_{ji}), \quad (15)$$

тобто, сумарне значення потоків по всіх дугах у потоковій моделі складається із суми потоків по дугах, які належить першому підграфу, суми потоків по дугах, які належить другому підграфу та суми потоків по дугах, по яких робиться розбивка і які виключаються з вихідного графа (доданок 1, 2 та 3 в (15) відповідно). Відповідно до формалізації задачі ієрархічної кластеризації ДО у термінах потокової моделі зв'язок між двома підграфами, які отримано в результаті мінімального розрізу вихідного графа, повинен бути мінімальним:  $\sum_{i \in N_s} \sum_{j \in N_t} (f_{ij} + f_{ji}) \rightarrow \min$ .

На основі отриманих розбивок запропоновано побудову ієрархічної структури діагностичних ГО, елементами якої є групи, у кожній з яких зібрані ГП, які сильно корелюють між собою, за умови мінімального зв'язку між ГП, які належать різним групам. При цьому рівні ієрархічної структури визначаються  $k$  рівнями кластеризації ГП. У групах останнього рівня даної структури запропоновано виконувати відбір по одному представнику кожної групи, який має максимальну діагностичну цінність відносно заданої системи діагнозів з урахуванням трудомісткості вимірювань, відповідно до:

$$I(x_j) = \frac{Z_D(x_j)}{T(x_j)}, \quad (16)$$

де  $T(x_j)$  – трудомісткість вимірювань ГП  $x_j$ , яка визначається експертною оцінкою з урахуванням наявності приладів вимірювань і часом вимірювання;

$Z_D(x_j)$  – діагностична цінність  $x_j$ -го ГП відносно заданої системи діагнозів.

При цьому розрахунок  $Z_D(x_j)$  виконується разом з розрахунком діагностичної цінності  $x_j$ -го ГП відносно кожного діагнозу  $D_i$  –  $Z_{D_i}(x_j)$  у відповідності з теоретико-інформаційним підходом:

$$Z_{D_i}(x_j) = \sum_{s=1}^m \left[ P(x_{js} / D_i) \cdot \log_2 \frac{P(x_{js} / D_i)}{P(x_{js})} \right], \quad (17)$$

$$Z_D(x_j) = \sum_{i=1}^n \sum_{s=1}^m \left[ P(D_i) \cdot P(x_{js} / D_i) \cdot \log_2 \frac{P(x_{js} / D_i)}{P(x_{js})} \right] \quad (18)$$

де  $n$  – кількість діагнозів;

$m$  – кількість діагностичних інтервалів показника  $x_j$ ;

$P(D_i)$  – вірогідність діагнозу  $D_i$  серед всієї групи діагнозів  $D$ ;

$P(x_{js}/D_i)$  – вірогідність наявності  $s$ -го діагностичного інтервалу показника  $x_j$  при діагнозі  $D_i$ ;

$P(x_{js})$  – вірогідність наявності  $s$ -го діагностичного інтервалу показника  $x_j$  у всієї групі діагнозів, що розглядаються.

Для застосування теоретико-інформаційного підходу ГП запропоновано представити трьома непересічними діагностичними інтервалами: нижче норми, норма, вище норми, відповідно до прийнятої у медицині класифікації значень ГП.

Таким чином, розроблено метод ієрархічної кластеризації ДО на основі потокових моделей і методика відбору діагностично цінних ГО, які дозволять з вихідної множини показників виконувати відбір діагностично цінних із них відносно заданої системи діагнозів з урахуванням трудомісткості вимірювань з метою забезпечення якісної діагностики захворювань.

**У четвертому розділі** розроблено ПЗ підсистеми формування мінімально необхідного набору ДО, проведений аналіз засобів створення ПЗ, а також розроблено організацію вхідних і вихідних даних цієї підсистеми та структуру спеціалізованої бази ГП. Підсистема формування мінімально необхідного набору ДО є частиною системи аналізу результатів ГД і частиною ІКСМД, вона реалізована за допомогою методу ієрархічної кластеризації ДО (див. рис. 1). При цьому результати роботи всіх підсистем ІКСМД тісно пов'язані між собою і при спільній взаємодії визначають якість роботи всієї системи, тобто якість діагнозу. У зв'язку із цим розроблено структуру ПЗ ІКСМД системи аналізу результатів ГД (рис. 3); визначені основні призначення блоків даної структури, а реалізацію ПЗ виконано для підсистеми формування мінімально необхідного набору ДО на мові високого рівня C++. Отримане ПЗ дозволяє виконувати настроювання всіх необхідних параметрів потокової моделі та ієрархічну кластеризацію ГО, а також вносити всі необхідні для подальшого використання відомості в базу знань.

**У п'ятому розділі** за допомогою тестового експерименту визначена мінімально необхідна кількість вершин, які підключаються до фіктивного  $s$  і  $t$  ( $N_{st}$ ), з метою формування потокової моделі, на основі якої виконується ієрархічна кластеризація ДО. Встановлено, що  $N_{st}$  лінійно залежить від вихідної кількості вершин графа, тобто від вихідної кількості ГП  $N$ . На підставі цього побудована лінійна функція регресії, яка описує залежність кількості вершин, що приєднуються до фіктивного джерела і стоку, від вихідної кількості ГП і має вигляд:

$$N_{st} = 0,774 + 0,0993 * N, \quad \text{для } N \geq 3 \quad (19)$$

Проведено апробацію розробленого методу ієрархічної кластеризації ДО на основі потокових моделей на реальних даних, у якості яких виступають 16 ГП: 8 показників динамічної в'язкості плазми: VP1, VP2, VP3, VP4, VP5, VP6, VP7, VP8 та 8 показників динамічної в'язкості крові: VK1, VK2, VK3, VK4, VK5, VK6, VK7, VK8, які вимірюються за допомогою капілярного віскозиметра при восьми різних напругах зсуву. За допомогою розробленого методу вихідний простір ГП розподілено в 5 груп, усередині яких дані показники мають тісний кореляційний зв'язок, а зв'язок між показниками різних груп – слабкий. На підставі отриманої розбивки побудована ієрархічна структура ДО, на нижньому рівні якої відібрані 5 діагностично цінних ГО відносно заданих кардіологічних захворювань (рис. 4).

При цьому даний відбір виконувався з урахуванням оцінки діагностичної цінності у відповідності з (17–19) в два етапи: відносно двох груп захворювань: „Здорові” ( $D_{11}$ ) і

„Кардіологічна патологія” ( $D_{12}$ ) – відібрано  $VK1, VP1, VP8, VK3, VK6$ ; та відносно п’яти груп – „Здорові” ( $D_{21}$ ), „Серцева недостатність” ( $D_{22}$ ), „Гіпертонія” ( $D_{23}$ ), „Ішемічна хвороба серця (ІХС)” ( $D_{24}$ ), і „Кардіоміопатія” ( $D_{25}$ ) – відібрано  $VK1, VP1, VP6, VK3, VK5$  (див. рис. 4 та табл. 1).

Таблиця 1

Діагностична цінність гемореологічних показників  
відносно п’яти груп кардіологічних захворювань

№ п/п	Показник, $x_j$	$T(x_j)$	$Z_{D_{21}}(x_j)$	$Z_{D_{22}}(x_j)$	$Z_{D_{23}}(x_j)$	$Z_{D_{24}}(x_j)$	$Z_{D_{25}}(x_j)$	$Z_D(x_j)$	$I(x_j)$	Група
1	<b>VK1</b>	0,4	0,1647	0,3136	0,4841	0,6719	0,1350	<b>0,2693</b>	<b>0,6733</b>	1
2	VK2	0,5	0,0264	0,0984	0,1088	0,2725	0,1656	0,0808	0,1616	
3	<b>VK3</b>	0,5	0,0024	0,0201	0,0040	0,1494	0,0955	<b>0,0266</b>	<b>0,0532</b>	4
4	VK4	0,5	0,0002	0,0241	0,0286	0,0262	0,1811	0,0229	0,0458	
5	<b>VK5</b>	0,5	0,0033	0,0092	0,2135	0,0621	0,0977	<b>0,0393</b>	<b>0,0786</b>	5
6	VK6	0,5	0,0046	0,0272	0,0727	0,0661	0,0613	0,0256	0,0512	
7	VK7	0,5	0,0007	0,0179	0,0445	0,0413	0,1114	0,0202	0,0404	
8	VK8	0,5	0,0024	0,0101	0,0505	0,0490	0,0507	0,0168	0,0336	
9	<b>VP1</b>	0,65	0,0487	0,1691	0,2006	0,3631	0,1051	<b>0,1186</b>	<b>0,1825</b>	2
10	VP2	0,7	0,0183	0,0364	0,1490	0,1132	0,1357	0,0530	0,0757	
11	VP3	0,7	0,0002	0,0442	0,0501	0,0122	0,0383	0,0163	0,0233	3
12	VP4	0,7	0,0003	0,0153	0,0106	0,0007	0,1045	0,0115	0,0164	
13	VP5	0,7	0,0254	0,0043	0,1517	0,0735	0,0027	0,0386	0,0551	
14	<b>VP6</b>	0,7	0,0220	0,0009	0,3544	0,0631	0,0120	<b>0,0572</b>	<b>0,0817</b>	
15	VP7	0,7	0,0002	0,0265	0,0911	0,0238	0,1626	0,0283	0,0404	
16	VP8	0,7	0,0416	0,0892	0,0530	0,0735	0,0027	0,0505	0,0721	

В групах 3 і 5 ГП, які є найціннішими відносно двох і п’яти груп захворювань виявилися різними:  $VP8$  і  $VK6$  для двох груп,  $VP6$  і  $VK5$  для п’яти груп захворювань. Трудомісткість вимірювань  $x_j$ -го ГП ( $T(x_j)$ ) визначено в умовних одиницях експертним шляхом з урахуванням складності методики вимірювань. Апріорні вірогідності даних захворювань за вихідною

вибіркою дорівнюють:  $P(D_{11})=0,567$ ,  $P(D_{12})=0,433$ ;  $P(D_{21})=0,567$ ,  $P(D_{22})=0,155$ ,  $P(D_{23})=0,106$ ,  $P(D_{24})=0,098$ ,  $P(D_{25})=0,074$ .

Для перевірки адекватності застосування методу ієрархічної кластеризації ДО для синтезу комп'ютерного діагнозу за допомогою дискримінантного аналізу виконано розпізнавання п'яти груп кардіологічних захворювань по вихідній множині із 16-ті ГП та по 5-ті відібраним діагностично цінним ознакам (табл. 2).

Таблиця 2

Порівняльний аналіз використання розробленого методу ієрархічної кластеризації діагностичних ознак з методами кластерного аналізу для розпізнавання п'яти груп кардіологічних захворювань

	N	Метод					
		ієрархічної кластеризації ДО	одинок-них зв'язків	повних зв'язків	попарного середнього	центроїд-ний	Ворда
Кількість відібраних ДО (з 16), шт.	16	5	6	5	6	5	5
Відібрані ГО	VK1-VK8, VP1-VP8	VK1, VK3, VK5, VP1, VP6	VK1, VK5, VP1, VP2, VP8, VP6	VK1, VK5, VP2, VP6, VP7	VK1, VK5, VP1, VP2, VP8, VP6	VK1, VK5, VP1, VP2, VP6	VK1, VK5, VP2, VP6, VP7
Доля ДО від N, %	100	31,25	37,5	31,25	37,5	31,25	37,5
Якість розпізнавання 5-ті груп кардіологічних захворювань							
Чуткість	0,792	0,698	0,623	0,642	0,623	0,642	0,642
Специфічність	1,00	0,957	0,928	0,942	0,928	0,928	0,942
Рівень позитивних рішень	1,00	0,925	0,868	0,895	0,868	0,872	0,895
Рівень негативних рішень	0,863	0,805	0,762	0,774	0,762	0,771	0,774
Надійність	0,91	0,844	0,795	0,812	0,795	0,803	0,812
Зниження надійності розпізнавання, %	0	6,6	11,5	9,8	11,5	10,7	9,8

Надійність розпізнавання за відібраними ознаками всього на 6,6% нижче в порівнянні з розпізнаванням тих же класів за вихідною множиною, у той час як число показників зменшилося в 3,2 рази, що свідчить про високу ефективність розробленого підходу з погляду класифікації та ранньої діагностики кардіологічних захворювань.

За отриманими результатами класифікації об'єктів до однієї із груп кардіологічних захворювань проведена порівняльна оцінка якості діагнозу, який розпізнано по діагностично цінним ГО, відібраним за допомогою розробленого методу та за допомогою стандартних методів



кластерного аналізу: одиночних і повних зв'язків, попарного середнього, центроїдного та Ворда. Надійність розпізнавання груп кардіологічних захворювань по діагностично цінним ГО, відібраним при використанні розробленого методу в середньому на 4,2% вище в порівнянні з використанням методів кластерного аналізу. При цьому по відношенню до розпізнавання даних захворювань по 16-ти ГП розроблений метод має найвищі показники якості розпізнавання в порівнянні з методами кластерного аналізу (див. табл. 2).

Таким чином, експериментальна перевірка методу ієрархічної кластеризації ДО на основі потокових моделей, на основі якого розроблено систему аналізу результатів ГД, показала ефективність і працездатність розробленого підходу з точки зору якісного розпізнавання та ранньої діагностики захворювань.

## ВИСНОВКИ

Дисертація спрямована на вирішення актуальної задачі – формування мінімально необхідного простору діагностично цінних ознак відносно заданої системи діагнозів. У роботі запропоновано метод ієрархічної кластеризації ДО на основі потокових моделей, в основі якого лежить ідея перетворення вихідного простору показників шляхом формування груп тісно зв'язаних показників і відбору в кожній з них по одному діагностично цінному відносно заданої системи діагнозів представнику з урахуванням трудомісткості вимірювань; розроблено принципи побудови та ПЗ підсистеми формування мінімально необхідного набору ДО.

1. Проведений аналіз існуючих методів зниження розмірності простору ДО, який показав, що не існує універсального методу, який дає оптимальне рішення задачі відбору діагностично цінних ознак відносно заданої системи діагнозів з урахуванням структури взаємозв'язків між ними без здійснення повного перебору.

2. За результатами оцінки ГД, які спрямовані на ранню діагностику захворювань кровоносної та серцево-судинної систем, встановлено, що вимірювання ГП є трудомістким процесом, який вимагає наявності відповідних приладів. Обґрунтовано необхідність дослідження структури зв'язків між ГП з метою відбору діагностично цінних з них відносно заданої системи діагнозів.

3. Розроблено метод ієрархічної кластеризації ДО на основі потокових моделей, який дозволяє одержувати оптимальне рішення задачі формування груп показників, які мають тісний статистичний зв'язок усередині груп при слабкому їхньому взаємозв'язку між групами. При цьому запропоновано розглядати показники як вершини графа, а задачу ієрархічної кластеризації звести до ітераційної процедури розрізу графа на підграфи по мінімальній сумарній пропускній здатності дуг, у якості якої виступає коефіцієнт кореляції.

4. Виконана адаптація алгоритму "дефекту" для реалізації методу ієрархічної кластеризації ДО, що дозволило зняти обмеження на розмірність вихідної задачі. У рамках адаптації визначені параметри дуг потокової моделі та мінімально необхідна кількість вузлів графа, які підключаються до фіктивного джерела і стоку.

5. Запропоновано методику побудови ієрархічної структури діагностичних ГО мінімально необхідного обсягу, що дозволило врахувати статистичну залежність між даними показниками, їхню діагностичну цінність відносно заданої системи діагнозів і трудомісткість їхніх вимірювань при забезпеченні якісного діагнозу.

6. Розроблена структурна схема системи аналізу результатів ГД на основі ієрархічної кластеризації, яка забезпечує формування мінімально необхідного набору діагностичних ГО відносно заданої системи діагнозів з урахуванням наявних приладів вимірювань, що знижує трудомісткість проведення ГД.

7. Розроблено ПЗ підсистеми формування мінімально необхідного простору діагностичних ГО, робота якої заснована на методі ієрархічної кластеризації ДО.

8. Запропоновано формувати вирішальне правило, яке дозволяє класифікувати об'єкти до одного з заданих класів кардіологічних захворювань, за допомогою дискримінантного аналізу. При цьому критерієм якості ієрархічної кластеризації діагностичних ГО прийнята надійність розпізнавання заданих захворювань по відібраним діагностично цінним ГО.

9. Виконана перевірка адекватності застосування розробленого методу ієрархічної кластеризації ДО на основі потокових моделей для синтезу діагнозу на прикладі п'яти груп кардіологічних захворювань. Надійність розпізнавання даних груп захворювань по відібраним 5-ті діагностично цінним ГО усього на 6,6% нижче в порівнянні з розпізнаванням тих самих класів по вихідній множині ГП, у той час як число показників зменшилося в 3,2 рази, що дозволяє зменшити трудомісткість ГД.

10. Проведений порівняльний аналіз використання методу ієрархічної кластеризації ДО зі стандартними методами кластерного аналізу для задачі класифікації кардіологічних захворювань, який показав ефективність і працездатність розробленого підходу. Якість розпізнавання кардіологічних захворювань по діагностично цінним ГО, відібраним при використанні розробленого методу в середньому на 4,2% вище в порівнянні з використанням методів кластерного аналізу.

### **СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ**

1. *Будянская Э.Н., Поворознюк А.И., Максюта Н.В.* Структурная идентификация диагностических признаков на основе алгоритма „дефекта” // Системи обробки інформації. Зб. наук. пр. – Х: ХВУ, 2003. – Вип. 3. – С. 159-164.

2. *Будянская Э.Н., Поворознюк А.И., Максюта Н.В.* Применение кластерного анализа для структурной идентификации диагностических признаков // Системи обробки інформації. Зб. наук. пр. – Х.: ХВУ, 2004. – Вип. 6. – С. 23-28.

3. *Максюта Н.В., Поворознюк А.И.* Алгоритмы и методы снижения пространства диагностических признаков // Вісник НТУ „ХПІ”. Зб. наук. пр. – Х.: НТУ „ХПІ”, 2005. – № 46. – С. 126–131.

4. *Максюта Н.В., Поворознюк А.И.* Точность постановки диагноза при использовании структурной идентификации диагностических признаков // Вісник НТУ „ХПІ”. Зб. наук. пр. – Х.: НТУ „ХПІ”, 2005. – №56. – С. 169–174.

5. *Максюта Н.В., Поворознюк А.И., Будянская Э.Н.* Структурная схема подсистемы анализа гемореологических показателей // Вісник НТУ „ХПІ”. Зб. наук. пр. – Х.: НТУ „ХПІ”, 2006. – № 40 – С. 116-124.

6. *Максюта Н.В., Поворознюк А.И.* Структурная идентификация диагностических признаков в интеллектуальных системах медицинской диагностики // Контрольно-измерительные приборы и автоматика. – Х.: „Эргос”, 2005. – № 10. – С. 4–14.

7. *Максюта Н.В., Поворознюк А.И.* Синтез диагностических правил на основе структурной идентификации диагностических признаков // Проблеми інформатики і моделювання: Матеріали четвертої міжн. наук.-техн. конф. (25–27 листопада 2004 р.). – Х.: НТУ „ХПІ”, 2004. – С. 6.

8. *Максюта Н.В., Поворознюк А.И.* Структурная идентификация диагностических признаков в интеллектуальных системах медицинской диагностики // Проблеми інформатики і моделювання: Матеріали п'ятої міжн. наук.-техн. конф. (24–26 листопада 2005 р.). – Х.: НТУ „ХПІ”, 2005. – С. 4–5.

9. *Баран (Максюта) Н.В.* Архитектура базы клинических и клинико-лабораторных данных пользователей ВДТ // Інформаційні технології: наука, техніка, технологія, освіта, здоров'я: Анотації доповідей міжн. наук.-техн. конф. (16–17 травня 2002 р.). – Х.: НТУ „ХПІ”, 2002. – С. 340-341.

## АНОТАЦІЇ

Максюта Н.В. Система аналізу результатів гемореологічних досліджень на основі ієрархічної кластеризації діагностичних ознак. – Рукопис.

Дисертація на здобуття наукового ступеня кандидата технічних наук за фахом 05.11.17 – біологічні та медичні прилади та системи. – Харківський національний університет радіоелектроніки, Харків, 2007.

Дисертація спрямована на вирішення актуальної наукової та технічної задачі – розвиток методів ієрархічної кластеризації ДО, які характеризують в'язкісні властивості крові, і проектування ІКСМД. У роботі вперше вирішено задачу ієрархічної кластеризації ДО на основі потокових моделей шляхом її зведення до ітераційної процедури пошуку мінімального розрізу; виконана адаптація параметрів потокової моделі до задачі ієрархічної кластеризації ДО; отримав подальший розвиток метод формування оптимального рішення потокових задач - алгоритм „дефекту”, а також розроблено критерій виходу з ітераційної процедури ієрархічної кластеризації ДО. У дисертації розроблено структуру системи аналізу результатів ГД, а також підсистему формування мінімально необхідного набору діагностичних ГО на основі їх ієрархічної

кластеризації з метою відбору діагностично цінних ГО відносно заданої системи діагнозів з урахуванням трудомісткості та можливості їхніх вимірювань при забезпеченні якісного діагнозу.

Ключові слова: інтелектуальна комп'ютерна система медичної діагностики, гемореологічні показники, ієрархічна кластеризація діагностичних ознак, відбір діагностично цінних ознак, класифікація, потокова модель, кореляційний зв'язок.

Maksyuta N.V. System of the analysis of results gemoreologycal researche on the basis of hierarchical clusterization of diagnostic attributes. – The manuscript.

The dissertation on competition of a scientific degree of technical sciences candidate on a speciality 05.11.17 – Biological and Medical Devices and Systems. – The Kharkov National University of Radioelectronics, Kharkov, 2007.

The dissertation is directed on the decision of an actual scientific and technical problem – development of methods hierarchical clusterization of diagnostic attributes which characterize viscosity properties of blood, and designing intellectual computer system of medical diagnostics. In work the problem of hierarchical clusterization of diagnostic attributes from a position of the stream models theory for the first time is solved by its data to iterative procedure of search of the minimal cut; adaptation of stream models parameters to a problem of hierarchical clusterization of diagnostic attributes is executed; has received the further development a method of reception of the optimum decision stream tasks – algorithm of "defect", and also the criterion of an exit from iterative procedure of hierarchical clusterization of diagnostic attributes is developed. In the dissertation the structure of system of the analysis of results gemoreologycal researche, and also a subsystem of formation of minimally necessary set diagnostical gemoreologycal attributes is developed on the basis of their hierarchical clusterization with the purpose of selection diagnostically valuable gemoreologycal attributes concerning the set system of diagnoses in view of labour input and an opportunity of their measurement at maintenance of the qualitative computer diagnosis.

Key words: intellectual computer system of medical diagnostics, gemoreologycal parameters, hierarchical clusterization of diagnostic attributes, selection diagnostically valuable attributes, classification, stream model, correlation interrelation.

Максюта Н.В. Система анализа результатов гемореологических исследований на основе иерархической кластеризации диагностических признаков. – Рукопись.

Диссертация на соискание ученой степени кандидата технических наук по специальности 05.11.17 – биологические и медицинские приборы и системы. – Харьковский национальный университет радиоэлектроники, Харьков, 2007.

Диссертация направлена на решение актуальной научной и технической задачи – развитие методов иерархической кластеризации диагностических признаков (ДП), которые характеризуют вязкостные свойства крови, и проектирование ИКСМД. Работа посвящена разработке системы анализа результатов гемореологических исследований (ГИ) с целью формирования пространства

ДП минимально необходимого объема с учетом их статистической зависимости, трудоемкости измерения и диагностической ценности относительно заданной системы диагнозов.

Для решения поставленной задачи разработан метод иерархической кластеризации ДП на основе потоковых моделей. При этом предложено рассматривать в виде потоковой модели структуру связей между гемореологическими показателями (ГП): показатели выступают в качестве вершин графа, лежащего в основе потоковой модели, а в качестве дуг графа – корреляционные связи между ними.

Проведен анализ существующих методов измерения ГП, который показал, что ГИ является трудоемким процессом, требующим использования специальных приборов, не всегда имеющихся в наличии в клинических лабораториях из-за своей дороговизны. В связи с этим показана необходимость исследования структуры связей между ГП с целью отбора диагностически ценных из них относительно заданной системы диагнозов, что позволяет уменьшить трудоемкость проводимых исследований.

Отбор диагностически ценных гемореологических признаков относительно заданной системы диагнозов предложено осуществлять с помощью теоретико-информационного подхода с учетом трудоемкости и возможности их измерения в группах тесно связанных ГП. Данные группы получены путем выполнения итерационной процедуры разреза графа, лежащего в основе потоковой модели, на подграфы по минимальной пропускной способности дуг, принадлежащих разрезу, в качестве которой выступает корреляция между ГП.

Для построения потоковой модели и выполнения разреза графа на подграфы выполнена адаптация алгоритма „дефекта“: определены параметры дуг потоковой модели и минимально необходимое количество узлов графа, подключаемых к фиктивному источнику и стоку.

Разработан критерий выхода из итерационной процедуры иерархической кластеризации ДП, основанный на вычислении межгрупповой и внутригрупповой корреляции признаков, и методика построения иерархической структуры диагностических гемореологических признаков с учетом характера связи между ними.

В диссертационной работе впервые разработаны структура, алгоритмы и программное обеспечение подсистемы формирования минимально необходимого набора диагностических гемореологических признаков, в основе которой лежит предложенный метод иерархической кластеризации ДП на основе потоковых моделей. Разработанная подсистема является составной частью системы анализа результатов ГИ и позволяет осуществлять поиск диагностически ценных гемореологических признаков относительно заданной системы диагнозов с учетом трудоемкости и возможности измерения данных показателей при обеспечении качественного диагноза.

Реализованный в подсистеме формирования минимально необходимого набора ДП метод иерархической кластеризации ДП на основе потоковых моделей обеспечивает надежность классификации групп кардиологических заболеваний по отобраным диагностически ценным гемореологическим признакам в среднем на 4,2% выше по сравнению с методами кластерного анализа, позволяющими решать аналогичную задачу.

Ключевые слова: интеллектуальная компьютерная система медицинской диагностики, гемореологические показатели, иерархическая кластеризация диагностических признаков, отбор диагностически ценных признаков, классификация, потоковая модель, корреляционная связь.

Підписано до друку 06.04.2007 р. Формат паперу 60x90/16.  
Папір офсетний. Друк – ризографія. Умовн. друк. арк. 0,9  
Гарнітура Times New Roman. Наклад 100 прим. Зам. № 594069

---

Надруковано у СПДФО Ізрайлев Є.М.  
Свідоцтво № 04058841Ф0050331 від 21.03.2001 р.  
61024, м. Харків, вул. Гуданова, 4/10

---