

КОМП'ЮТЕРНІ ЗАСОБИ, МЕРЕЖІ ТА СИСТЕМИ

У роботі проведено лінійний дискримінантний аналіз даних, отриманих за допомогою методу магнітокардіографії (МКГ). Знайдено діагностичні показники та їх оптимальний набір, що забезпечують найкращу дискримінацію "Здорові – Хворі на ішемічну хворобу серця (ІХС)". Вказано на основні шляхи підвищення достовірності діагностики ІХС за допомогою МКГ за рахунок збільшення потужності дискримінації.

© Д.С. Чернишева, М.М. Будник,
2004

УДК 612.171, 519.237, 519.233.3

Д.С. ЧЕРНИШЕВА, М.М. БУДНИК

ЗАСТОСУВАННЯ ДИСКРИМІНАНТНОГО АНАЛІЗУ ДО ОБРОБКИ МАГНІТОКАРДІОГРАФІЧНОЇ ІНФОРМАЦІЇ

Вступ. Дослідження по застосуванню магнітокардіографії (МКГ) як методу для гарантованої діагностики ішемічної хвороби серця (ІХС) проводяться в Інституті кардіології ім. М.Д. Стражеска на протязі ряду років [1–2]. До переваг МКГ належить безконтактність виконання вимірів; менша (порівняно з ЕКГ) залежність магнітного поля від впливу неоднорідності та анізотропії багатопровідного середовища, яким є тіло людини, висока чутливість до струмів (джерел активності) всередині тіла, можливість їх локалізації у просторі. Це дозволяє створити засоби обробки МКГ даних, що локалізують ішемізовані зони та оцінити вплив на них медикаментозної терапії й лікування [3].

Для реєстрації використано 4-канальний СКВІД-магнітокардіограф [4]. Його програмне забезпечення дозволяє записувати в реальному часі файли даних на жорсткому диску у вигляді набору 36 бінарних файлів для кожної карти, кожний вимір триває 30 с (всього 9 послідовних вимірів по 4 точки в кожному). Одночасно реєструється референтна ЕКГ та генерується реєстраційна картка пацієнта. Після реєстрації файли архівуються та накопичуються під керуванням БД.

Далі проводиться програмна обробка, що складається з трьох основних етапів: попередня, математична і медична [5–6]. Одним із напрямків медичної обробки є кількісний (кореляційний, статистичний) аналіз магнітних карт, що полягає в обчисленні за допомогою математичних методів кількісних параметрів [7].

Дискримінантний аналіз є одним із методів багатомірного статистичного аналізу. Його мета полягає у тому, щоб на основі вимірювань різних характеристик (параметрів) об'єкта класифікувати його, тобто віднести до однієї з декількох груп (класів, кількість яких відома) деяким оптимальним способом. Критерієм оптимальності є мінімум ймовірності хибної класифікації [8]. В роботі використано набір засобів пакету STATISTICA, які забезпечують проведення дискримінантного аналізу даних, візуалізацію та інтерпретацію результатів.

Основна ідея лінійного дискримінантного аналізу (ЛДА) – нехай є n об'єктів, кожний з яких має m характеристик, тобто описується вектором $X_1..X_m$, $m > 1$. Задача полягає в тому, щоб за результатами вимірів зарахувати об'єкт до деякої групи (класу) G_1, \dots, G_k , $k \geq 2$. Тобто, треба побудувати правило, за яким за результатами вимірів параметрів об'єкта вказати групу, до якої він належить. Кількість груп та розподіл об'єктів по групах є априорі відомими. За розміром втрат $r(i, j)$ за неправильної класифікації об'єкта вводяться середні втрати, до яких призводить використання даного правила, і визначається правило, яке мінімізує ці втрати. При цьому задаються ймовірності приналежності об'єкта до певного класу, що оцінюються із масиву експериментальних даних [8].

Постановка задачі. Для впровадження нових діагностичних методик, у тому числі МКГ, в клінічну практику необхідно відшукати діагностичні критерії та розробити вирішуюче правило, що дозволяє зарахувати обстеженого до здорових або до хворих певним захворюванням. З цією метою було досліджено можливості модуля ЛДА пакета STATISTICA для дискримінації хворих на ІХС. Для цього, по-перше, необхідно встановити особливості алгоритмів, реалізованих у зазначеному пакеті до обробки кардіологічних даних, та дослідити залежність потужності дискримінації від кількості та комбінацій параметрів. У кінцевому результаті за допомогою МКГ здорових та хворих на ІХС необхідно встановити залежність потужності дискримінації від числа та різних комбінацій комбінацій параметрів, за які використовуються кількісні показники ІХС.

Аналіз модельних даних. Спочатку досліджувались дві модельні числові вибірки без перекриття даних: в таблицю вхідних даних було набрано цілі числа від 1 до 10 з кроком 1 для першої змінної, та від 20 до 30 – для другої та робота програми по їх розділенню на дві групи. У цьому випадку потужність дискримінації за всіма трьома методами ЛДА – стандартному, покроковим вперед та назад, як і можна очікувати, досягає 100% для обох груп. Для дослідження якості дискримінації для наявності перекриття областей значень змінних у таблицю було введено числа від 1 до 10 для першої змінної та від 8,1 до 17,1 – для другої змінної. Результат дискримінації для обох груп становить 90%, що відповідає критичному значенню модельної змінної в межах $X_{кр} \in (9; 9,1)$.

На рис. 1 наведено залежність потужності від ступеня перекриття змінної. Далі першу та другу групи поміняли місцями за допомогою зміни областей значень змінної (зображено значення потужності дискримінації у відсотках для випадку, коли в таблицю даних занесена першою групою для рис. 1, а – № 1; для рис. 1, б –

№ 2). При цьому було виявлено явище асиметрії – потужність дискримінації залежить від того, яка група записана в таблицю першою (рис. 1, б).

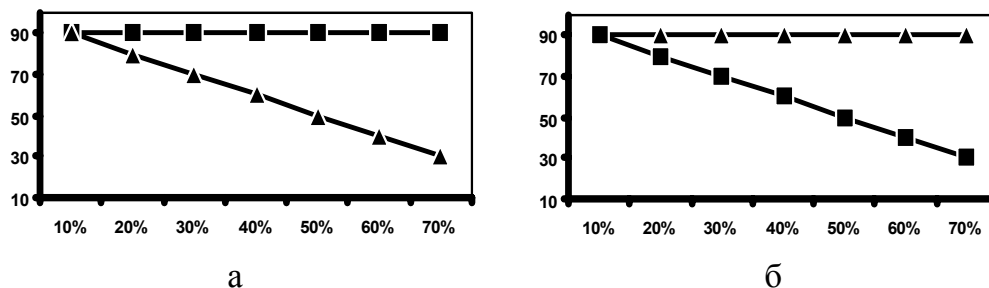


РИС. 1

Далі були проаналізовані тестові дані – три сорти квітів, що описуються 4 змінними. Всі параметри достовірно відрізняються за критерієм Стьюдента. На рис. 2 зображено результат аналізу по кожній змінній окремо, змінні розташовані у порядку зростання r : X1 – довжина пелюстка, X2 – ширина пелюстка, X3 – довжина чашолистика, X4 – ширина чашолистика. На рис. 2 наводиться потужність дискримінації для тестових даних: а) одна змінна; б) набори змінних ($Y2 = \{X1, X2\}$, $Y3 = \{X1, X2, X3\}$, $Y4 = \{X1, X2, X3, X4\}$). Умовні позначення для сортів квітів: \blacklozenge – VIRGINIC, \bullet – VERSICOL, \blacksquare – SETOSA. Як випливає з рис. 2, а, б, високий рівень дискримінації досягається вже за обмеженої кількості параметрів. Наприклад, при врахуванні однієї ознаки мінімальна потужність 92%, а при врахуванні всіх чотирьох ознак (набір Y4) – 96%.

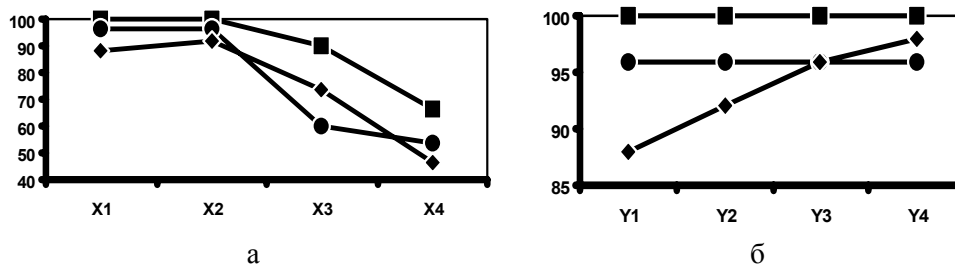


РИС. 2

Аналіз МКГ даних. Обстежений 81 хворий на ІХС та 71 здоровий – всього 152 особи. Для них обчислено 10 МКГ параметрів: МР, ПНС, ПВС, NST, LST, СВП, ПВП, Y, NT, NJ, що характеризують ІХС (див. [1]). Всі параметри впорядковані у порядку зростання r згідно з критерієм Стьюдента (табл. 1). Із них бу-

ло відібрано 6 параметрів X1-X6 (Nst, MP, ПВС, ПНС, Lst, Y), що задовольняють умові $p < 0.01$.

ТАБЛИЦЯ 1. Значення критерію p для МКГ показників

Позначення	X1	X2	X3	X4	X4	X5	X6	X7	X8	X9
Показник	MP	Nst	ПВС	ПНС	Lst	Y	Nj	СВП	Nt	ПВП
p	4,022 E-09	9,749 E-09	3,213 E-07	0,0002	0,002	0,0027	0,0105	0,0823	0,523	0,8059

На рис. 3 зображено залежність потужності дискримінації для різних змінних (рис. 3, а) та різних наборів змінних Zi (рис. 3, б). Спочатку ЛДА проводився по кожній змінній окремо (рис. 3, а). Умовні позначення: \blacklozenge – здорові, \blacksquare – ІХС. З огляду на експериментально отримане на модельних даних явище асиметрії, здорові були завжди 1-ю групою, а хворі ІХС – 2-ю. Параметри на осі абсцис розташовано у порядку зростання p (див. табл. 1). Видно, із збільшенням p для здорових потужність має тенденцію до погіршення, але навіть для найкращого показника вона не перевищує 76%. Така величина замала для надійної дискримінації ІХС щодо здорових. Для хворих такої тенденції немає, а тому необхідно переходити до аналізу у багатомірному просторі.

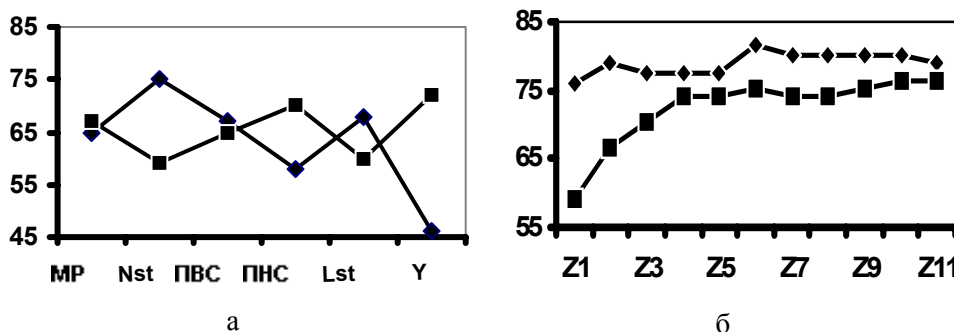


РИС. 3

Таким чином, для підвищення потужності логічно проаналізувати її залежність від кількості змінних. Для цього була побудована така залежність у порядку зростання p, сумарного за всіма змінними z наборів (рис. 3, б):

$$Z_i = \sum_i X_i \tag{1}$$

Із рис. 3 видно, що зі збільшенням кількості параметрів потужність дискримінації збільшується як для здорових, так і для хворих. Звідси випливає, що для досягнення найбільш достовірного розділення здорові – ІХС на основі МКГ показників потрібно 6 змінних, оскільки саме при шести параметрах спостерігається максимум кривої для здорових (81,69%) та локальний максимум для ІХС (75,3%). Проте отриманий набір шести показників може бути неоптимальний, тому що впорядкування згідно з критерієм Стьюдента не є коректним у випадку, коли розподіл вибірок для цих показників відрізняється від нормального.

Далі ЛДА проводився за різними сполученнями параметрів з шести по два, три, чотири та 5-змінним з шести (рис. 4), причому сполучення також впорядковані на осі абсцис по зростанню сумарної величини r .

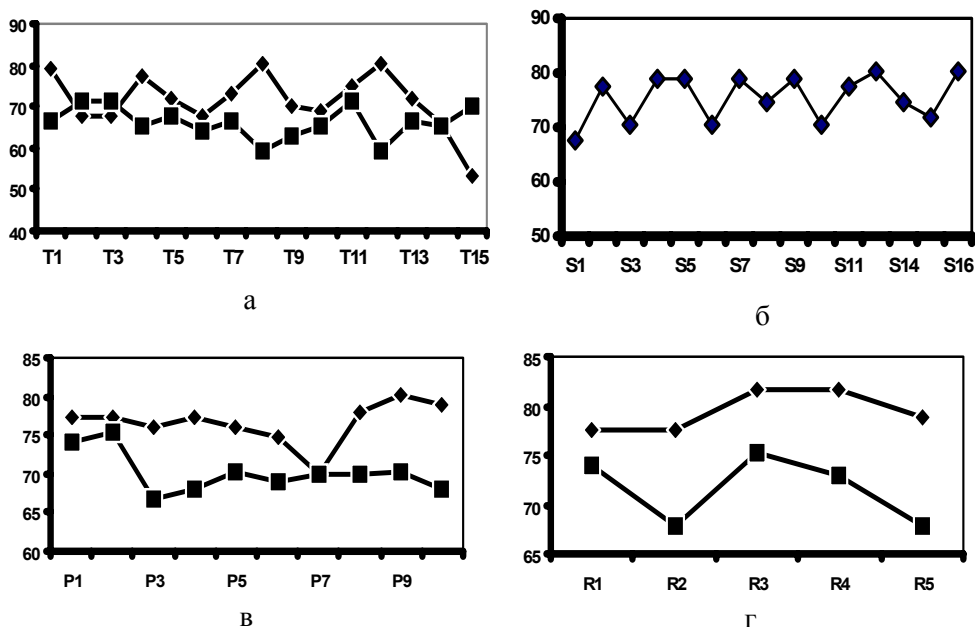


РИС. 4

Із збільшенням кількості змінних залежності набувають більш регулярного характеру (рис. 4). Використовуючи дані рис. 4, а,б,в,г, можна відібрати оптимальні сполучення показників, тобто такі, що дають максимальні значення потужності дискримінації. Оптимальні набори та максимальні значення потужностей наведені в табл. 2–3, а відповідні графіки – на рис. 5.

Найкраща потужність дискримінації при наборі з п'яти параметрів зростає для здорових з 77,46 до 81,69 %, для хворих – з 74,07 до 75,3 % (рис. 5). Це спричинено використанням експериментально визначених оптимальних наборів параметрів, а не наборів, отриманих простим упорядкуванням по r . Отримане

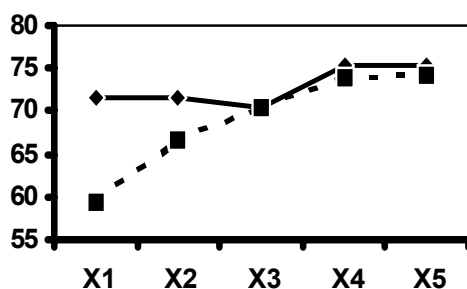
збільшення свідчить про те, що є сенс шукати такі оптимальні набори, особливо для дискримінації здорових.

ТАБЛИЦЯ 2. Потужність дискримінації при оптимізації набору параметрів для здорових

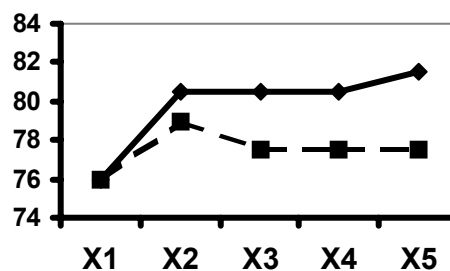
Позначення		T8	S13	P9	R3
Набір змінних	Nst	Nst Lst	MP Nst Y	Y Lst Nst MP	N _{ST} Y L _{ST} ПНС MP
Потужність, %	76,05	80,28	80,28	80,28	81,69

ТАБЛИЦЯ 3. Потужність дискримінації при оптимізації набору параметрів для хворих ІХС

Позначення		T11	S14	P2	R3
Набір змінних	ПНС	MP ПВС	MP ПНС Nst	MP ПВС Nst Lst	NST Y LST ПНС MP.
Потужність, %	71,6	71,6	70,37	75,3	75,3



а



б

РИС. 5

Висновки. Внаслідок досліджень модельних та тестових даних було показано, що потужність дискримінації зменшується при збільшенні зони перекриття областей значень змінних, а також при зменшенні статистичної розрізненості груп (збільшенні p). Такі особливості, як правило, мають місце при обробці медичних даних, у тому числі при обробці магнітокардіографічних показників у хворих на кардіологічні захворювання. Тому для досягнення не-обхідної достовірності дискримінації здорові-хворі ІХС визначалося: які і скільки МКГ показників слід вибрати як координати простору ознак?

Аналіз одномірної дискримінації показує, що її потужність недостатня для надійного виявлення хворих на фоні здорових і навпаки. Для її підвищення були проведені дослідження у двох напрямках, що призводять до збільшення розмірності простору ознак та відрізняються алгоритмом відбору показників. У 1-му

варіанті відбір проводився у порядку зростання p , тобто у порядку погіршення статистичної розрізненості вибірок. Такий простий метод призводить до покращення дискримінації, але, як виявилось, її рівень все ще недостатній для надійної клінічної діагностики цієї складної патології. Проте він дає можливість визначити максимальну кількість ознак, вище якої збільшувати їх кількість недоцільно. Так, знайдено, що потужність дискримінації досягає максимуму за шести параметрів. Це означає, що збільшення простору ознак при дискримінації “хворі ІХС – здорові” більше шести є недоцільним. Тому було використано більш складний варіант, в якому набір показників визначався на основі комбінаторного перебору всіх можливих сполучень з досліджуваного набору 6 МКГ показників. Критерієм відбору було максимальне значення потужності. Внаслідок цього отримано додаткове збільшення потужності – середня потужність становить 78 %.

У табл. 4 наведені оптимальні набори параметрів, за якими дискримінація здорові - ІХС дає найкращий результат. Звідси видно, що при збільшенні кількості параметрів у комбінаціях набори зближуються. Наприклад, якщо при одному параметрі та в наборі з двох параметрів маємо різні показники, то набори з трьох і чотирьох показників відрізняються тільки одним параметром. Набори з п'яти параметрів збігаються для обох типів дискримінації. Якщо порівняти значення потужності дискримінації для оптимального набору з п'яти параметрів та значення потужності дискримінації за всіма шістьма неоптимізованими параметрами, то вони збігаються, а для хворих відрізняються менше ніж на один відсоток. Це дозволяє очікувати, що оптимізація за шістьма параметрами і більше буде давати однакові набори.

ТАБЛИЦЯ 4. Оптимальні набори параметрів

Група	Кількість параметрів				
	1	2	3	4	5
Здорові	X1	T8	S13	P9	R3
	N _{ST}	N _{ST} -L _{ST}	MP-Y-N _{ST}	MP-L _{ST} -N _{ST} -Y	N _{ST} -Y-L _{ST} -ПНС-MP
Хворі ІХС	X4	T11	S14	P2	R3
	ПНС	MP-ПВС	MP-ПВС-N _{ST}	MP-N _{ST} -L _{ST} -ПВС	N _{ST} -Y-L _{ST} -ПНС-MP

З огляду на потреби практичної медицини, для надійної медичної діагностики ІХС необхідне значення потужності близько 90 %. Причина її менших значень полягає в поганих статистичних властивостях МКГ параметрів, які не забезпечують необхідного рівня розрізненості за критерієм Стьюдента. Шлях для вирішення цієї проблеми полягає у відшуканні нових числових МКГ показників, що можуть забезпечити необхідний рівень дискримінації. Інша причина полягає в тому, що людський організм є складною системою, відгук якої на вплив зовнішніх факторів є, в загальному випадку, нелінійним. Тому доцільно також дослідити можливості інших, крім ЛДА, методів дискримінації.

Крім того, результати проведених досліджень вказують на те, що в першому наближенні пошук нових МКГ показників ІХС достатньо проводити згідно з першим напрямком, тобто збільшення простору ознак для дискримінації здійснювати у порядку зростання сумарної величини p . Причина полягає в тому, що при використанні комбінаторного перебору сполучень збільшення потужності для хворих становило тільки 1 %.

1. *Минов Ю.Д.* Разработка и оптимизация импульсно-релаксационной СКВИД-магнетометрической системы для биомагнитных и кардиологических исследований: Дис. ... канд. техн. наук. – Киев, 1996. – 178 с.
2. *Діагностичні критерії хронічної ішемічної хвороби серця на основі реєстрації та аналізу магнітокардіограм* / М.М. Будник, І.Д. Войтович, В.І. Козловський та ін. – К., 2002. – 49 с. – (Препр. / НАН України. Ін-т кібернетики ім. В.М. Глушкова; 2002–5).
3. *Реєстрація та обробка магнітокардіограм у пацієнтів з ішемічною хворобою серця* / В.І. Козловський, М.М. Будник, Л.А. Стаднюк та ін. // Український кардіологічний журнал. – 2002. – № 5. – С. 25 – 34.
4. *First Ukrainian multi-channel magnetocardiograph: assembling and testing* / I. Voytovych, M. Budnyk, Yu. Minov et al // Proc. 2nd IEEE Workshop on Intelligent Data Acquisition and Advanced Computing System (IDAACS'2003). – Lviv, 2003. – P. 16 – 19.
5. *Сосницький В.Н., Будник Н.Н., Войтович І.Д., Минов Ю.Д.* Автоматизированные системы для биомагнитных исследований // УСиМ. – 1995. – № 3. – С. 31 – 46.
6. *Техніка та технологія магнітокардіографії* / Т.М. Березовська, М.М. Будник, О.С. Коваленко та ін. // Електроніка та зв'язок. – 2002. – № 3. – С. 11 – 14.
7. *Абакумов В.Г., Геранін В.О.* Біомедичні сигнали та їх обробка. – К.: БЕК, 1997. – С. 25 – 54.
8. *Феллер В.* Введение в теорию вероятностей и ее приложения. – М.: Мир, 1996. – Т. 2. – С. 354 – 389.

Отримано 02.02.2004