

## ОБМІН ДОСВІДОМ

УДК 616-037:004.008.1  
DOI 10.11603/1681-276X.2015.3.5204

©П. Р. Сельський

ДВНЗ “Тернопільський державний медичний університет імені І. Я. Горбачевського”

## АЛГОРИТМ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ ОПТИМІЗАЦІЇ ПРОГНОЗУВАННЯ ПЕРЕБІГУ ЗАХВОРЮВАНЬ НА ПЕРВИННОМУ РІВНІ З КОМПЛЕКСНИМ ЗАСТОСУВАННЯМ ІННОВАЦІЙНИХ ІНФОРМАЦІЙНИХ МЕТОДИК

АЛГОРИТМ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ ОПТИМІЗАЦІЇ ПРОГНОЗУВАННЯ ПЕРЕБІГУ ЗАХВОРЮВАНЬ НА ПЕРВИННОМУ РІВНІ З КОМПЛЕКСНИМ ЗАСТОСУВАННЯМ ІННОВАЦІЙНИХ ІНФОРМАЦІЙНИХ МЕТОДИК – У роботі запропоновано алгоритм прийняття рішень для оптимізації прогнозування перебігу захворювань на первинному рівні надання медико-санітарної допомоги. Підхід ґрунтується на комплексному застосуванні інформаційних методик: обчислення коефіцієнтів кореляції, багатопараметричної нейромережевої кластеризації, ROC-аналізу, індукції дерева рішень.

АЛГОРИТМ ПРИНЯТИЯ РЕШЕНИЙ ДЛЯ ОПТИМИЗАЦИИ ПРОГНОЗИРОВАНИЯ ТЕЧЕНИЯ ЗАБОЛЕВАНИЙ НА ПЕРВИЧНОМ УРОВНЕ С КОМПЛЕКСНЫМ ПРИМЕНЕНИЕМ ИННОВАЦИОННЫХ ИНФОРМАЦИОННЫХ МЕТОДИК – В работе предложен алгоритм принятия решений для оптимизации прогнозирования течения заболеваний на первичном уровне оказания медико-санитарной помощи. Подход основан на комплексном применении информационных методик: вычисления коэффициентов корреляции, мультипараметрической нейросетевой кластеризации, ROC-анализе, индукции дерева решений.

THE ALGORITHM OF DECISION-MAKING TO OPTIMIZE THE PROGNOSIS OF THE DISEASE AT THE PRIMARY LEVEL TO THE INTEGRATED USE OF INNOVATIVE METHODS OF INFORMATION – The article adduces the algorithm of decision-making to optimize the prognosis of the disease at the primary level of health-care services. The approach is based on the integrated use of information methods: calculating the correlation coefficients, multiparameter neural network clustering, ROC-analysis, decision tree induction.

**Ключові слова:** алгоритм прийняття рішень, первинний рівень медико-санітарної допомоги, нейромережева кластеризація, ROC-аналіз, дерево рішень.

**Ключевые слова:** алгоритм принятия решений, первичный уровень медико-санитарной помощи, нейросетевая кластеризация, ROC-анализ, дерево решений.

**Key words:** decision algorithm, the primary health care level, neural network clasterization, ROC-analysis, decision tree induction.

**ВСТУП** Інформатизація роботи сільських лікувальних закладів дає можливість суттєво підвищити якість надання первинної медичної допомоги [1]. Численні дослідження спрямовані на впровадження інноваційних медичних інформаційних технологій [2–5], проте не до кінця вирішеною залишається проблема їх використання у сільській медицині. При цьому важливим є оптимізація прогнозування перебігу захворювань з метою покращення лікувально-профілактичної роботи на первинному рівні із застосуванням недорогих та простих у використанні інформаційних методик.

Метою роботи є проаналізувати результати обстеження пацієнтів із гіпертензією на основі кореляційних показників та багатопараметричної нейромережевої кластеризації, ROC-аналізу та дерева рішень з метою розробки алгоритму прийняття рішень для оптимізації прогнозування перебігу захворювання на первинному рівні надання медико-санітарної допомоги.

**МАТЕРІАЛИ І МЕТОДИ** В якості даних для аналізу використано результати обстеження 63 хворих із гіпертензією у навчально-практичних центрах первинної медико-санітарної допомоги (НПЦПМСД), відкритих ДВНЗ “Тернопільський державний медичний університет імені І. Я. Горбачевського” у селах Гнилиці та Зарубинці Тернопільської області [6], внесених у базу програми “Реєстратура” протягом 2011 та 2012 років. Групу контролю склали 19 жителів даних населених пунктів, у яких, за даними бази програми “Реєстратура”, не зафіксовано жодної патології. Статистичну обробку матеріалу проводили з використанням пакета програм “Microsoft Excel” (Microsoft Office 2003). Статистична значущість різниці між середніми арифметичними та відносними величинами оцінювалась за критерієм Стюдента–Фішера (t). При порівнянні однотипних груп проводили кореляційний аналіз із врахуванням коефіцієнта кореляції (r) за допомогою методу квадратів Пірсона.

З метою оптимізації прогнозування перебігу захворювання у хворих із гіпертензією використано аналіз за допомогою ROC-кривих. ROC-крива (англ. receiver operating characteristic) – графік, що дозволяє оцінити якість бінарної класифікації, та відображає співвідношення між часткою вірних позитивних класифікацій від загального числа позитивних класифікацій (англ. true positive rate – TPR (sensitivity)) з часткою помилкових позитивних класифікацій від загального числа негативних класифікацій (англ. false positive rate – FPR (1-specificity)) при варіюванні порога вирішального правила [7, 8].

Для більш глибокого аналізу показників обстеження з метою прогнозування перебігу захворювання використано нейромережевий підхід з використанням надбудови “NeuroXL Classifier” для програми “Microsoft Excel”. Нейромережа є перевіреною і досить розповсюдженою технологією для розв’язування комплексних класифікаційних проблем. Вони моделюються на основі людського мозку та є взаємно пов’язаними мережами незалежних процесорів, які змінюючи зв’язки (цей процес відомий як навчання), навчаються розв’язуванню проблеми. Програма “NeuroXL Classifier” (розробка компанії “AnalyzerXL”) реалізує самоорганізаційні нейромережі, які виконують категоріювання шляхом вивчення трендів та взаємозв’язків усередині даних. Незважаючи на високу ефективність, нейромережі часто не використовуються в силу своєї складності й навчання, яке вимагається для їх правильної реалізації. “NeuroXL Classifier” усуває такі бар’єри, приховуючи складність методів на основі нейромережі і використовуючи переваги використання робочих книг “Microsoft Excel” [9].

Одним із підходів, що відображає природний процес мислення при диференційній діагностиці, є метод індукції дерева рішень. У результаті застосування методики можна отримати множину правил, що йдуть від кореня до кожного термінального вузла, містять нерівності для чисельних атрибутів та умови включення для категоріальних

атрибутив. За основу ми взяли саме таку рекурсивну процедуру роботи [10]. Метод реалізовано в середовищі розробки "Netbeans" мовою програмування "Java". Базу навчальних даних розгорнуто на сервері "MySQL".

**РЕЗУЛЬТАТИ ДОСЛІДЖЕНЬ ТА ЇХ ОБГОВОРЕННЯ** У НПЦПМСД сіл Гнилиці та Зарубинці проведено аналіз даних обстеження 63 хворих, серед яких – 15 чоловіків та 48 жінок. Пересічний вік хворих складав (64,30±1,81) року. Аналіз результатів обстеження пацієнтів із гіпертензією у НПЦПМСД на основі середніх значень, кореляційних показників виявив статистично достовірну відмінність частоти пульсу в групах із стабільним перебігом хвороби порівняно з групою з погіршенням стану (табл. 1). Встановлено, що рівень тахікардії є суттєвим вихідним показником, який вказує на більшу ймовірність погіршення стану та, зокрема, розвитку ускладнень. Кореляційний аналіз виявив прямий кореляційний зв'язок між показниками гемодинаміки при першому та повторному обстеженні пацієнтів у групах із стабільним перебігом (пульс - + 0,5, артеріальний тиск: нижній - + 0,3, пульсовий - + 0,1) та погіршенням стану (пульс - + 0,6, артеріальний тиск: верхній - + 0,5, нижній - + 0,7, пульсовий - + 0,3).

Таким чином, виявлений прямий кореляційний зв'язок між переважною більшістю показників гемодинаміки при першому та повторному обстеженнях пацієнтів у всіх групах із гіпертензією показав значення зміни частоти пульсу та артеріального тиску, як об'єктивних маркерів перебігу захворювання, а, отже, й ефективності лікування.

З метою встановлення значення поєднання зміни тих чи інших параметрів для прогнозу перебігу захворювання було здійснено нейромережеву кластеризацію пацієнтів на основі показників статевої приналежності, віку, змін електричної осі серця за результатами електрокардіографічного дослідження, гемодинамічних показників (пульс, верхній, нижній та пульсовий артеріальний тиск) при першому та повторному дослідженнях. Було використано комп'ютерну програму "NeuroXL Classifier". Для алгоритму нейромережевої кластеризації обрано параметри, запропоновані програмою, та кількість кластерів, рівну п'ятьом.

На рисунках 1,а та 1,б наведено деякі результати виконання програми для пацієнтів із гіпертензією.

Аналіз кластерних портретів виявив, що саме поєднання високих показників артеріального тиску (систоличного,

**Таблиця 1. Показники обстеження пацієнтів із гіпертензією у групах зі стабільним перебігом та з погіршенням стану за даними програми "Реєстратура" у НПЦПМСД сіл Гнилиці та Зарубинці**

Група пацієнтів	Кількість обстежених, n	Вік, роки	Положення електричної осі, °	Показник гемодинаміки			
				перше звернення			
				пульс, уд./хв	АТ, мм рт. ст.		
систоличний	діастолічний	пульсовий					
Стабільний перебіг	50	63,76±1,80	37,72±2,00	77,28±1,17	155,80±2,29	93,10±1,01	62,70±1,95
Погіршення стану, розвиток ускладнень	13	66,38±3,74	38,23±2,67	81,92±1,64****	150,77±4,28	92,31±2,40	58,46±3,74
Загальна група	63	64,30±1,81	37,83±1,92	78,24±1,15	154,76±2,29	92,94±1,04	61,83±1,95

Продовження табл. 1.

Група пацієнтів	Кількість обстежених, n	Вік, роки	Положення електричної осі, °	Показник гемодинаміки			
				повторне обстеження в процесі лікування			
				пульс, уд./хв	АТ, мм рт. ст.		
систоличний	діастолічний	пульсовий					
Стабільний перебіг	50	63,76±1,80	37,72±2,00	78,62±1,07	145,28±1,84**	88,30±1,04**	56,60±1,36*
Погіршення стану, розвиток ускладнень	13	66,38±3,74	38,23±2,67	78,46±2,20	148,08±5,51	89,23±2,49	58,85±4,86
Загальна група	63	64,30±1,81	37,83±1,92	78,59±1,07	145,86±2,01**	88,49±1,08***	57,06±1,57

- Примітки: 1)\* – p<0,05;  
 2) \*\* – p<0,01;  
 3) \*\*\* – p<0,001 – порівняно з першим зверненням;  
 4) \*\*\*\* – p<0,05 – порівняно з групою із стабільним перебігом.

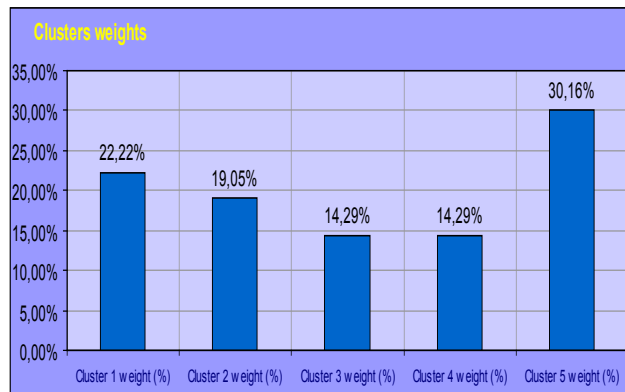
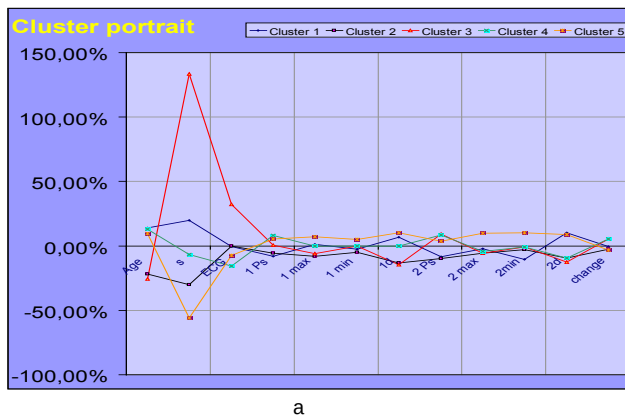


Рис. 1. Результати кластеризації для хворих із гіпертензією за результатами обстеження у НПЦПМСД (усього 63 пацієнти): а) кластерний портрет – середні значення параметрів, включно із показниками гемодинаміки при першому та повторному обстеженнях, у межах розподілених кластерів; б) частки кластерів – відсотки пацієнтів, що потрапили у певний кластер.

діастолічного та пульсового) дає підставу прогнозувати погіршення стану пацієнтів, тоді як поєднання високих показників віку та пульсу (тахікардія) мають суттєве, проте не першочергове значення для прогнозу.

Проведено визначення чутливості та специфічності показників гемодинаміки, зокрема артеріального тиску, при первинному і повторному обстеженнях. З метою дослідження даних показників в якості маркерів включення до груп ризику погіршення стану та розвитку ускладнень у хворих із гіпертензією проведено ROC-аналіз. Встанов-

лено, що при прогнозуванні перебігу гіпертензії на основі поєднаних змін показників артеріального тиску в якості маркерів включення до групи ризику погіршення стану та розвитку ускладнень доцільніше використовувати дані обстеження до призначеного лікування. Так, ROC-аналіз результатів дослідження показників систолічного артеріального тиску показав більшу чутливість при першому обстеженні хворих порівняно з другим. Водночас, на більшості рівнів показники специфічності не різнилися за результатами першого та другого обстежень (рис. 2).

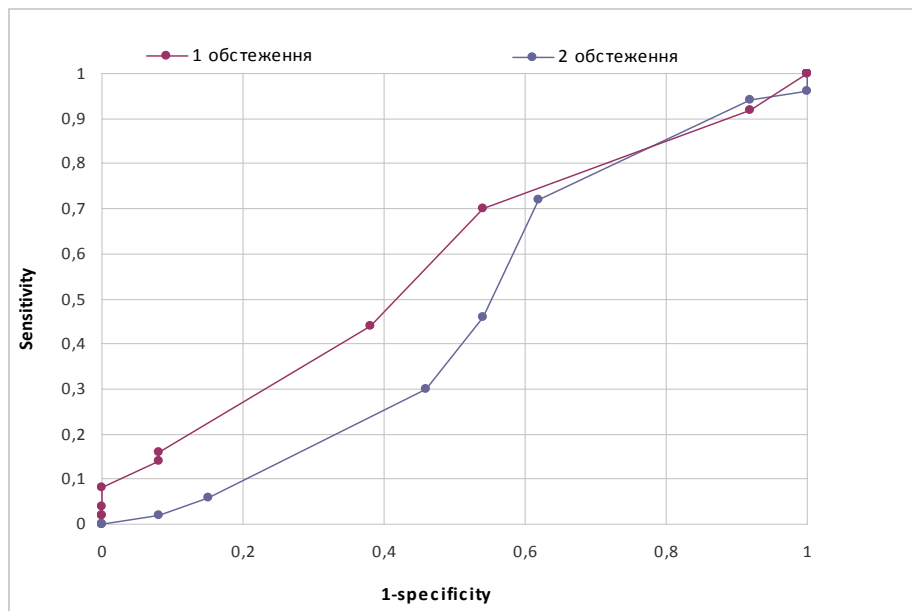


Рис. 2. ROC-аналіз результатів дослідження показників систолічного артеріального тиску при першому та повторному обстеженнях в якості маркерів включення до групи ризику погіршення стану у хворих із гіпертензією.

На заключному етапі дослідження було застосовано методику індукування дерева рішень. Для оптимізації прогнозування перебігу захворювання у НПЦПМСД було побудовано дерево рішень на основі масиву навчальних даних, які включали дані про вік (age), стать (sex), а також наступні показники гемодинаміки при першому та другому обстеженнях: частота пульсу (pulse), систолічний (SAP) та діастолічний (DAP) артеріальний тиск. Використано таку таблицю атрибутів:

```
INSERT INTO mysql.attribute (id, attribute_name, attribute_field_name) VALUES (1, 'What is age?', 'A1'), (2, 'What is sex?', 'A2'), (3, 'What is pulse?', 'A3'), (4, 'What is SAP?', 'A4'), (5, 'What is DAP?', 'A5');
```

Набори включали лише категоріальні дані (попередньо оброблені), наприклад:

```
INSERT INTO mysql.categorised_data (id, A1, A2, A3, A4, A5, class) VALUES (1, 'senior', 'female', 'normal', 'high', 'high', 'healthy');
```

На рисунку 3 представлено побудоване дерево рішень. Час, затрачений на індукування дерева, – 860 мілісекунд.

Тактику поетапного аналізу показників обстеження хворих із гіпертензією запропоновано у вигляді алгоритму прийняття рішень для оптимізації прогнозування захворювання з метою корекції обстеження та лікування (рис. 4).

**ВИСНОВКИ** У роботі запропоновано метод аналізу результатів обстеження пацієнтів із гіпертензією у навчально-практичних центрах первинної медико-санітарної допомоги на основі середніх значень, кореляційних показників, алгоритмів нейромережевої кластеризації, ROC-аналізу та дерева рішень.

З'ясовано, що аналіз на основі середніх значень та обчислення коефіцієнтів кореляції пересічних показників віку, положення електричної осі серця, ряду гемодинамічних показників є первинним інструментом, який не дає встановити значення поєднання зміни тих чи інших параметрів, включно із достовірно відмінними показниками, для прогнозування перебігу захворювання в сторону погіршення чи покращення. Водночас, ефективно

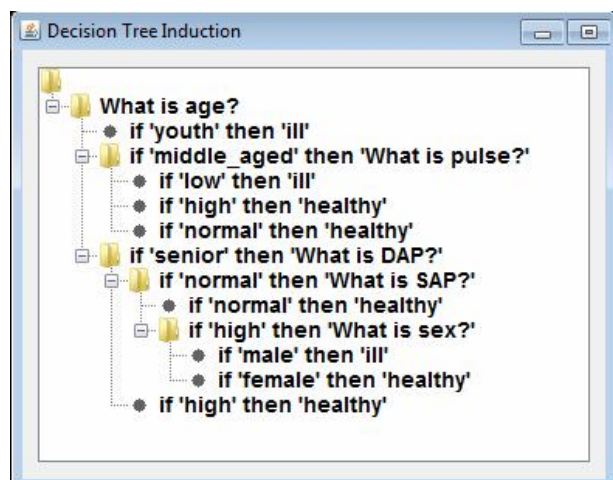


Рис. 3. Дерево рішень для прогнозування перебігу захворювання у хворих із гіпертензією.

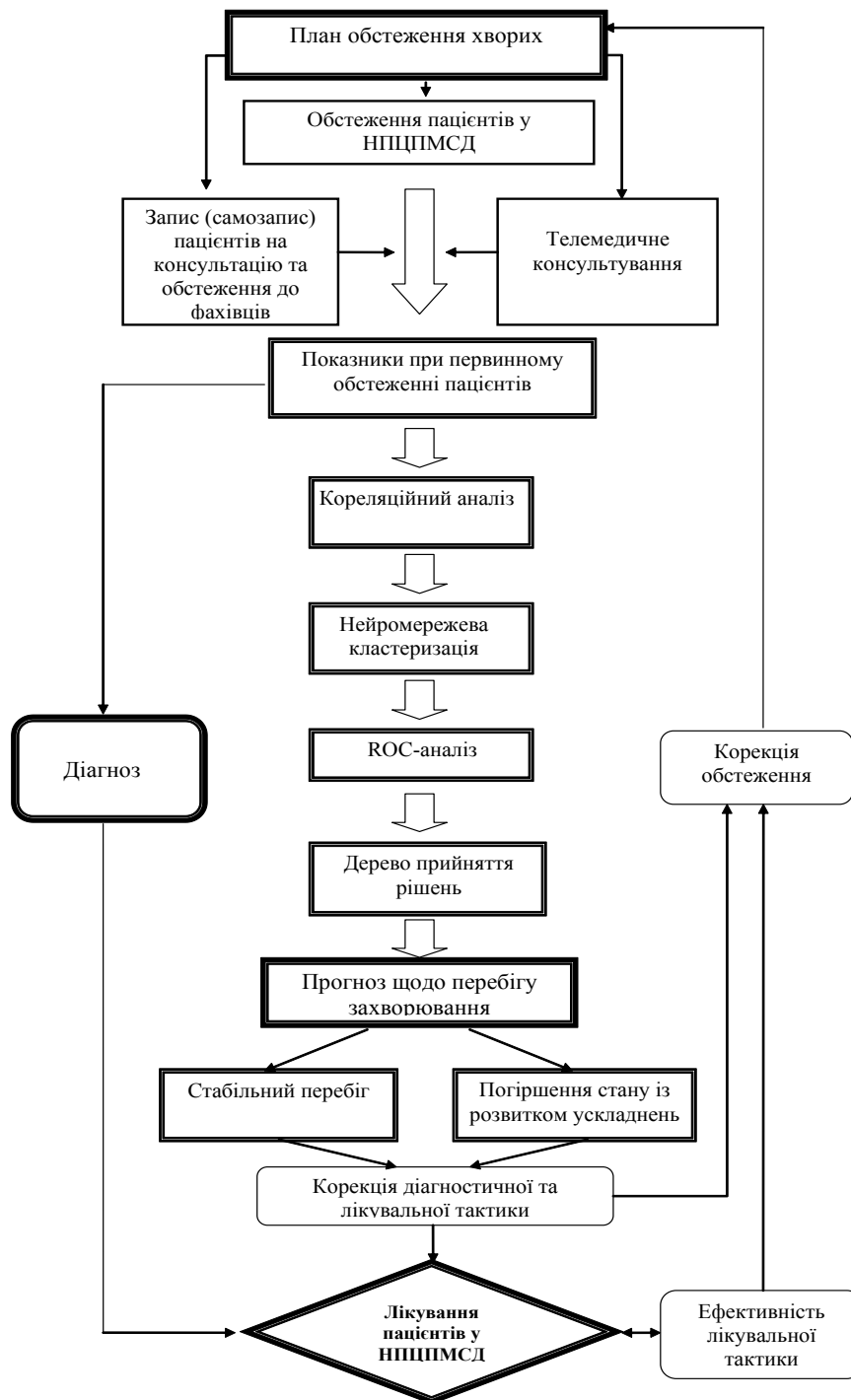


Рис. 4. Алгоритм прийняття рішень для оптимізації прогнозування захворювань на первинному рівні надання медичної допомоги з метою корекції обстеження та лікування.

та об'єктивно розподілити пацієнтів у відповідні категорії за рівнем пересічних показників результатів обстеження дозволяє нейромережева кластеризація.

Запровадження методики ROC-аналізу для прогнозування перебігу захворювань на первинному рівні надання медичної допомоги з метою визначення чутливості та специфічності досліджуваних показників у різні періоди та за різними методиками є перспективним, зважаючи на доступність та простоту у використанні.

Методика індукції дерева рішень впроваджена з метою отримання правил в процесі прийняття рішень, що

відповідають алгоритмам клінічного мислення, і може бути використана для прогнозування перебігу й іншої патології.

Запропонований алгоритм прийняття рішень із комплексним та поетапним застосуванням інформаційних методик, на нашу думку, може бути використаний і для визначення груп ризику розвитку ускладнень при різних захворюваннях, які є найпоширенішими в тому чи іншому регіоні.

Бачиться перспективним комплексне застосування інноваційних інформаційних методик для прогнозування

перебігу захворювань, а, отже, і покращення лікувально-профілактичної роботи на первинному рівні.

СПИСОК ЛІТЕРАТУРИ

1. Концепция информатизации здравоохранения в Украине / О. П. Минцер, Ю. В. Вороненко, Л. Ю. Бабинцева [и др.] // Медична інформатика та інженерія. – 2012. – № 3. – С. 5–29.
2. Минцер О. П. Інформаційно-технологічні проблеми організації телемедичних консультацій / О. П. Минцер, В. В. Краснов, Г. Тахере // Медична інформатика та інженерія. – 2011. – № 4. – С. 32–37.
3. Measuring patient-centered communication in patient-physician consultations: theoretical and practical issues / R. M. Epstein, P. Franks, K. Fiscella [et al.] // Soc. Sci. Med. – 2005. – Vol. 61. P. 1516–1528.
4. Марценюк В. П. О программной среде проектирования интеллектуальных медицинских баз данных / В. П. Марценюк, Н. О. Кравец // Клиническая информатика и телемедицина – 2004. – №1. – С. 47–53.
5. Востров Г. Н. Інформаційна модель надання дистанційних медичних послуг населенню. Перше повідомлення / Г. Н. Востров, О. П. Минцер, О. О. Павлов [та ін.] // Медична інформатика та інженерія. – 2010. – № 3. – С. 37–47.
6. Ковальчук Л. Я. Результати реалізації новітніх методик навчального процесу в Тернопільському державному медичному університеті імені І. Я. Горбачевського та плани на майбутнє / Л. Я. Ковальчук. // Медична освіта. – 2012. – № 2. – С. 11–17.
7. Hanley J. A. Sampling variability of nonparametric estimates of the areas under receiver operating characteristic curves: an update / J. A. Hanley, K. O. Hajian-Tilaki // Academic Radiology. – 1997. – Vol. 4. – P. 49–58.
8. Hilgers R. A. Distribution-free confidence bounds for ROC curves / R. A. Hilgers // Methods of Information in Medicine. – 1991. – Vol. 30. – P. 96–101.
9. Марценюк В. П. Нейромережеве прогнозування складання студентами-медиками ліцензійного інтегрованого іспиту “Крок 1” на основі результатів поточної успішності та семестрового комплексного тестового іспиту / В. П. Марценюк, А. В. Семенець, О. О. Стаханська // Медична інформатика та інженерія. – 2010. – № 2. – С. 57–62.
10. Han J. Data Mining: Concepts and Techniques / J. Han, M. Kamber. – Morgan Kaufmann, San Francisco, 2006. – 800 p.

Отримано 15.06.15