

## АНАЛІЗ РЕЗУЛЬТАТІВ ОБСТЕЖЕННЯ ПАЦІЄНТІВ З ГІПЕРТЕНЗІЄЮ НА ОСНОВІ КОРЕЛЯЦІЙНИХ ПОКАЗНИКІВ ТА ІНДУКУВАННЯ ДЕРЕВА РІШЕНЬ З МЕТОЮ ОПТИМІЗАЦІЇ ПРОГНОЗУВАННЯ ПЕРЕБІГУ ЗАХВОРЮВАННЯ НА ПЕРВИННОМУ РІВНІ

П. Р. Сельський

*ДВНЗ «Тернопільський державний медичний університет  
імені І. Я. Горбачевського МОЗ України»*

У роботі запропоновано методику оптимізації прогнозування розвитку захворювань на первинному рівні. В якості маркерів включення до груп ризику погіршення стану та розвитку ускладнень у хворих з гіпертензією досліджено показники гемодинаміки при первинному і повторному обстеженнях. Підхід ґрунтується на обчисленні коефіцієнтів кореляції та використанні індукування дерева рішень.

**Ключові слова:** первинна медико-санітарна допомога, артеріальна гіпертензія, дерево рішень.

## АНАЛИЗ РЕЗУЛЬТАТОВ ОБСЛЕДОВАНИЯ ПАЦИЕНТОВ С АРТЕРИАЛЬНОЙ ГИПЕРТЕНЗИЕЙ НА ОСНОВЕ КОРЕЛЯЦИОННЫХ ПОКАЗАТЕЛЕЙ И ИНДУКЦИИ ДЕРЕВА РЕШЕНИЙ С ЦЕЛЬЮ ОПТИМИЗАЦИИ ПРОГНОЗИРОВАНИЯ ТЕЧЕНИЯ ЗАБОЛЕВАНИЯ НА ПЕРВИЧНОМ УРОВНЕ

П. Р. Сельский

*ГБУЗ «Тернопольский государственный медицинский университет  
имени И. Я. Горбачевского МЗ Украины»*

В работе предложена методика оптимизации прогнозирования развития заболеваний на первичном уровне оказания медицинской помощи. В качестве маркеров включения в группы риска ухудшения состояния и развития осложнений у больных артериальной гипертензией исследованы показатели гемодинамики при первичном и повторном обследованиях. Подход основан на вычислении коэффициентов корреляции и индукции дерева реше-

**Ключевые слова:** первичная медико-санитарная помощь, артериальная гипертензия, дерево решений.

## ANALYSIS OF THE EXAMINATION RESULTS OF PATIENTS WITH HYPERTENSION BASED ON CORRELATION INDICES AND DECISION TREE TO OPTIMIZE THE PROGNOSIS OF THE DISEASE AT THE PRIMARY LEVEL

P. R. Selskyy

*SHEI «Ternopil State Medical University by I. Ya. Horbachevsky  
Ministry of Public Health of Ukraine»*

The paper suggests an optimization method for the developing disease prognosis at the primary health care level. The hemodynamic indices in primary and re-examination as token inclusion to vulnerable groups of health deterioration and developing the progression in patients with hypertension have been investigated. Approach is based on calculation of correlation coefficients and decision tree.

**Key words:** primary health care, hypertension, decision tree.

**Вступ.** Ефективна інформатизація роботи закладів охорони здоров'я дає можливість суттєво підвищити якість та доступність надання медичних послуг

сільському населенню [1]. Численні дослідження спрямовані на вирішення проблем упровадження інноваційних медичних інформаційних технологій

[2-4], проте не до кінця вирішеною залишаються проблеми їх використання у сільській медицині та відповідної підготовки висококваліфікованих спеціалістів. При цьому важливою є оптимізація прогнозування розвитку ускладнень захворювань з метою корекції обстеження та лікування.

**Мета дослідження.** Запропонувати методіку оптимізації прогнозування перебігу захворювання у хворих на гіпертензію за допомогою кореляційного аналізу та індукування дерева рішень.

**Матеріали та методи.** В якості даних для аналізу використано результати обстеження 63 пацієнтів з гіпертензією у навчально-практичних центрах первинної медико-санітарної допомоги (НПЦПМСД) с. Гнилиця та с. Зарубинці Тернопільської області [5]. У 50 (79,37 %) хворих спостерігали стабільний перебіг хвороби, а у 13 (20,63 %) пацієнтів виявляли погіршення стану та розвиток ускладнень. Здійснювалося дослідження показників пульсу та артеріального тиску за результатами першого (первинного, до призначеного лікування) та другого (повторного, в процесі лікування) обстежень хворих. Статистична обробка матеріалу проводилася з використанням пакета програм "Microsoft Excel" (Microsoft Office 2003).

З метою оптимізації прогнозування перебігу захворювання у хворих на гіпертензію використано метод індукції дерева рішень, що відображає природний процес мислення при диференціальній діагностиці. Протягом кінця 1970-х на початку 1980-х років минулого століття J. R. Quinlan [6] розробив алгоритм побудови дерева рішень ШЗ (ітеративний дихотомайзер). Пізніше J. R. Quinlan представив алгоритм C4.5 (наступник ШЗ), який став еталоном, з яким часто порівнюють новітні алгоритми в галузі машинних знань. У 1984 році група статистиків (L. Breiman, J. Friedman, R. Olshen, C. Stone) опублікували роботу щодо Classification and Regression Trees (CART), в якій описали побудову бінарних дерев рішень. Алгоритми ШЗ та CART, незважаючи на те, що були розроблені незалежно і приблизно у той же час, реалізують подібний підхід до навчання дерев рішень на основі навчальних даних. При цьому дерева рішень будуються в результаті рекурсивної процедури типу «зверху-вниз». Математично задача індукції дерева рішень формулюється таким чином. Маємо множину  $D$ , що містить  $N$  наборів навчальних даних. При цьому кожен  $i$ -й набір  $(A_1^i, A_2^i, \dots, A_p^i, C^i)$  складається з вхідних даних - атрибутів  $A_1, \dots, A_p$  та вихідних даних - атрибута класу

$C$ . Атрибути  $A_1, \dots, A_p$  можуть набувати як чисельних, так і категоріальних значень. Атрибут класу  $C$  приймає одне з  $K$  дискретних значень:  $C \in \{1, \dots, K\}$ . Метою є прогнозування деревом рішень значення атрибута класу  $C$  на основі значень атрибутів  $A_1, \dots, A_p$ . При цьому слід максимізувати точність прогнозування атрибута класу, а саме  $P\{C=c\}$  на термінальних вузлах для довільного  $c \in \{1, \dots, K\}$ . Алгоритми індукції дерев рішень автоматично розбивають на вузлах значення чисельних атрибутів  $A_i$  на два інтервали:  $A_i \leq x_j$  та  $A_i > x_j$ , а категоріальних атрибутів  $A_i$  на дві підмножини:  $A_j \in S_j, A_j \notin S_j$ . Розбиття чисельних атрибутів ґрунтується на мірах на основі ентропії. Процес розбиття рекурсивно повторюється доти, поки не спостерігатиметься покращення точності прогнозування. Останній крок передбачає вилучення вузлів для уникнення оверфітінгу моделі. У результаті можна отримати множину правил, які йдуть від кореня до кожного термінального вузла, містять нерівності для чисельних атрибутів та умови включення для категоріальних атрибутів. За основу взято таку рекурсивну процедуру роботи [7]. Метод реалізовано в середовищі розробки Netbeans мовою програмування Java. Базу навчальних даних розгорнуто на сервері MySQL.

**Результати та їх обговорення.** У НПЦПМСД загалом було обстежено 63 хворих, серед яких - 15 чоловіків та 48 жінок. Пересічний вік хворих склав (64,30+1,81) років. Пересічне значення положення електричної осі серця було нормальним і становило (37,83+1,92)°. Пересічний показник пульсу склав (78,24+1,15) ударів за хвилину. Показники артеріального тиску при першому зверненні пацієнтів були наступними: систолічний - (154,76+2,29) мм рт. ст., діастолічний - (92,94+1,04) мм рт. ст., пульсовий тиск - (61,83+1,95) мм рт. ст. При повторному обстеженні пересічний показник пульсу ((78,59+1,07) ударів за хвилину) та пульсовий тиск ((57,06+1,57) мм рт. ст.) статистично не різнилися ( $p > 0,05$ ). Показники систолічного і діастолічного артеріального тиску були значно нижчими, порівняно з аналогічними показниками до лікування: систолічний - (145,86+2,01) мм рт. ст. ( $p < 0,01$ ), діастолічний - (88,49+1,08) мм рт. ст. ( $p < 0,001$ ). У 13 (19,40 %) пацієнтів спостерігалось погіршення стану та розвиток ускладнень.

Кореляційний аналіз виявив прямий кореляційний зв'язок між показниками пульсу (+0,5), систолічного (+0,1), діастолічного (+0,4) та пульсового (+0,1) тисків при першому та повторному обстеженнях.

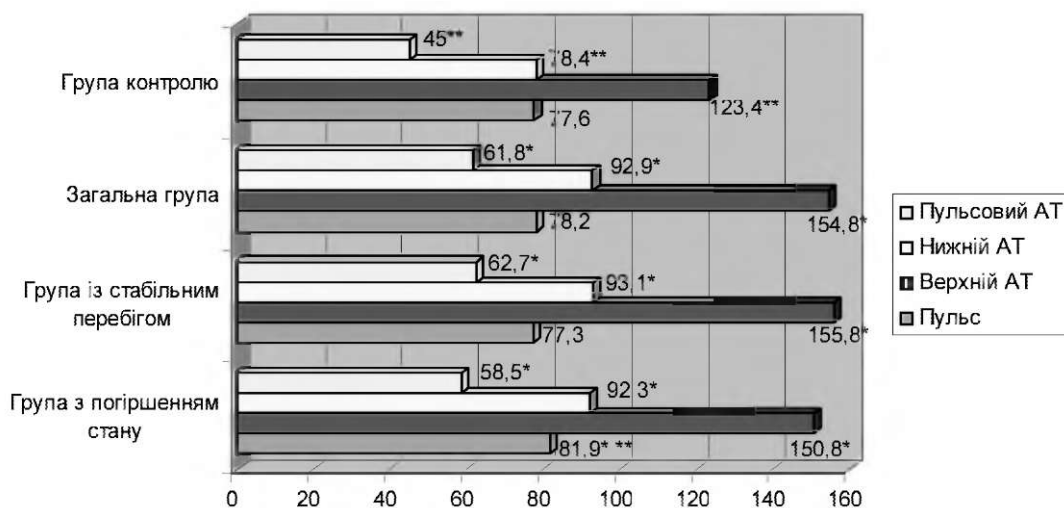
При дослідженні групи 50 пацієнтів із стабільним перебігом захворювання встановлено, що серед них переважали жінки ((72,00+6,35) %). Пересічний вік хворих складав (63,76+1,80) років. Пересічне значення положення електричної осі серця було нормальним і становило (37,72+2,00)°. Пересічний показник пульсу складав (77,28+1,17) ударів за хвилину. Показники артеріального тиску при першому зверненні пацієнтів були наступними: систолічний - (155,80+2,29) мм рт. ст., діастолічний - (93,10+1,01) мм рт. ст., пульсовий тиск - (62,70+1,95) мм рт. ст. При повторному обстеженні пересічний показник пульсу ((78,62+1,07) ударів за хвилину) статистично не різнився ( $p > 0,05$ ). Показники артеріального тиску були значно нижчими, порівняно з аналогічними показниками до лікування: систолічний - (145,28+1,84) мм рт. ст. ( $p < 0,01$ ), діастолічний - (88,30+1,04) мм рт. ст. ( $p < 0,01$ ), та пульсовий тиск - ((56,60+1,36) мм рт. ст. ( $p < 0,05$ )).

Проводили також порівняльний аналіз групи 13 пацієнтів із погіршенням стану, яке визначали на основі таких проявів, як встановлені при обстеженні і зафіксовані в діагнозі тяжча стадія чи ступінь розвитку захворювання або поява ускладнень. Встановлено, що серед даної групи також переважали жінки ((92,31+7,69) %), проте ця частка була достовірно вищою, порівняно з групою із стабільним перебігом захворювання ( $p < 0,05$ ). Пересічний вік хворих суттєво не різнився і складав (66,38+3,74) років. Пересічне значення положення електричної осі серця було нормальним ((38,23+2,67)°) і також суттєво не різни-

лося. Пересічний показник пульсу був значно вищий ( $p < 0,05$ ), порівняно з групою із стабільним перебігом, і складав (81,92+ 1,64) ударів за хвилину. Інші показники гемодинаміки не різнилися, порівняно з аналогічними показниками при стабільному перебігу хвороби ( $p > 0,05$ ). Показники артеріального тиску при першому зверненні пацієнтів були наступними: систолічний - (150,77+4,28) мм рт. ст., діастолічний - (92,31+2,40) мм рт. ст., пульсовий тиск - (58,46+3,74) мм рт. ст. При повторному обстеженні пересічний показник пульсу ((78,46+2,20) ударів за хвилину) статистично не різнився ( $p > 0,05$ ). Не різнилися, порівняно з аналогічними показниками до лікування, і показники артеріального тиску: систолічний - (148,08+5,51) мм рт. ст., діастолічний - (89,23+2,49) мм рт. ст. та пульсовий тиск - ((58,85+4,86) мм рт. ст. ( $p > 0,05$ )).

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).

Показники артеріального тиску при першому обстеженні в усіх групах з гіпертензією виявлялись вищими ( $p < 0,05$ ), порівняно з аналогічними показниками в групі контролю (рис. 1), які становили: систолічний - (123,42+1,49) мм рт. ст., діастолічний - (78,42+1,32) мм рт. ст. та пульсовий тиск - ((45,00+1,58) мм рт. ст. Пересічний показник пульсу



**Рис. 1.** Показники гемодинаміки пацієнтів з гіпертензією в групах із стабільним перебігом, погіршенням стану та загальній групі при першому обстеженні в НПЦПМСД. \* - $p < 0,05$  порівняно з показником групи контролю, \*\* — $p < 0,05$  порівняно з показником групи із стабільним перебігом хвороби.

був суттєво вищим ( $p < 0,05$ ) в групі з погіршенням стану пацієнтів, проте не суттєво різнився ( $p > 0,05$ ) в групі із стабільним перебігом захворювання та загальній групі, порівняно із аналогічним показником групи контролю ((77,58±1,02) ударів за хвилину). Пересічний вік у групі контролю виявлявся нижчим ( $p < 0,05$ ), порівняно з групами із гіпертензією, і складав (38,21±3,91) років.

На заключному етапі вирішення питання щодо включення пацієнтів до групи ризику погіршення стану та розвитку ускладнень було упроваджено методику індукування дерева рішень. Для оптимізації прогнозування перебігу захворювання у НПЦПМСД було використано масив 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 2 представлено побудоване дерево рішень. Час, затрачений на індукування дерева - 860 мілісекунд.

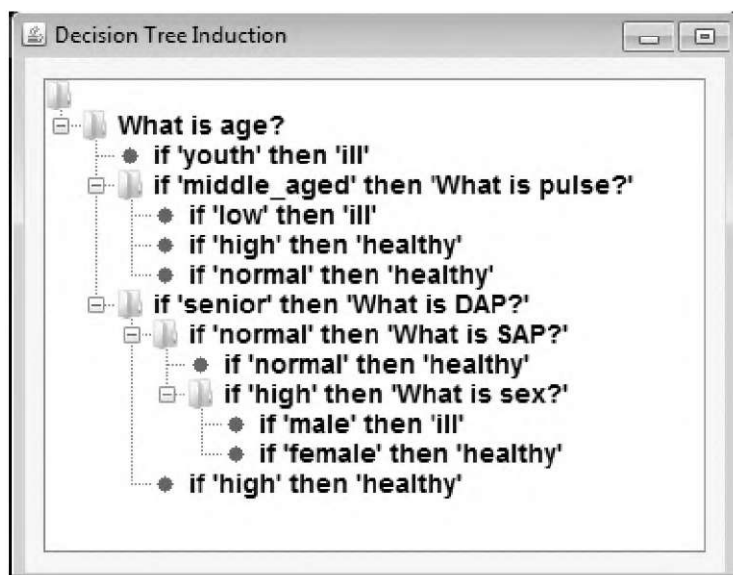


Рис. 2. Дерево рішень для прогнозування перебігу захворювання у хворих на гіпертензію.

**Висновки.** 1. Для оптимізації прогнозування розвитку ускладнень захворювань на первинному рівні з метою корекції обстеження та лікування запропоновано методику аналізу результатів обстеження пацієнтів з гіпертензією у навчально-практичних центрах первинної медико-санітарної допомоги на основі середніх значень, кореляційних показників та індукування дерева рішень.

2. Виявлена статистично достовірна відмінність частоти пульсу в групах із стабільним перебігом хвороби та із погіршенням стану свідчить про те, що рівень тахікардії є суттєвим вихідним показником, який вказує на більшу ймовірність погіршення стану та, зокрема, розвитку ускладнень. Прямий коре-

ляційний зв'язок між переважною більшістю показників гемодинаміки при першому та повторному обстеженнях пацієнтів у всіх групах з гіпертензією показав значення зміни частоти пульсу та артеріального тиску, як об'єктивних маркерів перебігу захворювання, а, отже, і ефективності лікування.

3. Для оптимізації прогнозування перебігу захворювання застосовано індукування дерева рішень на основі масиву навчальних даних, які включали дані про вік, стать та показники гемодинаміки. Отримані правила в процесі прийняття рішень, що відповідають алгоритмам клінічного мислення, можуть бути використані і для прогнозування перебігу іншої патології.

**Перспективи подальших досліджень.** Бачиться перспективним запровадження поєднаного використання кореляційного аналізу та дерева рішень для

прогнозування перебігу захворювань, а, отже, і покращення лікувально-профілактичної роботи у сільських закладах охорони здоров'я.

#### **Література**

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. 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.
3. Марценюк В. П. Ефективність використання інформаційних та телемедичних технологій на первинному рівні надання медичної допомоги / В. П. Марценюк, П. Р. - Сельський // Науково-практична конференція з міжнародною участю "Інформатизація реабілітаційного процесу" : матеріали конференції. - Киев : КВИЦ, 2013. - С. 66-67.
4. 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.
5. 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.
6. Quinlan J. R. Induction of decision trees / J. R. Quinlan // Machine Learning. - 1986,-№ 1.-P. 81-106.
7. Han J. Data Mining: Concepts and Techniques / J. Han, M. Kamber. - Morgan Kaufmann, San Francisco, 2006. - 800 p.