



М.М. Савула¹, Я.М. Матвійчук², В.К. Паучок³

¹Тернопільський державний медичний університет імені І.Я. Горбачевського

²Національний університет «Львівська політехніка»

³Тернопільський національний економічний університет

Диференціальне діагностування деяких захворювань органів дихання з використанням дискримінантного аналізу даних первинного медичного обстеження

Мета дослідження — вивчити придатність дискримінантного аналізу для діагностування найпоширеніших захворювань органів дихання.

Матеріали та методи. Значення 26 показників обстеження 230 пацієнтів, методи дискримінантного аналізу та регуляризованої апроксимації.

Результати та обговорення. За даними первинного медичного обстеження, зміненими нелінійним перетворенням координат і доповненими апроксимаційними значеннями показників мінеральної щільності кісткової тканини, за допомогою дискримінантного аналізу отримано задовільні результати діагностування чотирьох найпоширеніших захворювань органів дихання (рак, туберкульоз, саркоїдоз, пневмонія).

Висновки. Запропонований метод і його програмне забезпечення придатні для практичного застосування в закладах первинної медичної допомоги.

Ключові слова

Диференціальна діагностика, захворювання органів дихання, дискримінантний аналіз, регуляризація.

Математичні методи все ширше використовують для розв'язання медичних проблем, зокрема — для діагностування різних захворювань [2, 7]. Вони не можуть замінити досвід і клінічне мислення лікаря, але, до певної міри, вилучають суб'єктивізм, допомагають молодому лікареві швидко зорієнтуватися, визначити подальшу тактику.

Попередньо ми застосували модифікований кластерний аналіз для діагностування чотирьох найпоширеніших захворювань органів дихання — туберкульозу, саркоїдозу, затяжних неспецифічних пневмоній і раку легень [8]. Отримано задовільні результати, проте вдалося розділити хворих не на чотири (за кількістю захворювань), а на три групи. Поза тим, крім 23 показників стандартного обстеження, було використано два показники мінеральної щільності кісткової тка-

нини (МЩКТ) поперекового відділу хребта (визначені в плані наукового дослідження), які недоступні на рівні первинної медичної допомоги. Тому вирішено вилучити показники МЩКТ з даних обстеження, додати три показники, встановлені простими доступними методами, і застосувати дискримінантний аналіз, який в одиничних випадках використовували для вирішення медичних задач [1, 3].

Мета роботи — вивчити придатність дискримінантного аналізу доступних даних клініко-лабораторних, радіологічних та інструментальних методів досліджень для діагностування найчастіших захворювань органів дихання.

Матеріали та методи

Використано дані обстеження 230 хворих. Рак легень встановлено у 59 (25,65 %), туберкульоз легень — у 67 (29,13 %), саркоїдоз — у 56 (24,34 %), затяжні неспецифічні запальні про-

цеси в легенях — у 48 (20,86 %) хворих. Для характеристик хворих взято набір якісних і кількісних ознак, що стосуються статі, віку, функції зовнішнього дихання (ФЗД), результатів простих лабораторних і загальнодоступних рентгенологічних досліджень (без комп'ютерної томографії).

Для розроблення методу діагностування вибрано дані обстеження 208 хворих з відомим діагнозом. Для його перевірки обрано дані 22 хворих із початково невідомим діагнозом. Виходили з припущення, що на основі даних пацієнтів із відомими діагнозами вдасться встановити невідомі діагнози.

Для обчислень застосовано програмну систему Matlab фірми MatWorks та програмне забезпечення, розроблене авторами. Для введення і зберігання даних — табличний процесор Excel фірми Microsoft.

Результати та обговорення

Для діагностування враховано 26 показників обстеження 230 пацієнтів. Беручи до уваги позитивні результати попереднього дослідження, додатково взято два показники МШКТ (табл. 1).

Поставлено таке завдання. Нам відомі значення 26 показників (див. табл. 1) для 230 пацієнтів. Для 208 пацієнтів також відомі діагнози: рак, саркоїдоз, туберкульоз або затяжна неспецифічна пневмонія. Також відомі значення тих самих 26 показників 22 хворих, але не відомі їхні діагнози. Належить встановити ці невідомі діагнози. Застосовано для розв'язання цієї задачі дискримінантний аналіз [4].

Суть дискримінантного аналізу. Значення 26 показників кожного з 230 пацієнтів задають точку в 26-вимірному просторі. Є підстави припускати, що для хворих на одні й ті ж недуги ці точки утворюють скупчення. А скупчення, утворені групами пацієнтів з різними хворобами, розташовані в різних ділянках цього простору. Центри таких скупчень відображають своєрідні «середні значення показників», притаманних тій чи тій хворобі. Отже, якщо у вибірці даних 230 пацієнтів є пацієнти з чотирма захворюваннями (рак, саркоїдоз, туберкульоз, запалення), тоді за ними можна визначити «центри» цих чотирьох скупчень точок.

Значення 26 показників кожного з хворих із невідомим діагнозом утворюють точки, які лежать на певній відстані від «центрів скупчень» (середніх значень показників) відповідних захворювань. Такі відстані називають «дискримінантами», що дало назву методу аналізу. Якщо точка, утворена значеннями показників окремого пацієнта, лежить ближче до певного «центру

Таблиця 1. Перелік використаних показників

Показник		
1	Вік, роки	
2	Стать, жіноча (0), чоловіча (1)	
3	Маса тіла, кг	
4	Зріст, см	
5	Індекс маси тіла, (маса, кг)/(зріст, м) ²	
6	Початок: поступовий (0), гострий (1)	
7	Тривалість хвороби до звернення, міс	
8	Куріння: ні (0), так (1)	
9	Супровідні хвороби: немає (0), є (1)	
10	Артралгія: немає (0), є (1)	
Клінічний аналіз крові		
11	ШОЕ, мм/год	
12	Лейкоцити, 10 ⁹ /л	
13	Лімфоцити, %	
14	Паличкоядерні гранулоцити, %	
15	Сегментоядерні гранулоцити, %	
16	МБТ у мазку: «—» (0), «+» (1)	
17	Проба Манту, мм	
Рентгенологічні зміни		
18	В одній легені (0), у двох (1)	
19	Вогнища немає (0), є (1)	
20	Інфільтрати: немає (0), є (1)	
21	Деструкція: немає (0), є (1)	
22	Розмір деструкції, см	
23	Збільш. лімфовузлів: немає (0), є (1)	
Показники ФЗД		
24	ЖЄЛ, %	Відносно належних значень
25	ОФВ ₁ , %	
26	МВЛ, %	
МШКТ		
27	L1–2, %	Відносно «молодих, дорослих»
28	L4, %	

скупчень», тоді є більша ймовірність, що у нього відповідна хвороба. І навпаки, якщо дані окремого пацієнта далекі від середніх значень захворювання на певну хворобу, тоді мала ймовірність у нього цієї хвороби.

У нашому прикладі множину 26 показників 208 пацієнтів із відомими діагнозами називають «ідентифікаційною вибіркою». Множину таких же самих показників 22 пацієнтів без встановлених діагнозів називають «тестовою вибіркою». Сукупність значення всіх показників кожного пацієнта називають «спостереженням».

Розгляньмо формальний опис застосування дискримінантного аналізу для поставленої задачі діагностики. Нехай ідентифікаційну вибірку становлять дані m пацієнтів. Тестову вибірку становлять дані M пацієнтів. У всіх спостереженнях є $n = 26$ показників (їх називають «величинами»).

Нехай відомі значення n показників медичного обстеження:

$$x_{ki} (k = \overline{1, m}; i = \overline{1, n}), \quad (1)$$

де: x_{ki} — значення i -го показника для k -го хворого з відомим діагнозом, n — кількість показників; $m = 208$ — кількість хворих з відомим діагнозом. Позначимо ці відомі діагнози:

$$s_k (k = \overline{1, m}), \quad (2)$$

де s_k — назва хвороби, яка має одне з таких значень: «рак», «туберкульоз», «саркоїдоз», «запалення». Дані (1), (2) становлять ідентифікаційну вибірку.

Розгляньмо тестову вибірку. Позначимо результати медичного обстеження хворих з невідомими діагнозами:

$$y_{ki} (k = \overline{1, M}; i = \overline{1, n}), \quad (3)$$

де: y_{ki} — значення i -го показника для k -го хворого з невідомим діагнозом; M — кількість хворих з невідомим діагнозом.

Належить, опираючись на дані (1), (2), встановити невідомі захворювання, які відповідають даним (3) пацієнтів із тестової вибірки:

$$z_k (j = \overline{1, M}), \quad (4)$$

де: z_k — назва хвороби, яка також має одне зі значень: «рак», «туберкульоз», «саркоїдоз», «запалення». Тобто потрібно за даними (1)–(3) встановити діагнози (4).

Спершу було розв'язано описану задачу для $m = 208$ пацієнтів лише на основі 26 показників первинного медичного обстеження (див. табл. 1). Потім розв'язано цю задачу для $m = 140$ пацієнтів із додатковим урахуванням двох показників МЩКТ, тобто на основі 28 показників (див. табл. 1). З'ясувалося, що додаткове урахування показників МЩКТ підвищує точність діагностування. Проте безпосередньо застосувати показники МЩКТ в задачі діагностування небажано, адже це дуже звужує сферу практичного застосування в закладах первинної медичної допомоги.

Тому до 26 величин у спостереженнях (1) з ідентифікаційної вибірки додано два модельні показники:

$$\tilde{x}_{k27}, \tilde{x}_{k28} (k = \overline{1, m}), \quad (5)$$

де $\tilde{x}_{k27}, \tilde{x}_{k28}$ — модельні значення МЩКТ L1–2 і L4 пацієнтів у ідентифікаційній вибірці. Так само до спостережень (3) додано два модельні показники:

$$\tilde{y}_{k27}, \tilde{y}_{k28} (k = \overline{1, M}), \quad (6)$$

де $\tilde{y}_{k27}, \tilde{y}_{k28}$ — модельні значення МЩКТ L1–2 і L4 пацієнтів в тестовій вибірці.

Значення (5), (6) вираховано за допомогою поліноміальної апроксимації показників МЩКТ L1–2, L4 (відомих для 140 пацієнтів), регуляризованої методом за А.Н. Тихоновим [9] та редукцією апроксимаційного базису [5, 6, 10]. Аргументами апроксимаційних поліномів вибрано такі кількісні показники: вік (1); маса тіла (3); зріст (4); тривалість захворювання до звернення

(7); ШОЕ (10); лейкоцити (11); лімфоцити (12); паличкоядерні (13) і сегментоядерні гранулоцити (14); проба Манту (16); розмір деструкції легеневої тканини (21); ЖЄЛ (23); ОФВ₁ (24); МВЛ (25). Позначимо масив індексів цих показників символом $I = \{1, 3, 4, 7, 10, 11, 12, 13, 14, 16, 21, 23, 24, 25\}$ (див. табл. 1).

Показники $\tilde{x}_{k27}, \tilde{x}_{k28} (k = \overline{1, m})$ додано до 26 змінних (1) ідентифікаційної вибірки. Показники $\tilde{y}_{k27}, \tilde{y}_{k28} (k = \overline{1, M})$ додано до 26 змінних (3) тестової вибірки. Кількість величин, взятих для дискримінантного аналізу, збільшено до $\tilde{n} = n + 2$. Таким чином підготовлено дані для розв'язування задачі класифікації з допомогою дискримінантного аналізу на основі 28 показників. Проте цього виявилось недостатньо для точності діагностування.

Щоб підвищити точність розв'язування задачі діагностики, всі кількісні показники $x_{ki} (k = \overline{1, m}), y_{ki} (j = \overline{1, M}), i \in \{1, 5, 27, 28\}$ замінено на такі значення:

$$\begin{aligned} \bar{x}_{ki} &= x_{ki}(1 + F_i(\tilde{x}_{ki})), k = \overline{1, m}, \\ \bar{y}_{ki} &= x_{ki}(1 + F_i(\tilde{y}_{ki})), (j = \overline{1, M}), \end{aligned}$$

де F_i — функція розподілу того показника, визначена за даними (1), (3). Таке нелінійне перетворення координат призводить до взаємного віддалення «скупчень» точок, котрі відповідають пацієнтам з різними захворюваннями. Експериментально встановлено, що це поліпшує результат діагностування. Також значення змінних $i \in \{1, 5, 27, 28\}$ приведено до спільного масштабу нормуванням на одиницю.

За цими даними розв'язано задачу встановлення діагнозів (4) за допомогою дискримінантного аналізу. Результати розв'язання цієї задачі наведено в табл. 2.

Із матеріалів, наведених у табл. 2, випливає, що за допомогою дискримінантного аналізу отримано задовільні результати діагностування у пацієнтів з найпоширенішими захворюваннями легень. Про це свідчать високі показники чутливості і специфічності. Найкращі позитивні результати отримано під час встановлення діагнозу саркоїдозу (показник чутливості 98,2 %), значно нижчі в разі пневмонії (77,8 %) і туберкульозу (79,1 %). Однак ефективність діагностування зменшує значна кількість випадків гіпердіагностики пневмонії (13 хворих) і раку (13 хворих), що утруднює розмежування цих хвороб. Висока також кількість випадків гіподіагностики туберкульозу (14 осіб).

Із 67 хворих на туберкульоз хибно діагностовано рак у 6 (8,96 %), пневмонію — у 6 (8,96 %), саркоїдоз — у 2 (2,99 %) пацієнтів. Із 59 хворих на рак хибно діагностовано пневмонію у 6 (10,17 %), саркоїдоз — у 3 (5,08 %), туберкульоз — у 2 (3,39 %) пацієнтів. Із 48 хворих на пневмонію хибно діагностовано рак у 7 (14,58 %), туберкульоз — у

Таблиця 2. Результати діагностування чотирьох захворювань легень за допомогою дискримінантного аналізу в усіх пацієнтів (230 хворих)

Результати діагностування	Захворювання				Разом
	Рак	Туберкульоз	Саркоїдоз	Пневмонія	
Кількість осіб із захворюванням	59	67	56	48	230
Встановлено правильний діагноз	48	53	55	37	193
Показник чутливості, %	81,36	79,10	98,21	77,08	83,91
Хибно-негативний діагноз (гіподіагностика)	11	14	1	11	37
Хибно-позитивний діагноз (гіпердіагностика)	13	6	5	13	37
Не було у всій групі цього захворювання	171	163	174	182	690
Не діагностовано захворювання в осіб, у яких його не було	158	157	169	169	653
Показник специфічності, %	92,39	96,32	97,13	92,86	94,64

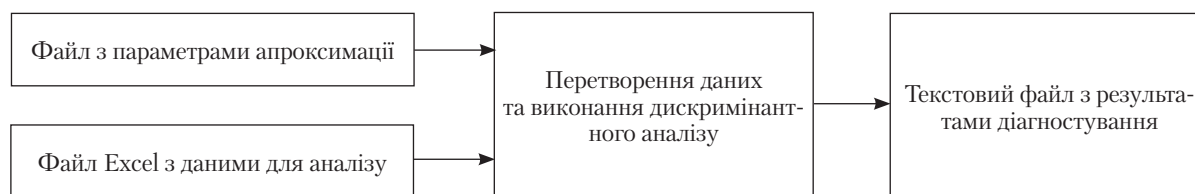


Рисунок. Схема програмної системи діагностики захворювання органів дихання з допомогою дискримінантного аналізу

4 (8,33 %) пацієнтів. Найвищий був результат діагностування саркоїдозу (лише 1 хибний діагноз).

Отже, отримані результати не завжди звільняють лікаря від додаткових обстежень: повторні пошуки МТБ за підозри на туберкульоз, пробне лікування ефективними антибіотиками для заперечення неспецифічного запального процесу, проведення бронхоскопії та КТ в разі безуспішної антибіотикотерапії.

Досі, на жаль, жоден метод не забезпечує 100 % встановлення правильного діагнозу. Тому запропонований метод діагностування найчастіших захворювань легень (рисунок) із використанням показників стандартного медичного обстеження, доступних для первинної ланки медичної допомоги, може бути використаний під час встановлення попереднього діагнозу і для визначення подальшої тактики обстеження та лікування хворого.

На основі описаного методу розроблено програмне забезпечення, готове для практичного використання в медичних закладах. Це програм-

не забезпечення надано для вільного доступу (звертатися за адресою myrko@list.ru).

Висновки

1. Для діагностики чотирьох найпоширеніших легеневих захворювань (рак легень, саркоїдоз, туберкульоз, затяжна пневмонія) застосовано дискримінантний аналіз 26 медичних показників, доступних для первинної медичної ланки.

2. Для вдосконалення методу діагностики застосовано модельні значення показників, вираховані за допомогою регуляризованої поліноміальної апроксимації.

3. Метод може допомогти швидко встановити попередній діагноз на основі доступних показників медичного обстеження.

3. Розроблене програмне забезпечення діагностування найчастіших захворювань органів дихання готове до впровадження і застосування у медичних закладах.

Список літератури

- Атаман Ю.О., Гарбузова В.Ю., Атаман О.В. Застосування дискримінантного аналізу для оцінок кальцифікації артерій нижніх кінцівок // Здобутки клінічної і експериментальної медицини.— 2011.— № 2.— С. 162–165.
- Афанасьєва Е.А. Алгоритмизация процессов ранней дифференциальной диагностики синдрома диабетической стопы: Автореф. дис. ...канд. мед. наук.— Воронеж, 2008.— 20 с.
- Волчегорский И.А., Новиков П.Н., Дударова Т.П., Боло-

тов А.А. Предикторы неэффективности комплексного лечения больных инфильтративным туберкулезом легких // Тер. арх.— 2012.— Т. 84, № 11.— С. 18–25.

- Ким О.Дж., Мьюллер Ч.У., Клекка У.Р. и др. Факторный, дискриминантный и кластерный анализ / Под ред. И.С. Енюкова.— М.: Финансы и статистика, 1989.— 215 с.
- Курганевич А., Матвійчук Я. Регуляризація задач ідентифікації макромоделей нелінійних динамічних систем методом редукції апроксимаційного базису // Теоретична електротехніка.— 2000.— Вип. 55.— С. 31–36.

6. Матвійчук Я.М. Математичне макромодельовання динамічних систем: теорія та практика.— Л.: Видавничий центр ЛНУ ім. Івана Франка, 2000.— 215 с.
7. Насонов С.В. Диагностика и дифференциальная диагностика заболеваний желудка с использованием высокоэффективной жидкостной хроматографии сыворотки крови: Автореф. дис. ...канд. мед. наук.— Нижний Новгород, 2009.— 32 с.
8. Савула М.М., Кравченко Н.С., Паучок В.К. Застосування модифікованого кластерного аналізу даних стандартного медичного обстеження для диференціальної діагностики деяких легеневих захворювань // Укр. пульмонолог. журн.— 2010.— № 4.— С. 39–41.
9. Тихонов А.Н., Гончаровский А.В., Степанов В.В., Ягола А.Г. Регуляризирующие алгоритмы и априорная информация.— М.: Наука, 1983.— 200 с.
10. Янків-Вітковська Л.М., Матвійчук Я.М., Савчук С.Г., Паучок В.К. Дослідження змін координат GNSS-станцій методом макро-модельовання // Вісник геодезії та картографії.— 2012.— № 3.— С. 9–17.

М.М. Савула¹, Я.Н. Матвійчук², В.К. Паучок³

¹Тернопольский государственный медицинский университет имени И.Я. Горбачевского

²Национальный университет «Львовская политехника»

³Тернопольский национальный экономический университет

Дифференциальное диагностирование некоторых заболеваний органов дыхания с помощью дискриминантного анализа данных первичного медицинского обследования

Цель исследования — изучить пригодность дискриминантного анализа для диагностики наиболее распространенных заболеваний органов дыхания.

Материалы и методы. Значения 26 показателей обследования 230 пациентов, методы дискриминантного анализа, методы регуляризованной аппроксимации.

Результаты и обсуждение. По данным первичного медицинского обследования, измененным при помощи нелинейного преобразования координат и дополненным аппроксимационными значениями показателей минеральной плотности костной ткани, получены удовлетворительные результаты диагностирования четырех наиболее распространенных заболеваний органов дыхания (рак, туберкулез, саркоидоз, пневмония) с применением дискриминантного анализа.

Выводы. Предложенный метод и его программное обеспечение пригодны для практического применения в учреждениях первичной медицинской помощи.

Ключевые слова: дифференциальная диагностика, заболевания органов дыхания, дискриминантный анализ, регуляризация.

М.М. Savula¹, Ya.M. Matvijchuk², V.K. Pauchok³

¹Horbachevsky Ternopil State Medical University, Ternopil, Ukraine

²National University Lviv Polytechnic, Lviv, Ukraine

³Ternopil National Economic University, Ternopil, Ukraine

Differential diagnosing of some respiratory diseases by the discriminative analysis of data of primary medical examination

Objective: to study applicability of discriminative analysis for diagnostics of the most widespread respiratory diseases.

Materials and methods. 26 examination indicators data for 230 patients, methods of the discriminative analysis, methods of regularized approximation.

Results and discussion. Recieved data of primary medical examination modified by nonlinear transformation of coordinates and supplemented with approximating values of indicators of mineral density of bones showed satisfactory results of diagnosing of the four most widespread respiratory diseases (cancer, tuberculosis, sarcoidosis, pneumonia) with application of the discriminate analysis.

Conclusions. The method and its software are practicaly applicable in primary healthcare facilities.

Key words: differential diagnosis, respiratory diseases, discriminative analysis, regularization.

Контактна інформація:

Савула Марія Михайлівна, д. мед. н., проф. кафедри пропедевтики внутрішньої медицини та фтизіатрії
46000, м. Тернопіль, Майдан Волі, 3. Тел. (0352) 52-02-41. E-mail: savula@mail.ru

Стаття надійшла до редакції 4 лютого 2013 р.