

Medical and Biological Cybernetics

DOI: <https://doi.org/10.15407/kvt211.01.077>

CC BY-NC

КОВАЛЕНКО О.С., д-р мед. наук, проф.,
зав. відд. медичних інформаційних технологій
<https://orcid.org/0000-0001-6635-0124>, e-mail: askov49@gmail.com

КОЗАК Л.М., д-р біол. наук, старш. наук. співроб.,
провід. наук. співроб., відд. медичних інформаційних технологій
<https://orcid.org/0000-0002-7412-3041>, e-mail: lmkozak52@gmail.com

КРИВОВА О.А.,
наук. співроб., відд. медичних інформаційних технологій
<https://orcid.org/0000-0002-4407-5990>, e-mail: ol.kryvova@gmail.com

БИЧКОВ В.В., д-р мед. наук,
старш. наук. співроб., відд. медичних інформаційних технологій
<https://orcid.org/0009-0004-8385-9925>, e-mail: vvb0949@gmail.com

НЕНАШЕВА Л.В.,
молод. наук. співроб., відд. медичних інформаційних технологій
<https://orcid.org/0000-0003-1760-2801>, e-mail: larnen@ukr.net

Міжнародний науково-навчальний центр
Інформаційних технологій та систем
НАН України та МОН України,
40, пр. Акад. Глушкова, м. Київ, 03187, України

ЗАСТОСУВАННЯ КЛАСИФІКАЦІЙНИХ МОДЕЛЕЙ ЗА МЕТОДАМИ DATA MINING ТА ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ДЛЯ АНАЛІЗУ РЕЗУЛЬТАТІВ ЛІКУВАННЯ ПАЦІЄНТІВ КАРДІОЛОГІЧНОГО ТА ДІАБЕТИЧНОГО ПРОФІЛІВ

Вступ. Протягом останніх років наукове співтовариство, особливо у медичній галузі, прикладає багато зусиль задля розроблення технологій та систем електронної охорони здоров'я. Здійснено розроблення різних методів інтелектуальної підтримки, які необхідні для забезпечення високої якості медичної допомоги. Дослідження результативності застосування різних методів діагностування, лікування пацієнтів та відновлення їхнього здоров'я забезпечує один з вагомих складників оцінювання якості медичної допомоги.

Метою статті є аналіз результатів надання медичної допомоги із застосуванням розроблених моделей за методами Data Mining для виявлення факторів, які впливають на результати лікування.

© Видавець ВД «Академперіодика» НАН України, 2023

ISSN 2663-2586 (Online), ISSN 2663-2578 (Print). *Cyb. and Comp. Eng.* 2023. № 1 (211)

77

Результати. Розроблено метод визначення результатів надання медичної допомоги із застосуванням методів *Data Mining*, особливістю якого є поєднання алгоритмів фільтрації, кластеризації та методів класифікації. Побудовано моделі результату надання медичної допомоги залежно від значимих факторів. Для апробації розробленого методу здійснено ретроспективний аналіз з використанням бази даних госпітальних пацієнтів різних відділень клінічних установ. Проаналізовано розподіл результатів лікування пацієнтів кардіологічного та діабетичного профілів, здійснено аналіз наявних супутніх захворювань та ускладнень. Розроблено модель визначення факторів впливу на одержаний результат лікування, побудовану за методом дерева рішень (*CART*). Аналіз структури такого дерева дає можливість зробити висновки щодо логіки прийняття рішень конкретним лікарем. За допомогою моделей дерев рішень проаналізовано зв'язок між ускладненнями, основним діагнозом та іншими чинниками (зокрема, супутні діагнози, повторність госпіталізації тощо).

Висновки. Поєднання статистичних методів аналізу баз медичних госпітальних даних із застосуванням розробленого методу та моделей за методами *Data Mining* (дерево рішень, розрахованого за алгоритмом *CART* та 10-кратної кросс-валідації) дало змогу виявити частотні характеристики супутніх захворювань та ускладнень, типових для пацієнтів кардіологічного та діабетичного профілів, а також уможливило визначення основних чинників, від яких залежить прийняття рішень лікарями про результат лікування.

Ключові слова: електронна охорона здоров'я, методи *Data Mining*, алгоритм *CART*, результати лікування

ВСТУП

Протягом останніх років наукове співтовариство, особливо у медичній галузі, прикладає багато зусиль та ресурсів задля розроблення технологій та систем електронної охорони здоров'я. Інтелектуальні інформаційні системи розглядаються як розширення можливостей інформаційних технологій для підвищення якості та ефективності діяльності системи охорони здоров'я. Незважаючи на певні прогалини у реалізаціях цих проблем, було проведено багато досліджень та здійснено розроблення у напрямку використання різних методів інтелектуальної підтримки, які необхідні для забезпечення високої якості медичної допомоги.

ПОСТАНОВКА ПРОБЛЕМИ

Дослідження результативності застосування різних методів діагностування, лікування пацієнтів та відновлення їхнього здоров'я забезпечує один з вагомих складників оцінювання якості медичної допомоги. Велика когорта дослідників вже більше п'ятдесяти років використовують структуру Донабедіана «Структура-Процес-Результат» для аналізу, оцінювання та керування якістю надання медичної допомоги на різних рівнях: окремих медичних працівників, відділень та закладів системи охорони здоров'я [1–3].

Багато дослідників вбачають, що центральну позицію серед критеріїв оцінювання якості надання медичної допомоги займає критерій результативності лікування, який охоплює такі варіанти оцінювання: повне чи часткове одужання, необхідність подальшого амбулаторного лікування чи реабілітації, набуття інвалідності чи повернення до праці у повному обсязі, а також такі негативні результати, як смертельні випадки. Дістав поширення метод

анкетування пацієнтів щодо результатів лікування, який також дає змогу виявити критерії якості життя та їхнє значення для визначення результатів і загальну оцінку отриманої медичної допомоги [4]. Багато питань щодо якості надання медичної допомоги ґрунтуються на всеосяжній інформаційній обізнаності усіх учасників цього процесу про поточний стан пацієнта, про можливі ризики внаслідок виконання лікувальних процедур тощо [5]. Одна з характеристик систем оцінок якості, що зазнають критики, це те, що багато таких індикаторів належать вимірюваним показникам, і не всі є важливими для кінцевої оцінки результативності, що може мати високу ймовірність загрози процесу лікування [6].

Для оцінювання якості лікування на всіх етапах необхідно визначати об'єктивний стан пацієнта (або ступень тяжкості стану пацієнта, групу ризику). Це дасть змогу на засадах доказової медицини порівняти повноту та результативність лікування, встановити відмінності між різними лікарнями, госпітальними центрами (лікарнями).

В основі більшості моделей оцінювання захворюваності та результатів надання медичної допомоги лежать статистичні моделі (логістична регресія — LR) та методи Data Mining: штучна нейронна мережа (ANN), дерева рішень (DT), метод опорних векторів (SVM). Для оцінювання поточного стану хворого, встановлення прогнозів лікування, оцінювання можливих ризиків (смертності, ускладнень) використовують такі класичні інструменти, як шкали тяжкості (Severity scores). Системи оцінювання кількісно або якісно визначають ступінь тяжкості стану та класифікують пацієнта до конкретних груп ризику на основі аналізу відхилень анатомічних, фізіологічних, біохімічних показників тощо [7–10]. Наразі розроблено понад два десятки шкал оцінювання ступеня тяжкості, однак лише деякі з них можна вважати загальноприйнятими.

За умов епідемії COVID-19 для класифікації та прогнозу тяжкості пацієнтів використовують міжнародну шкалу National Early Warning Score (NEWS2) [11], яку було розроблено для оцінювання тяжкості хворих з гострими респіраторними синдромами [12].

Переважає більшість досліджень вказує на позитивний досвід використання інформаційних технологій (IT), що має значний вплив на результативність та ефективність медичної допомоги, надає перелік об'єктивних критеріїв оцінювання [13–15].

Основою моніторингу поточних результатів лікування часто слугують математичні моделі тяжкості стану, задля їх побудови та валідації може бути використано логістичну регресію. У [16] автори зазначають, що, зважаючи на гетерогенність досліджень, неможливо сказати, яка з математичних моделей є оптимальною для всіх завдань та застосованих шкал тяжкості. Останнім часом все частіше для верифікації та моделювання тяжкості стану пацієнтів стали використовувати методи Data Mining: дерева рішень [17], штучна нейронна мережа (ANN) [18], метод опорних векторів (SVM) [19].

Методи дерев рішень, які було розроблено наприкінці минулого століття, в останнє десятиліття набули широкого розповсюдження в медичних дослідженнях. Вони легко сприймаються професіоналами клінічної практики, тому що є наочними та перетворюються на логічні

умови (вирішувальні правила, правила класифікації). Класифікаційні дерева використовувались, наприклад, для розрахунку ймовірності смерті від коронарної патології [20], внутрішньо-мозкових кровотеч або черепно-мозкових травм [21], для прогнозування стійких вегетативних станів, для стратифікації груп пацієнтів за ймовірністю смертності серед загальної популяції хворих інтенсивної терапії [22].

Зазначено, що основними перевагами дерев рішень є те, що отримані правила прийняття рішень можна легко інтерпретувати, а одержаний у кожному кінцевому вузлу склад групи пацієнтів ("листка" дерева) є відносно однорідним [23].

Метою роботи було аналіз результатів надання медичної допомоги із застосуванням розроблених моделей за методами Data Mining для виявлення факторів, які впливають на результати лікування.

РОЗРОБЛЕННЯ МЕТОДУ ТА МОДЕЛЕЙ ВИЗНАЧЕННЯ РЕЗУЛЬТАТІВ НАДАННЯ МЕДИЧНОЇ ДОПОМОГИ З ВИКОРИСТАННЯМ МЕТОДІВ DATA MINING

Особливий інтерес до проблеми відбору ознак для оцінювання результатів лікування виникає у випадку, характерному для медичних ретроспективних досліджень, коли маємо велику кількість неструктурованих ознак для обмеженої вибірки пацієнтів. Особливістю запропонованого методу є поєднання алгоритмів фільтрації, кластеризації та методів класифікації для побудови моделі результату надання медичної допомоги залежно від значимих факторів. Метод має два основні модулі побудови моделі: модуль А для випадку, коли цільовий атрибут (стандарт оцінки тяжкості стану) є відсутнім, модуль Б — коли для навчання моделі використано одну з стандартизованих шкал оцінок стану пацієнта (Рис. 1).

За розробленим методом виявлення залежності результату надання медичної допомоги, на кінцевому етапі за допомогою дерева рішень формуються діагностичні правила «якщо, то» для граничних значень залежності результату лікування від досліджуваних факторів.

АПРОБАЦІЯ РОЗРОБЛЕНОГО МЕТОДУ ВИЯВЛЕННЯ ЗАЛЕЖНОСТІ РЕЗУЛЬТАТУ НАДАННЯ МЕДИЧНОЇ ДОПОМОГИ ВІД ДОСЛІДЖУВАНИХ ФАКТОРІВ

Для апробації розробленого методу здійснено ретроспективний аналіз з використанням бази даних госпітальних пацієнтів різних відділень клініки ДУ ННЦ "Інститут кардіології ім. акад. М.Д. Стражеска" НАМН України (2018–2021 рр.) та Хмельницької обласної клінічної лікарні (2012–2018 рр.).

Оцінювання результатів лікування пацієнтів кардіологічного профілю

Характеристика використаних масивів даних. Проаналізовано результати надання медичної допомоги у відділеннях гіпертонії, аритмії клініки ДУ ННЦ "Інститут кардіології ім. акад. М.Д. Стражеска" НАМН України.

У відділенні гіпертонії в 2019 році проліковано 999 пацієнтів. Найчастіше траплявся основний діагноз — І11.9 Гіпертензивна (гіпертонічна) хвороба з переважним ураженням серця без (застійної) серцевої недостатності (66,7%). Середня кількість проведених ліжко-днів — 10,8 ліжко-днів (від 1 до 29). Основний діагноз у відділенні аритмії у 2019 році був як під час госпіталізації, так і під час виписки І48 — Фібриляція та тріпотіння передсердь, що складало

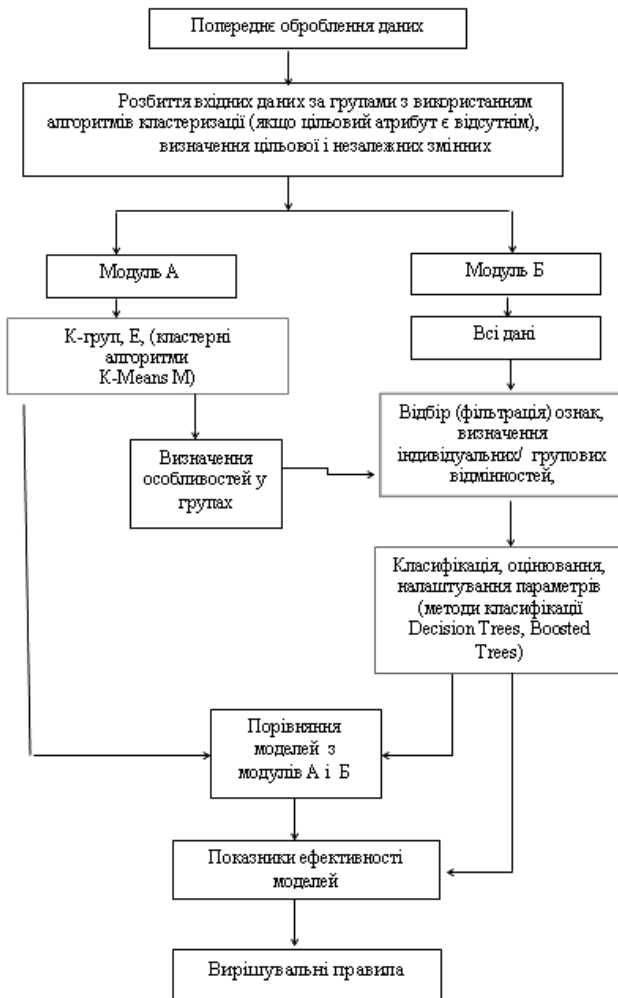


Рис. 1. Блок-схема методу визначення результатів надання медичної допомоги

56,7% від загальної кількості пацієнтів (944). Більшість інших випадків це — діагнози уточнювальних причин порушень серцевого ритму.

Основний діагноз супроводжувався супутніми хворобами. Найчастіше цей діагноз І48 супроводжувався набором таких захворювань: І25 — Хронічна ішемічна хвороба серця (18,7%), І25.1.7 — Атеросклеротична хвороба серця з гіпертонічною хворобою (9,3%), І51.4 — Міокардит, неуточнений (7,7%), або І11 — Гіпертензивна (гіпертонічна) хвороба з переважним ураженням серця (3,8%).

Як видно з наведеного у таблиці 1 розподілу результатів лікування, упорядкованого за частотою для відділень гіпертонії (ВГ) та аритмії (ВА), переважну більшість пацієнтів виписано з поліпшенням (ВГ — 98,2%, ВА — 62,7%).

Таблиця 1. Результат лікування у відділеннях гіпертонії та аритмії

Ранг	Результат	Частка пацієнтів (%)	
		ВГ	ВА
1	Виписаний(а) з поліпшенням	98,2	62,7
2	Виписаний(а) без змін	0,6	34,7
3	Виписаний(а) з одужанням	0,4	0,95
4	Виписаний(а) з погіршенням	0,2	0,74
5	Переведений(а) в інший лікувальний заклад	0,2	0,31
6	Амбулаторне лікування	0,2	0,31
7	Помер(ла)	0,2	0,1

Таблиця 2. Внесок досліджуваних факторів в розроблену модель

Фактори	Вага фактору
Супутній діагноз	100
Лікар	77
Основний діагноз	49
Проведено ліжко-днів	20

Аналіз супутніх захворювань та ускладнень. Треба зазначити, що діагностовано значна кількість супутніх захворювань, від одного (основне супутнє) до семи. Наприклад, діагноз Гіпертонічна хвороба з переважним ураженням серця траплявся майже у 50% пролікованих.

Серед супутніх захворювань значну частину складають діагнози, які не належать до основного профілю лікування — захворювань ССС. Наприклад, цукровий діабет 1, 2 типів та розлади ендокринної системи (152 пацієнти), хвороби легень (23 випадки), травної системи (13), нирок (10). В 2019 р. загальний перелік супутніх захворювань складався з 2692 захворювань за різними кодами МКБ-10.

Модель визначення факторів впливу на одержаний результат лікування побудовано за методом дерева рішень (CART — Classification and Regression Trees) з використанням модуля Data Miner пакету STATISICA. Серед доступних для аналізу факторів у досліджуваній базі даних були такі: основний діагноз, кількість ліжко-днів, супутній діагноз, лікар. Повне дерево, отримане після 10-кратної кросс-валідації, наведено на рис. 2. Внесок в модель кожної змінної (фактору) наведено у таблиці 2, за якою можна зробити висновок про найбільший вплив супутнього діагнозу та конкретного фахівця на одержаний результат лікування.

Така модель дає прогноз за матрицею порівняння класифікацій з середньою точністю — 89%. Наведемо прогнозні значення для деяких класів результату лікування: «1. Амбулаторне лікування» — у 327 пацієнтів (34,7%), для «4 — Виписаний(а) з поліпшенням» — у 596 пацієнтів (62,7%). Класи незбалансовані, малочисельні класи погано передбачаються.

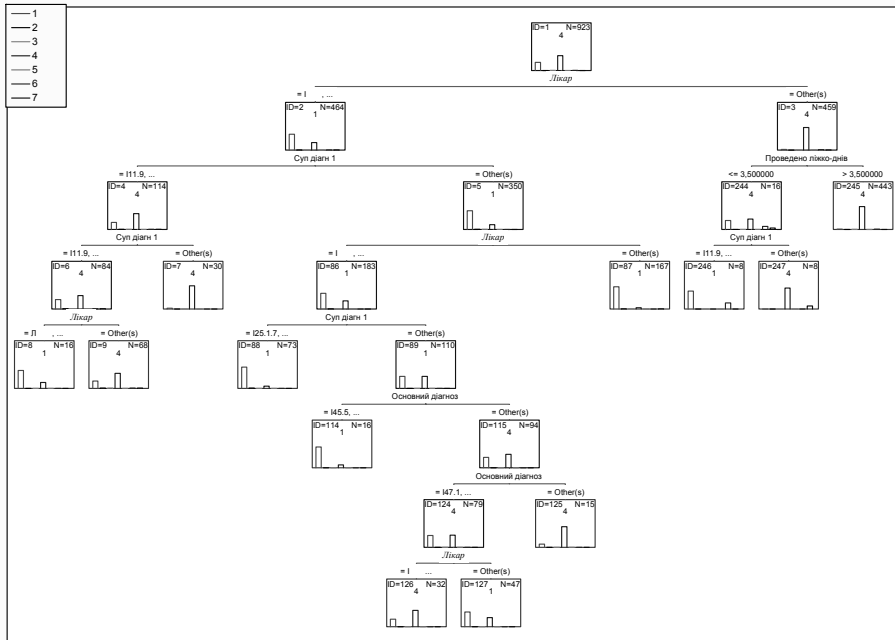


Рис. 2. Повне дерево результатів лікування

Аналіз структури такого дерева (Рис. 2) дає можливість зробити висновки щодо логіки прийняття рішень конкретним лікарем: наприклад, врахування наявності супутніх діагнозів пацієнта для прийняття рішення конкретним лікарем про результат лікування «виписаний з поліпшенням»; направлення пацієнтів на амбулаторне лікування тощо. Залежність між змінними «лікар» — «результат лікування» статистично значима: Chi-square = 588,7; df = 72, p = 0,00001.

За допомогою моделей дерев рішень проаналізовано зв'язок між ускладненнями, основним діагнозом та іншими факторами. Наприклад, застосування алгоритму CHAID (Chi-square automatic interaction detection) дає дерево рішень для моделі залежності ускладнення від основного діагнозу, супутніх діагнозів та персони лікаря (Рис. 3). Особливість алгоритму CHAID полягає в тому, що в якості критерію вибору найвагомішого фактору для сегментації даних використовується статистичний критерій Chi-square. Застосування алгоритму CHAID приводить до такої структури дерева (Рис. 3), коли на кожному кроці (рівні розгалуження дерева) з'являється кінцевий вузол (лист дерева) з ускладненням, що рідко трапляється. Як бачимо, серед переліку основних та супутніх діагнозів найчастішим ускладненням є діагноз I50 — серцева недостатність. Точність прогнозу за алгоритмом CHAID після 10-кратної кросс-валідації — 80%.

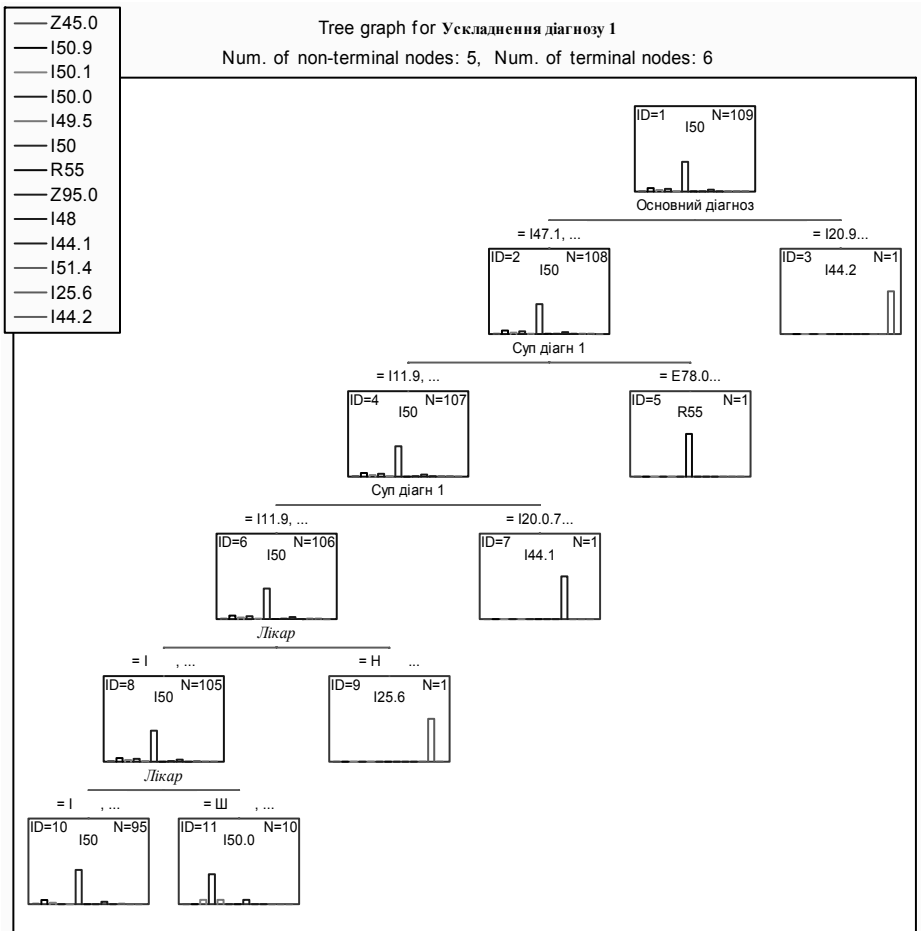


Рис. 3. Дерево рішень для залежності ускладнення від основного діагнозу (алгоритм CHAID)

Алгоритм CART для дерева рішень «ускладнення» після 10-кратної валідації приводить до компактнішої структури моделі порівняно з попередньою моделлю (CHAID), але точність класифікації ускладнень за алгоритмом CART також має значення — 80%. За аналізом моделі за CART визначено, що переважають такі ускладнення: Серцева недостатність (I50), Застійна серцева недостатність (I50.0), Серцева недостатність, не уточнена (I50.9).

Оцінювання результатів лікування пацієнтів діабетичного профілю

З використанням запропонованої інформаційної технології проаналізовано дані відділення ендокринології, нефрології, судинної хірургії, детоксикації Хмельницької обласної клінічної лікарні, які охоплювали 6267 записів за 2012–2018 рр. пацієнтів з захворюваннями на цукровий діабет. Визначено 20 класів діагностики цукрового діабету за кодами МКБ-10. Найчастіше зустрічалися діагнози цукровий діабет 2 типу та типу 1 з ускладненнями: E11.7. Інсулінонезалежний цукровий діабет з множинними ускладненнями) — 44,6% пацієнтів, а також 1 типу E10.7. Інсулінозалежний цукровий діабет з

множинними ускладненнями — 27,8%. Зазначимо, що крім первинно пролікованих, кількість яких складала 0,5%, більшість пацієнтів зверталися повторно, наприклад, два рази на рік — 85,6% пацієнтів.

Розраховано розподіл записів про результати лікування пацієнтів, упорядковані за рангом частоти трапляння (Табл. 3). На першому місці запис згідно з формою Ф-066 Результат лікування — «Амбулаторне лікування». Захворювання на діабет хронічне, з багатьма ускладненнями, зрозуміло, що переважна більшість (97,9%) пацієнтів направляються саме на амбулаторне лікування.

Розбудовано модель дерева рішень (алгоритм CART) для прогнозу результатів лікування в залежності від факторів: діагноз (icd_end), кількість ліжко-днів (history_bed_day), повторність госпіталізації (Рис.4).

Таблиця 3. Результати лікування хворих на діабет, упорядковані за рангом

Ранг	Результат лікування	Код	Кількість	%
1	Амбулаторне лікування	1	6133	97,9
2	Виписаний(а) з одужанням	2	44	0,7
3	Виписаний(а) з поліпшенням	3	38	0,6
4	Денний стац. поліклініки	4	30	0,5
5	Виписаний(а) без змін	6	18	0,3
6	Виписаний(а) з погіршенням	5	4	0,1

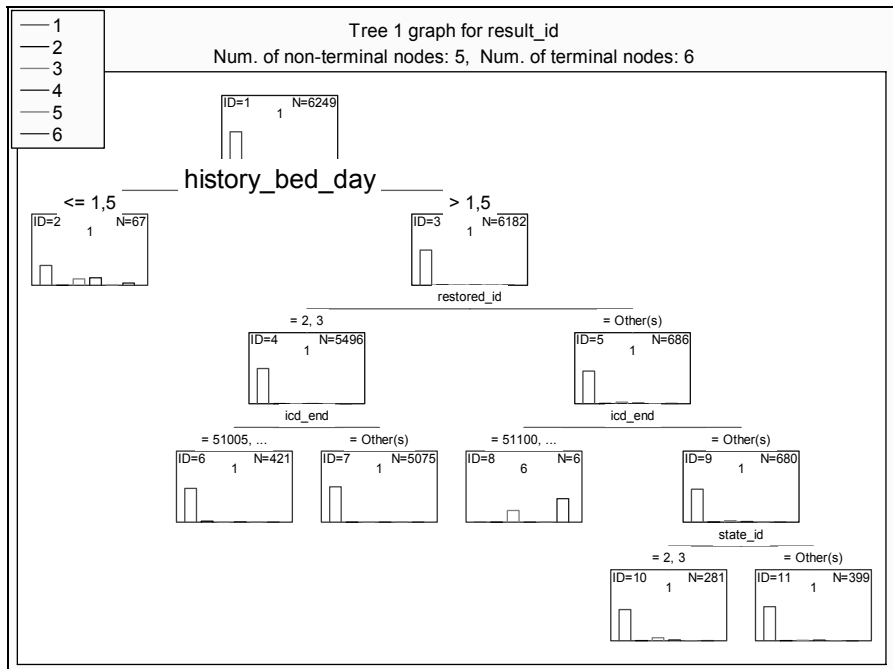


Рис. 4. Дерево рішень (алгоритм CART) залежність результату лікування від факторів: діагноз, кількість ліжко-днів, повторність госпіталізації

Отже, застосування розроблених моделей за методами Data Mining і технології дослідження та оцінювання результатів лікування дають змогу визначити вплив основних характеристик процесу надання медичної допомоги та забезпечують можливість інтелектуальної підтримки прийняття рішень щодо ефективності результатів медичного обслуговування.

ВИСНОВКИ

Розроблена концепція оцінювання якості надання медичної допомоги, яка охоплює принципи, систему показників оцінювання якості надання медичної допомоги, методи та моделі Data Mining визначення результатів медичного втручання за клінічними та статистичними показниками, а також за локальними клінічними протоколами, слугує методичним забезпеченням для оцінювання результатів здійснення медичних послуг на різних рівнях закладів охорони здоров'я.

Побудований метод оцінювання результатів надання медичної допомоги за розробленими моделями із застосуванням методів Data Mining, зокрема методу дерев рішень, дає змогу визначити розподіл стандартизованих результатів лікування в структурах закладів охорони здоров'я, що може бути використано для оцінювання якості надання медичної допомоги.

Поєднання статистичних методів аналізу баз медичних госпітальних даних із застосуванням методів Data Mining (дерево рішень, розрахованого за алгоритмом CART та 10-кратної кросс-валідації) дало змогу виявити перелік супутніх захворювань та ускладнень, характерних для пацієнтів кардіологічного та діабетичного профілей, а також уможливило визначення основних чинників, від яких залежить прийняття рішень лікарями про результат лікування.

ЛІТЕРАТУРА

1. Rademakers J, Delnoij D, de Boer D. Structure, process or outcome: which contributes most to patients' overall assessment of healthcare quality? *BMJ Qual Saf.* 2011 Apr;20(4):326-331. [doi: 10.1136/bmjqs.2010.042358],
2. Ossebaard HC, Van Gemert-Pijnen L. eHealth and quality in health care: implementation time. *Int J Qual Health Care.* 2016 Jun;28(3):415-419. [doi: 10.1093/intqhc/mzw032],
3. Tossaint-Schoenmakers R., Versluis A., Chavannes N., Talboom-Kamp E., Kasteleyn M. The Challenge of Integrating eHealth Into Health Care: Systematic Literature Review of the Donabedian Model of Structure, Process, and Outcome. *J Med Internet Res.* 2021;23(5):e27180 doi: 10.2196/27180
4. Triberti S., Savioni L., Sebrì V., Pravettoni G. eHealth for improving quality of life in breast cancer patients: A systematic review. *Cancer Treatment Reviews*, 2019, Vol. 74, pp. 1-14.
5. Rybarczyk-Szwajkowska A., Marczak M. Quality assessment of health care services in patients and medical staff opinion. January 2011. URL: https://www.researchgate.net/publication/273335138_Quality_assessment_of_health_care_services_in_patients_and_medical_staff_opinion
6. Legido-Quigley H., McKee M., Walshe K., Suñol R., Ellen Nolte E., Klazinga N. How can quality of health care be safeguarded across the European Union? *BMJ.* 2008 Apr 6; 336(7650): 920-923. doi: 10.1136/bmj.39538.584190.47
7. ASA physical status classification system. URL: <https://www.asahq.org/standards-and-guidelines/asa-physical-status-classification-system>

8. Owens W.D., Felts J.A., et al. A physical status classifications: A study of consistency of ratings. *Anesthesiology*. 1978. Vol. 49. P. 239–243.
9. Мельченко М.Г., Елій Л.Б. Можливості оцінки стану пацієнта. *Хірургія дитячого віку*. 2007. №2(55). С. 102–108.
10. Havens J.M., Columbus A.B., Seshadri A.J., et al. Risk stratification tools in emergency general surgery. *Trauma Surg. Acute Care Open*. 2018. № 3. P. 1–8.
13. Royal College of Physicians. NEWS2 and deterioration in COVID-19. URL: <https://www.rcplondon.ac.uk/news/news2-and-deterioration-covid-19>
14. MediCalc. National Early Warning Score 2. URL: <http://www.scymed.com/en/smnxpw/pwfhc210.htm>
15. Frimpong J., Jackson B. et al. Health information technology capacity at federally qualified health centers: a mechanism for improving quality of care. *BMC Health Services Research*. 2013. № 1. P. 13–35. <https://bmchealthservres.biomedcentral.com/articles/10.1186/1472-6963-13-35>
16. Kisekka V., Giboney J. S. The Effectiveness of Health Care Information Technologies: Evaluation of Trust, Security Beliefs, and Privacy as Determinants of Health Care Outcomes. *J Med Internet Res*. 2018; V.20, №4. P. 1–11.
17. Kushniruk A., Hall S., Baylis T., Borycki E., Kannry J. Approaches to demonstrating the effectiveness and impact of usability testing of healthcare information technology. *Studies in health technology and informatics*. 2019. 257. P. 244–249.
18. Minne L., Abu-Hanna A., de Jonge, E. Evaluation of SOFA-based models for predicting mortality in the ICU: A systematic review. *Crit Care*. 2008. 12. P.161.
19. Last M., Tosas O., Cassarino T.G., et al. Evolving classification of intensive care patients from event data. *Artif Intell Med*. 2016; 69:22-32.
20. Houthoofd R., Ruyssinck J., van der Hertten J., et al. Predictive modelling of survival and length of stay in critically ill patients using sequential organ failure scores. *Artif Intell Med*. 2015;63:191-207.
21. Kim S., Kim W., Park R.W. A Comparison of intensive care unit mortality prediction models through the use of Data Mining Techniques. *Health Inform Res*. 2011; 17:232-243.
22. Sameera B. Comparative study of scoring systems in ICU and emergency department in predicting mortality of critically ill. *Int J Res Med Sci*. 2017; 5(4):135.
23. Austin P.C. A comparison of regression trees, logistic regression, generalized additive models, and multivariate adaptive regression splines for predicting AMI mortality. *Stat Med*. 2007, 26:2937-57. <https://doi.org/10.1002/sim.2770>
24. Takahashi O., Cook E.F., Nakamura T., Saito J., Ikawa F., Fukui T. Risk stratification for in-hospital mortality in spontaneous intracerebral haemorrhage: a Classification and Regression Tree analysis. *QJM*. 2006, 99:743-50.
25. Trujillano, J., Badia, M., Serviá, L. et al. Stratification of the severity of critically ill patients with classification trees. *BMC Med Res Methodol*. 2009. 9. 83. P. 1–80.

Отримано 05.01.2023

REFERENCES

1. Rademakers J, Delnoij D, de Boer D. Structure, process or outcome: which contributes most to patients' overall assessment of healthcare quality? *BMJ Qual Saf*. 2011 Apr;20(4):326-331. [doi: 10.1136/bmjqs.2010.042358].
2. Ossebaard HC, Van Gemert-Pijnen L. eHealth and quality in health care: implementation time. *Int J Qual Health Care*. 2016 Jun;28(3):415-419. [doi: 10.1093/intqhc/mzw032]
3. Tossaint-Schoenmakers R., Versluis A., Chavannes N., Talboom-Kamp E., Kasteleyn M. The Challenge of Integrating eHealth Into Health Care: Systematic Literature Review of the Donabedian Model of Structure, Process, and Outcome. *J Med Internet Res*. 2021; 23(5):e27180 doi: 10.2196/27180
4. Triberti S., Savioni L., Sebri V., Pravettoni G. eHealth for improving quality of life in breast cancer patients: A systematic review. *Cancer Treatment Reviews*, 2019, Vol. 74, pp. 1–14.

5. Rybarczyk-Szwajkowska A., Marczak M. Quality assessment of health care services in patients and medical staff opinion. January 2011. URL: https://www.researchgate.net/publication/273335138_Quality_assessment_of_health_care_services_in_patients_and_medical_staff_opinion
6. Legido-Quigley H., McKee M., Walshe K., Suñol R., Ellen Nolte E., Klazinga N. How can quality of health care be safeguarded across the European Union? *BMJ*. 2008 Apr 26; 336(7650): 920–923. doi: 10.1136/bmj.39538.584190.47
7. ASA physical status classification system. URL: <https://www.asahq.org/standards-and-guidelines/asa-physical-status-classification-system>
8. Owens W.D., Felts J.A., et al. A physical status classifications: A study of consistency of ratings. *Anesthesiology*. 1978, Vol. 49, pp. 239–243.
9. Melchenko M.G., Eliy L.B. Possibilities of assessing the patient's condition. *Child. surgery*, 2007, no. 2(55), pp. 102-108 (in Ukrainian).
10. Havens J.M., Columbus A.B., Seshadri A.J., et al. Risk stratification tools in emergency general surgery. *Trauma Surg. Acute Care Open*. 2018, № 3, pp. 1–8.
11. Royal College of Physicians. NEWS2 and deterioration in COVID-19. URL: <https://www.rcplondon.ac.uk/news/news2-and-deterioration-covid-19>
12. MediCalc. National Early Warning Score 2. URL: <http://www.scymed.com/en/smnpw/pwfhc210.htm>
13. Frimpong J., Jackson B. et al. Health information technology capacity at federally qualified health centers: a mechanism for improving quality of care. *BMC Health Services Research*. 2013, № 1, pp. 13–35. URL: <https://bmchealthservres.biomedcentral.com/articles/10.1186/1472-6963-13-35>
14. Kisekka V., Giboney J. S. The Effectiveness of Health Care Information Technologies: Evaluation of Trust, Security Beliefs, and Privacy as Determinants of Health Care Outcomes. *J Med Internet Res*. 2018, V.20, №4, pp. 1–11.
15. Kushniruk A., Hall S., Baylis T., Borycki E., Kannry J. Approaches to demonstrating the effectiveness and impact of usability testing of healthcare information technology. *Studies in health technology and informatics*. 2019, 257, pp. 244–249.
16. Minne L., Abu-Hanna A., de Jonge, E. Evaluation of SOFA-based models for predicting mortality in the ICU: A systematic review. *Crit Care*. 2008, 12, 161.
17. Last M., Tosas O., Cassarino T.G., et al. Evolving classification of intensive care patients from event data. *Artif Intell Med*. 2016; 69:22-32.
18. Houthoofd R., Ruyssinck J., van der Hertten J., et al. Predictive modelling of survival and length of stay in critically ill patients using sequential organ failure scores. *Artif Intell Med*. 2015;63:191-207.
19. Kim S., Kim W., Park R.W. A Comparison of intensive care unit mortality prediction models through the use of Data Mining Techniques. *Health Inform Res*. 2011; 17:232-243.
20. Sameera B. Comparative study of scoring systems in ICU and emergency department in predicting mortality of critically ill. *Int J Res Med Sci*. 2017; 5(4):135.
21. Austin P.C. A comparison of regression trees, logistic regression, generalized additive models, and multivariate adaptive regression splines for predicting AMI mortality. *Stat Med*. 2007, 26:2937-57. <https://doi.org/10.1002/sim.2770>
22. Takahashi O., Cook E.F., Nakamura T., Saito J., Ikawa F., Fukui T. Risk stratification for in-hospital mortality in spontaneous intracerebral haemorrhage: a Classification and Regression Tree analysis. *QJM*. 2006, 99:743-50.
23. Trujillano, J., Badia, M., Serviá, L. et al. Stratification of the severity of critically ill patients with classification trees. *BMC Med Res Methodol*. 2009, 9, 83, 1–80.

Received 05.01.2023

Kovalenko O.S., DSc (Medicine), Professor,
Head of the Medical Information Technologies Department
<https://orcid.org/0000-0001-6635-0124>, e-mail: askov49@gmail.com
Kozak L.M., DSc (Biology), Senior Researcher,
Leading Researcher of the Medical Information Technologies Department
ORCID: 0000-0002-7412-3041, e-mail: lmkozak52@gmail.com
Kryvova O.A.,
Researcher of the Medical Information Technologies Department
<https://orcid.org/0000-0002-4407-5990>, e-mail: ol.kryvova@gmail.com
Bychkov V.V., DSc (Medicine),
Senior Researcher of the Medical Information Technologies Department
<https://orcid.org/0009-0004-8385-9925>, e-mail: lmkozak52@gmail.com
Nenasheva L.V.,
Junior Researcher of the Medical Information Technologies Department
<https://orcid.org/0000-0003-1760-2801>, e-mail: larnen@ukr.net
International Research and Training Center for Information
Technologies and Systems of the National Academy of Sciences
of Ukraine and Ministry of Education and Science of Ukraine,
40, Acad. Glushkov av., Kyiv, 03187, Ukraine

APPLICATION OF CLASSIFICATION MODELS BY DATA MINING AND INFORMATION TECHNOLOGY FOR ANALYZE THE RESULTS OF TREATMENT OF CARDIAC AND DIABETIC PATIENTS

Introduction. *In recent years, the scientific community, especially in the medical field, has been putting a lot of efforts and resources into the development of eHealth technologies and systems. Various methods of intellectual support, which are necessary to ensure high quality of medical care, have been developed. The study of the effectiveness of the application of various methods of diagnosis, treatment of patients and restoration of their health is one of the important components of the assessment of the quality of medical care.*

The purpose of the paper is to analyze the results of providing medical care with the use of developed models based on Data Mining methods to identify factors that affect the results of treatment.

The results. *A method of estimating the medical care results using Data Mining methods has been developed, the feature of which is the combination of filtering algorithms, clustering and classification methods. Models of the medical care result depending on significant factors were built. To test the developed method, a retrospective analysis was carried out using a database of hospital patients of various departments of clinical facilities. The distribution of treatment results evaluation (according to the standardized formulation) of cardiac and diabetic patients was obtained, and concomitant diseases and complications were analyzed. A model for determining the factors influencing the treatment outcome, based on the decision tree method (CART), has been developed. Analysis of the decision tree structure makes it possible to draw conclusions about the decision-making logic by a specific doctor. With the help of decision tree models, the relationship between complications, the main diagnosis and other factors, in particular, concomitant diagnoses, recurrence of hospitalization etc., was analyzed.*

Conclusions. *The combination of statistical methods and the developed method and models based on Data Mining (a decision tree calculated according to the CART algorithm and 10-fold cross-validation) for analysis of medical hospital databases made it possible to identify the frequency characteristics of concomitant diseases and complications typical for cardiac and diabetic patients, and also allowed to determine the main factors that depend on the decision-making by doctors about the outcome of treatment.*

Keywords: *eHealth, Data Mining methods, CART algorithm, information technology, treatment outcomes.*