

DOI: <https://doi.org/10.15276/ict.03.2026.22>

УДК 004.67:616.98

Порівняння байєсівських мереж різної структурної складності для прогнозування летальності при COVID-19

Бралатан Роман Анатолійович¹⁾

ORCID: <https://orcid.org/0009-0002-3439-5234>; bralatan.roman@gmail.com

Жуков Сергій Олександрович¹⁾

ORCID: <https://orcid.org/0009-0002-3443-6139>; sazhukov@gmail.com. Scopus Author ID: 57191729653

¹⁾ Вінницький національний технічний університет, Хмельницьке шосе, 95. Вінниця, 21021, Україна

АНОТАЦІЯ

У роботі розглянуто методологію побудови, налаштування та порівняння трьох імовірнісних моделей – байєсівських мереж різних класів зі зростаючою структурною складністю – для задачі прогнозування летальності пацієнтів із коронавірусною хворобою. Порівняно наївний байєсів класифікатор, деревоподібно доповнений наївний байєсів класифікатор та структурно обмежену байєсівську мережу, структуру якої відновлюють із даних алгоритмом структурного навчання під експертними клінічними обмеженнями. Щоб відокремити вплив структурної складності від відбору ознак, налаштування параметрів і вибору порогів, застосовано порівняння за однакових умов із послідовним спрощенням та ускладненням структури моделі. Особливу увагу приділено розмежуванню двох аспектів якості моделей – розпізнавальної здатності та якості (узгодженості) прогнозованих імовірностей, оскільки для клінічної системи підтримки прийняття рішень визначальною є саме достовірність оцінки ризику, а не лише впорядкування пацієнтів за ступенем ризику. Дослідження проведено на відкритому реєстрі пацієнтів із коронавірусною хворобою за клінічними ознаками з оцінюванням на вибірці, що відтворює природну поширеність летальних випадків. Розглянуто перенесення прогнозованих імовірностей у природну популяцію коригуванням апріорної ймовірності, а також чутливість клінічних рішень до співвідношення вартостей помилок та аналіз чистої клінічної користі. Мета роботи – обґрунтувати доцільність застосування структурно обмеженої байєсівської мережі як інтерпретованого інструмента оцінювання ризику в системах підтримки прийняття клінічних рішень.

Ключові слова: Байєсівські мережі; наївний Байєс; TAN; МІС; PyAgum; COVID-19; прогнозування летальності; калібрування ймовірностей; інтерпретоване машинне навчання; структурне навчання; системи підтримки прийняття рішень.

Для цитування: Бралатан Р. А., Жуков С. О. «Порівняння байєсівських мереж різної структурної складності для прогнозування летальності при COVID-19». *Інформатика. Культура. Техніка.* 2026; Том 3 № 1 (3): 260–276. DOI: <https://doi.org/10.15276/ict.03.2026.22>

Актуальність. Пандемія COVID-19 стала одним із найбільших викликів для систем охорони здоров'я за останнє століття. Вона зумовила безпрецедентне навантаження на медичні заклади, вимагаючи миттєвого прийняття життєво важливих рішень: сортування пацієнтів, визначення доцільності термінової госпіталізації, переведення у відділення інтенсивної терапії та раціонального розподілу обмежених ресурсів. У цих умовах виникла гостра потреба у системах підтримки прийняття клінічних рішень (СППР), що спираються на великі обсяги гетерогенних медичних даних. Класичні методи статистичного аналізу часто виявляються недостатньо гнучкими для виявлення складних нелінійних та багатовимірних взаємозв'язків, тоді як методи машинного навчання забезпечують суттєво вищу розпізнавальну здатність. Проте ключовою вимогою для медичних моделей залишається їхня внутрішня інтерпретованість: лікарю необхідно розуміти не лише фінальний маркер ризику, а й ті чинники, що вплинули на його формування [1].

Широкомасштабне застосування машинного навчання у клінічній рутині стикається з фундаментальною перешкодою – проблемою «чорної скриньки». Складні нелінійні алгоритми (глибокі нейронні мережі, градієнтний бустинг, ансамблі випадкових лісів) демонструють високі метрики дискримінації, проте їх складно впровадити без чіткого пояснення механізму прийняття рішення, бо лікар несе пряму юридичну відповідальність за пацієнта [3].

Важливо також, що для клінічної СППР цінність мають не лише високі значення метрик розпізнавання, а насамперед якість прогнозованих ймовірностей: у природній популяції з рідкісним летальним класом саме добре калібрована ймовірність, а не «сирий» бал ризику, є придатною для триажу. Через це виникає потреба в імовірнісних моделях, які явно

відображають ієрархію умовних (ймовірнісних) залежностей між ознаками, дають робастний прогноз за невизначеності та дозволяють інтегрувати експертний клінічний досвід безпосередньо у процес навчання алгоритму [5].

Аналіз останніх досліджень і публікацій. Питання розробки надійних, прозорих та інтерпретованих моделей для медичної діагностики й прогнозування активно обговорюється у світовій науковій спільноті [2]. Аналіз тенденцій останнього десятиліття, і особливо періоду після 2019 року, свідчить, що Байєсівські підходи залишаються одним із найбільш придатних, математично обґрунтованих інструментів для розв’язання медичних задач [4], оскільки поєднують інтерпретованість вивідних правил, здатність працювати зі стохастичною невизначеністю та можливість пояснити вплив конкретних маркерів на прогноз [5]. Дослідження ключових викликів інтеграції штучного інтелекту у клінічну практику показали, що головною перешкодою для реального використання таких систем є відсутність довіри клініцистів до алгоритмічного висновку та слабка адаптивність «чорних скриньок» за зсувів розподілу даних [3], [4].

Систематичні оглядові роботи показують, що Байєсівські мережі (Bayesian Networks, BN) є високоефективними для складних задач клінічної діагностики та стратифікації ризику, коли від системи вимагається інтерпретована структура умовних залежностей між симптомами [5]. Цей апарат успішно застосовують для складних онкологічних та інфекційних задач, зокрема для моделювання геопросторового та клінічного ризику раку товстої кишки [6] та передопераційної стратифікації ризику [7]. Базовий класифікатор наївного Байєса тривалий час залишався золотим стандартом розпізнавання [8], проте його припущення про повну умовну незалежність ознак є надто грубим спрощенням. Задля усунення цього недоліку запропоновано клас розширених структурних моделей [9], серед яких деревоподібно доповнений наївний Байєс (TAN) став популярним компромісом. Водночас у більшості прикладних медичних робіт імплементація обмежується найпростішими структурами [4], а поетапне зіставлення сучасних методів структурного навчання – від наївного Байєса через TAN до алгоритму МПС [10], [14] з примусовим впровадженням клінічних обмежень – у літературі практично відсутнє. Усунення саме цієї прогалини і зумовило фокус даної статті.

Мета дослідження. Мета роботи полягає у послідовному розробленні, алгоритмічному налаштуванні та порівняльному аналізі трьох імовірнісних моделей Байєсівських мереж різних класів зі зростаючою структурною складністю для задачі прогнозування летальності пацієнтів із COVID-19.

Ці моделі включають:

- 1) наївний Байєс (Naive Bayes) як слабку базову лінію [8];
- 2) покращений TAN із відбором ознак, тюнінгом згладжування та калібруванням порогу [9];
- 3) структурно обмежену байєсівську мережу, побудовану алгоритмом МПС із упровадженням категоричних клінічних експертних обмежень [14].

Окремою метою є показати, що вирішальна перевага складніших моделей проявляється не в розпізнавальній здатності, а в якості (калібруванні) ймовірностей на природній клінічній вибірці – за умови коректного перенесення ймовірностей у цільову популяцію шляхом корекції апіорної ймовірності. Усі експерименти проведено на реальному масштабному реєстрі COVID-19 [12]. З погляду системного аналізу внеском роботи є методика поетапного нарощування структурної складності моделі ($A \rightarrow B \rightarrow C$) як інструмент дослідження компромісу між інтерпретованістю структури, якістю калібрування ймовірностей і вартістю клінічних рішень у системі підтримки прийняття рішень.

Матеріали та методи. Емпіричне дослідження ґрунтується на відкритому реєстрі COVID-19 Уряду Мексики (Kaggle, набір `meirizri/covid19-dataset`), сформованому на основі даних національної системи епідеміологічного нагляду і оприлюдненому у 2021 році [12]. Вихідний файл містить 1 048 575 знеособлених записів пацієнтів із 21 полем. Цільову бінарну змінну летальності (`target`) сформовано з поля `DATE_DIED`: службове значення «9999-99-99» (відсутність зареєстрованої дати смерті) трактовано як відсутність летального

наслідку на момент зрізу реєстру (0), будь-яку конкретну дату – як летальний наслідок (1). Оскільки реєстр є фіксованим зрізом, така мітка відповідає внутрішньореєстровій (госпітальній) летальності; можливе праве цензурування пацієнтів, для яких спостереження на момент зрізу не завершилося, визнано обмеженням дослідження. Окремого поля з точною календарною датою зрізу реєстру відкритий набір не містить – доступні лише індивідуальні дати в полі DATE_DIED, а сам набір оприлюднено у 2021 році; тому за орієнтовний момент зрізу приймаємо час формування його публічної версії (2021 рік). Прогноз формується на клінічний момент первинного оцінювання пацієнта, а всі предиктори (стать, вік, хронічні коморбідності та наявність пневмонії, зафіксована при зверненні) доступні саме на цей момент і не містять інформації про подальший перебіг, що виключає витік інформації про результат.

Кожен запис описано 11 клінічними ознаками: SEX (стать), PNEUMONIA (пневмонія), AGE (вік), DIABETES (діабет), COPD (хронічне обструктивне захворювання легень), INMSUPR (імуносупресія), HIPERTENSION (гіпертензія), CARDIOVASCULAR (серцево-судинні хвороби), OBESITY (ожиріння), RENAL_CHRONIC (хронічна ниркова недостатність), TOBACCO (тютюнопаління) [1], [5].

Формування когорти. Реєстр кодує клінічні прапорці як «1 = так, 2 = ні», а значення 97/98/99 позначають невідоме. Із вихідних 1 048 575 записів вилучено спостереження з невідомим значенням хоча б однієї клінічної ознаки або з відсутнім віком – разом 21 626 записів (менше 2,1 %); після цього залишилося 1 026 949 записів із часткою летальних випадків 7,30 %. Найбільше пропусків припало на пневмонію (16 003), тоді як решта клінічних ознак мали приблизно по 3 000–3 400 пропусків кожна (значення частково перетинаються в одних записах); записів із невідомим віком не виявлено. Перевірка дублікатів показала, що значна частка записів має однакові вектори ознак – очікуваний наслідок огрубленого категоріального кодування невеликого числа полів; такі записи відповідають різним пацієнтам з однаковим клінічним профілем і їх збережено, оскільки вилучення спотворило б емпіричний розподіл класів. Синтетичного оверсемплінгу не застосовували.

Передобробку зведено до логічного мінімуму без агресивного оверсемплінгу [13]. Десять бінарних клінічних прапорців перекодовано у 0/1; ознаку AGE дискретизовано у три клінічні діапазони (0 – до 40 років, 1 – 40–59 років, 2 – 60 і більше) відповідно до фізіологічних норм [8]. Дискретизація є необхідним кроком для Байєсівських мереж класичного типу, оскільки для коректного навчання таблиць умовних імовірностей потрібні дискретні категорії [10], [11].

Для об'єктивного оцінювання побудовано незалежні вибірки. Природну клінічну тестову вибірку (близько 40 000 записів із реальною поширеністю летальності 7,3 %) зарезервовано ще до балансування – саме вона відображає реальну популяцію розгортання і формує головне порівняння. З решти даних (вибірка етапу розробки) сформовано робочу збалансовану вибірку з приблизно 15 000 записів (9 000 тих, хто вижив, і 6 000 летальних випадків, співвідношення 60:40) шляхом недосемплювання мажоритарного класу без синтетичного оверсемплінгу; апіорна частка позитивного класу в ній становить $\pi \approx 0,40$. Збалансовану вибірку поділено стратифіковано 80:20 на навчальну частину та збалансовану контрольну вибірку (holdout 20 %). Усі рішення, що залежать від даних (відбір ознак, тюнінг ваги згладжування, вибір порогів для всіх моделей за єдиним вартісним протоколом, структурне навчання МІС та bootstrap-аналіз), приймалися виключно на навчальній частині; природну тестову вибірку відокремлено до балансування й використано лише для одноразового фінального оцінювання. Природну поширеність летальності, потрібну для корекції апіорної ймовірності, оцінено на окремій калібрувальній підвибірці етапу розробки (записи, що не увійшли до збалансованої вибірки; $\pi' \approx 6,7\%$), а не за мітками тестової вибірки, тож тестова вибірка лишається повністю незалежною. Вагові коефіцієнти функції клінічної вартості зафіксовано у консервативному співвідношенні cFN : cFP = 3 : 1 (докладне клінічне обґрунтування наведено під формулою (4)). Відтворюваність забезпечено

фіксованим зерном генератора випадкових чисел (`random_state = 42`). Повний перелік обов'язкових та заборонених дуг наведено у відкритому репозиторії коду.

Теоретичною основою всіх трьох класифікаторів слугує факторизація спільного розподілу ймовірностей усіх змінних у добуток локальних умовних розподілів кожного вузла за його безпосередніми батьківськими вузлами.

Для набору клінічних ознак $X_1, X_2, \dots, X_m$ та бінарної цільової змінної Y (летальність) спільний розподіл подається у вигляді (1):

$$P(X_1, \dots, X_m, Y) = P(Y | \text{Pa}(Y)) \cdot \prod_{i=1}^m P(X_i | \text{Pa}(X_i)), \quad (1)$$

де $P(X_1, \dots, X_m, Y)$ – спільна ймовірність повного вектора ознак та цільового класу; $P(Y | \text{Pa}(Y))$ – умовний розподіл цільової змінної летальності за її батьками; $P(X_i | \text{Pa}(X_i))$ – локальний умовний розподіл ознаки X_i за множиною її батьківських вузлів $\text{Pa}(X_i)$ у графі мережі G . Саме конфігурація множин $\text{Pa}(X_i)$, тобто топологія спрямованого графа G , є ключовою відмінністю між трьома порівнюваними моделями (наївний Байєс, покращений TAN та МПІС) і виступає головним об'єктом структурної оптимізації.

Наївний Байєс є окремим (найпростішим) випадком (1), у якому кожна ознака має єдиного батька – цільову змінну, а TAN – розширенням, де кожна ознака отримує ще одного батька-ознаку (2):

$$\text{Naive Bayes: } P(Y, X_1, \dots, X_m) = P(Y) \cdot \prod_{i=1}^m P(X_i | Y); \quad (2)$$

$$\text{TAN: } P(Y, X_1, \dots, X_m) = P(Y) \cdot \prod_{i=1}^m P(X_i | Y, X_{\pi(i)})$$

де $X_{\pi(i)}$ – додатковий батько-ознака вузла X_i , дозволений у TAN (у наївному Байєсі така множина порожня). Саме ця відмінність робить наївний Байєс (етап А) та TAN (етап В) різними класами моделей: у наївного Байєса корельовані коморбідності враховуються як незалежні свідчення й «подвоюються», тоді як TAN частково знімає це припущення через попарні зв'язки.

Ступінь статистичних асоціацій на етапі відбору фіксувався за допомогою нормалізованої міри зв'язку – коефіцієнта Cramér's V , який нормалізує статистику χ^2 у діапазон від 0 (відсутність зв'язку) до 1 (функціональна залежність) і застосовується для дискретних категоріальних ознак (3):

$$V = \sqrt{\frac{\chi^2}{N(\min(r, c) - 1)}}, \quad (3)$$

де χ^2 – статистика хі-квадрат Пірсона, обчислена за таблицею спряженості двох ознак; N – загальна кількість спостережень; r та c – кількість рядків і стовпців таблиці спряженості (кількість градацій кожної ознаки); а $(\min(r, c) - 1)$ – коригувальний множник на число ступенів свободи, що забезпечує коректне масштабування оцінки між парами ознак різної розмірності.

Оптимальний поріг відсікання τ для етапів В і С обирався мінімізацією асиметричної функції клінічної вартості (4):

$$L(\tau) = c_{FN} FN(\tau) + c_{FP} FP(\tau), \quad (4)$$

де $L(\tau)$ – сумарна клінічна вартість помилок при порозі τ ; $FN(\tau)$ та $FP(\tau)$ – кількість хибнонегативних та хибнопозитивних прогнозів як функції порога; c_{FN} та c_{FP} – штрафи за відповідний тип помилки. Асиметричний штраф (3:1) свідомо зміщує оптимальний поріг τ у бік вищої чутливості, мінімізуючи ризик пропуску потенційно летальних пацієнтів. Співвідношення $c_{FN} : c_{FP} = 3 : 1$ обґрунтоване клінічно: пропуск летального результату позбавляє пацієнта своєчасної ескалації допомоги (госпіталізації, переведення у відділення інтенсивної терапії) і коштує значно дорожче за хибну тривогу, яка спричиняє лише додаткове спостереження [18]; значення 3:1 узятو як консервативну нижню оцінку, а його

вплив досліджено окремим аналізом чутливості. Штрафи є відносними (безрозмірними) вагами, а не грошовими величинами. Пороги для всіх моделей, включно з наївним Байєсом, визначаються за цією функцією вартості виключно на навчальних даних (єдиний вартісний протокол).

Для зручності інтерпретації в таблицях наведено середню (нормовану на число пацієнтів) клінічну вартість.

$$I(X_i; X_j | Z) = \sum p(x_i, x_j, z) \log \frac{p(x_i, x_j | z)}{p(x_i | z) p(x_j | z)}, \quad (5)$$

де $p(x_i, x_j | z)$ – умовна взаємна інформація між ознаками X_i та X_j за умови відомого Z ; $p(x_i, x_j, z)$ – спільна ймовірність відповідних градацій; $p(x_i, x_j | z)$ – умовна спільна ймовірність; $p(x_i | z)$, $p(x_j | z)$ – умовні маргінальні ймовірності; сумування ведеться по всіх комбінаціях дискретних значень. Якщо $I(X_i; X_j | Z) \rightarrow 0$, ознаки вважаються умовно незалежними, і відповідне ребро в графі видаляється. Особливістю алгоритму МПС є використання не класичної, а скінченнорозмірно скоригованої взаємної інформації: з емпіричної оцінки $I(X_i; X_j | Z)$ віднімається член складності, що штрафує за обмежений обсяг вибірки; саме ця регуляризація відрізняє МПС від класичних тестів умовної незалежності (наприклад, РС/ІС) і визначає, чи зберігається ребро у графі.

Модель А: наївний Байєс. На першому етапі побудовано класифікатор наївного Байєса на всіх 11 ознаках зі згладжуванням Лапласа [8]. Ця модель трактує кожен коморбідність як умовно незалежну за цільовою змінною, тому корельовані фактори ризику (пневмонія, вік, діабет, гіпертензія) враховуються багаторазово, що продукує надмірно впевнені ймовірності. Саме тому наївний Байєс слугує коректною слабкою базовою лінією: на природній вибірці він демонструє найгірше калібрування (найбільші значення оцінки Брайєра та лог-втрат), попри прийнятну розпізнавальну здатність. Для коректного зіставлення робочий поріг наївного Байєса, як і решти моделей, визначено за єдиним вартісним протоколом на навчальних даних ($\tau = 0,234$), а не фіксованим на рівні 0,5; номінальний поріг 0,5 подано лише як класичне посилання.

Модель В: покращений TAN із відбором ознак і тюнінгом гіперпараметрів. Другий етап підвищує клас моделі до TAN і водночас скорочує простір вхідних ознак на основі їх статистичної значущості [2], [10]. Попередній розвідувальний аналіз (Рис. 1) показує, що летальність різко зростає з градаціями основних клінічних ознак – віку, пневмонії, хронічної ниркової недостатності та діабету, – причому ці ознаки взаємно корельовані, що якраз і робить припущення незалежності наївного Байєса шкідливим для калібрування.

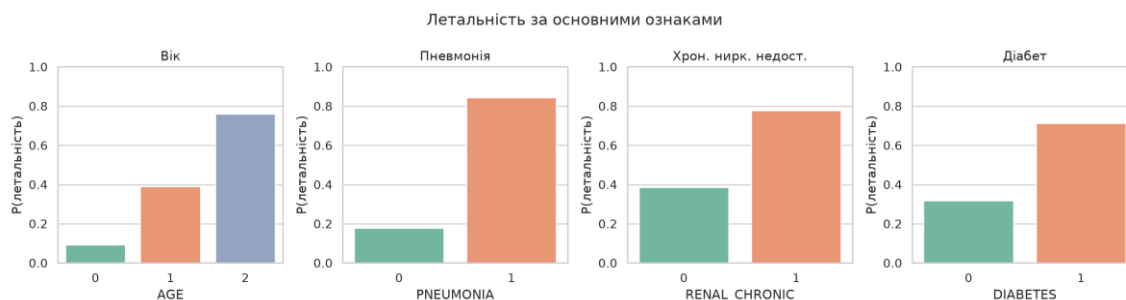


Рис. 1. Летальність за основними клінічними ознаками (збалансована вибірка)

Джерело: розроблено авторами

Силу зв'язку кожної ознаки з цільовою змінною оцінено коефіцієнтом Cramér's V на навчальній частині даних (Рис. 2). Найбільший інформаційний зв'язок із летальністю утворюють пневмонія ($V = 0,64$) та вік ($V = 0,55$), далі – гіпертензія й діабет (по 0,33), хронічна ниркова недостатність (0,15) та стать (0,14). Застосовано відтворюваний поріг

відбору на рівні медіани коефіцієнта, обчисленої виключно на навчальній частині даних ($V \geq 0,136$): для етапів В і С відібрано шість ознак – PNEUMONIA (0,64), AGE (0,55), HIPERTENSION (0,33), DIABETES (0,33), RENAL_CHRONIC (0,15) та SEX (0,14); решту ознак з нижчим маргінальним значенням Cramér's V не включено до подальшого моделювання (COPD 0,11; OBESITY 0,10; CARDIOVASCULAR 0,10; INMSUPR 0,08; TOBACCO 0,01). Слід зауважити, що коефіцієнт Cramér's V оцінює лише маргінальний (попарний) зв'язок ознаки з цільовою змінною і не враховує взаємну надлишковість ознак; тому такий відбір може відкинути ознаку, слабку маргінально, але потенційно інформативну умовно (у поєднанні з іншими ознаками), що є визнаним обмеженням цього етапу. Отримані пріоритети узгоджуються з клінічною практикою, підтверджуючи вік та наявність пневмонії як важливі предиктори прогнозу в межах цього реєстру [6], [7].

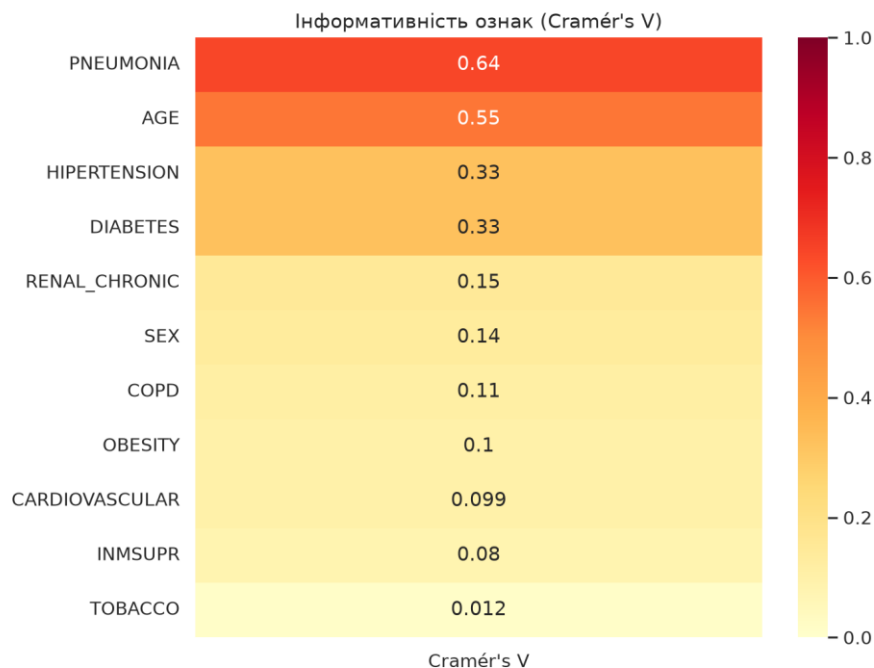


Рис. 2. Інформативність клінічних ознак за коефіцієнтом Cramér's V

Джерело: розроблено авторами

Вагу згладжування Лапласа (priorWeight) для TAN тюнінговано на сітці значень {0,5; 1,0; 2,0; 3,0; 5,0} за 5-кратною перехресною перевіркою за критерієм ROC AUC (Рис. 3); оптимальним виявилось значення 1,0. Далі за формулою (4) на тренувальних даних встановлено вартісно-оптимальний поріг відсікання $\tau = 0,177$. Поріг τ оптимізували на навчальній вибірці за заданою функцією клінічної вартості; незалежна тестова вибірка не використовувалася під час його визначення. Один додатковий батько на ознаку вже послаблює припущення незалежності наївного Байєса, помітно поліпшуючи калібрування та чутливість порівняно з етапом А.

Модель С: обмежена мережа МІС (PyAgrum). Третій етап використовує найвиразніший клас моделі – повну Байєсівську мережу, навчену алгоритмом МІС (Multivariate Information-based Inductive Causation) на відібраному наборі ознак у бібліотеці PyAgrum [14], [15]. Слово “Causation” у назві алгоритму є історичним і не свідчить про причинну валідність відновлених дуг. На відміну від фіксованої «зірки» наївного Байєса чи фіксованого дерева TAN, тут структура графа відновлюється з даних під експертними клінічними обмеженнями.

По-перше, заблоковано будь-які спрямовані дуги від target до вхідних фізіологічних ознак (унеможливлення біологічно абсурдних петель).

По-друге, введено обов'язкові (експертно задані) дуги PNEUMONIA → target та AGE → target [5], [7].

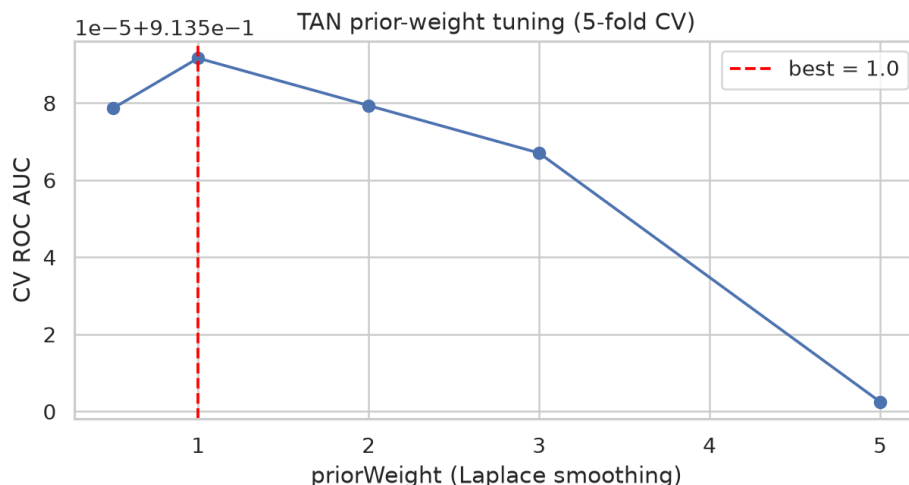


Рис. 3. Тюнінг ваги згладжування Лапласа TAN за 5-кратною перехресною перевіркою
Джерело: розроблено авторами

Формально множина клінічних обмежень C задається як об'єднання двох підмножин спрямованих дуг (6):

$$\begin{aligned} C &= C_{\text{mand}} \cup C_{\text{forb}}, \\ C_{\text{mand}} &= \{ \text{PNEUMONIA} \rightarrow Y, \text{AGE} \rightarrow Y \}, \\ C_{\text{forb}} &= \{ Y \rightarrow X_i, \forall i \} \end{aligned} \quad (6)$$

де C – повний набір структурних обмежень мережі; C_{mand} – множина обов'язкових (експертно заданих) дуг, що фіксують у моделі припущення про прямі спрямовані залежності від ознак пневмонії та віку до цільової змінної Y ; C_{forb} – множина заборонених дуг, що виключає будь-які зворотні зв'язки від цільової змінної Y до вхідних ознак X_i .

Таке накладання підмножин під час роботи МПС узгоджує кінцеву структуру з клінічними обмеженнями (без твердження про причинний ефект). Для стійкості оцінок таблиць умовних імовірностей застосовано згладжування Лапласа з параметром 1,0.

Навчена мережа містить 12 дуг, а прямими батьківськими вузлами цільової змінної летальності виступають одразу п'ять критичних індикаторів: пневмонія, вік, стать, діабет та хронічна ниркова недостатність. Візуалізовано також опосередковані каскадні зв'язки, наприклад $\text{AGE} \rightarrow \text{HYPERTENSION}$, $\text{DIABETES} \rightarrow \text{HYPERTENSION}$ та $\text{DIABETES} \rightarrow \text{RENAL_CHRONIC}$, що відповідають відомим патофізіологічним ланцюгам. Контрольне навчання повністю необмеженої МПС-мережі (без експертних обов'язкових дуг) відновило ту саму структуру з 12 дуг і той самий перелік батьків цільової змінної, тому експертні обмеження є узгодженими з даними, а не визначальними для результату (детальний аналіз стабільності – у розділі результатів). Апостеріорну ймовірність $P(\text{target} = 1 \mid \text{свідчення})$ обчислювали точним висновком (Lazy Propagation) із кешуванням за унікальними патернами свідчень, що робить оцінювання десятків тисяч рядків природної вибірки обчислювально ефективним.

Оцінювання якості, метрик та топологічної стабільності. Для порівняльного оцінювання трьох моделей застосовано комплексний набір метрик, обчислених на основі елементів матриці плутанини: TP (правильно виявлені летальні випадки), TN (правильно виявлені ті, хто вижив), FP (хибна тривога) та FN (пропущена загроза). Окрім F1-міри (гармонійне середнє точності й повноти), чутливості ($\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN})$) та специфічності ($\text{Specificity} = \text{TN} / (\text{TN} + \text{FP})$), особливу увагу приділено метрикам якості ймовірностей, оскільки саме вони розрізняють три класи моделей.

Оцінку Брайера (Brier Score) та лог-втрати (log-loss) обчислюють за формулами (7)-(8):

$$BS = \frac{1}{N} \sum (p_i - y_i)^2, \quad (7)$$

$$LL = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)], \quad (8)$$

де p_i – прогнозована ймовірність летальності для i -го пацієнта; $y_i \in \{0, 1\}$ – фактичний клас; N – кількість спостережень. Обидві метрики є чутливими до калібрування: що менші BS та LL , то ближчі прогнозовані ймовірності до фактичних частот. Додатково використано PR-AUC (площу під кривою «точність–повнота»), яка є коректним показником для рідкісного позитивного класу у природній популяції.

Для забезпечення довіри до топології згенерованої мережі проведено bootstrap-аналіз: мережу повторно навчали на 500 перевибірках із заміщенням і фіксували частоту появи кожної дуги (9):

$$S_{i \rightarrow j} = \frac{1}{B} \sum_{b=1}^B 1 \{(i \rightarrow j) \in G_b\}, \quad (9)$$

де $S_{i \rightarrow j}$ – частка появи спрямованої дуги $i \rightarrow j$; B – загальна кількість bootstrap-ітерацій ($B = 500$); G_b – структура, навчена на b -й перевибірці. Частоту появи слід інтерпретувати лише для дуг, які алгоритм міг як додати, так і не додати; обов'язкові експертні дуги PNEUMONIA $\rightarrow$ target та AGE $\rightarrow$ target фіксуються за конструкцією, тому їхня частота 100 % не є незалежною оцінкою стабільності.

Щоб урахувати можливу Марківську еквівалентність, стабільність структури оцінено окремо за двома складовими: частотою суміжності (наявністю ребра незалежно від напрямку) та умовною частотою орієнтації (часткою того самого напрямку серед перевибірок, у яких ребро присутнє). Детальні результати цього розділення, а також порівняння структурно обмеженої та повністю необмеженої мереж, наведено у розділі результатів (Рис. 4).

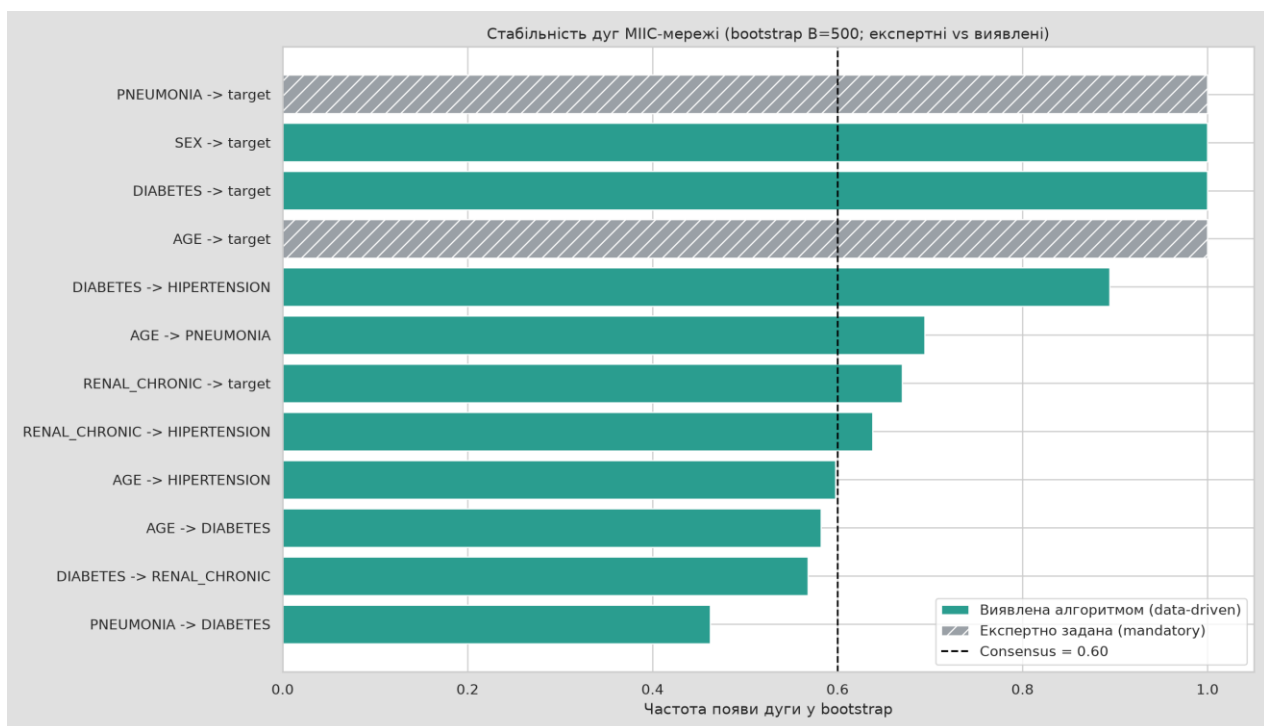


Рис. 4. Стабільність дуг МІС-мережі за bootstrap-аналізом ($B = 500$)

Джерело: розроблено авторами

Результати та їх обговорення. Порівняльний аналіз трьох послідовно ускладнюваних моделей свідчить про поліпшення якості ймовірнісного прогнозу за переходом $A \rightarrow B \rightarrow C$, хоча ця перевага не є універсальною за всіма метриками. Ключове спостереження полягає в тому, що розпізнавальна здатність ($ROC\ AUC \approx 0,91$) майже однакова для всіх трьох моделей, оскільки кілька домінуючих предикторів (пневмонія, вік) несуть основний ранжувальний сигнал. Натомість справжні відмінності проявляються у якості ймовірностей – саме в тому, що потрібно клінічному інструменту підтримки рішень. Підсумкові метрики на

природній клінічній вибірці (поширеність летальності 7,3 %) наведено в Таблиці 1. Стовпці «сир./кор.» подають метрики калібрування до і після корекції апіорної ймовірності до незалежно оціненої поширеності; пороги всіх моделей визначено за єдиним вартісним протоколом на навчальних даних; середню клінічну вартість подано у відносних (безрозмірних) одиницях штрафу на пацієнта.

Таблиця 1. Метрики моделей на природній клінічній вибірці (поширеність 7,3 %; N ≈ 40 000, летальних випадків ≈ 2 900)

Модель	ROC AUC	PR-AUC	Sens	Brier (сир./кор.)	Log-loss (сир./кор.)	ECE (сир./кор.)	Серед. вартість	Поріг τ
А: наївний Байєс	0.9092	0.4155	0.9291	0.131 / 0.065	0.466 / 0.224	0.186 / 0.046	0.266	0.2335
В: покращений TAN	0.9139	0.4396	0.9209	0.106 / 0.050	0.342 / 0.168	0.164 / 0.010	0.243	0.1771
С: обмежений МПС	0.9149	0.4506	0.8987	0.097 / 0.048	0.308 / 0.163	0.152 / 0.004	0.217	0.2234

Джерело: розроблено авторами

Як видно з Таблиці 1, показники калібрування монотонно поліпшуються за переходом $A \rightarrow B \rightarrow C$: оцінка Брайера знижується з 0,131 до 0,106 і далі до 0,097, лог-втрати – з 0,466 до 0,342 і до 0,308, а очікувана калібрувальна похибка (ECE, 10 рівношироких бінів) – з 0,186 до 0,164 і до 0,152; одночасно зростає PR-AUC (0,416 $\rightarrow$ 0,440 $\rightarrow$ 0,451). Метрики якості ймовірностей (Brier, log-loss, ECE, PR-AUC) не залежать від порога і тому зіставні безпосередньо. Пороги всіх трьох моделей, включно з наївним Байєсом, визначено за єдиним вартісним протоколом (4) виключно на навчальних даних, тому чутливість і середня клінічна вартість відображають спільний ефект структури й порога, а не лише структурну складність. Найкраще калібрування має модель С, найгірше – наївний Байєс (А). Абсолютні значення Brier і log-loss слід інтерпретувати як порівняльні між моделями, оскільки за рідкісного позитивного класу (7,3 %) вони значною мірою визначаються переважним негативним класом.

Сирі ймовірності, отримані на збалансованій навчальній вибірці (апіорна частка позитивного класу $\pi \approx 0,40$), систематично переоцінюють ризик на природній превалентності, тому їх перенесено у цільову популяцію корекцією апіорної ймовірності за процедурою Саренса–Елкана (10) [17]:

$$p' = \frac{(\pi'/\pi)p}{(\pi'/\pi)p + ((1-\pi')/(1-\pi))(1-p)}, \quad (10)$$

де p – ймовірність, отримана на збалансованій вибірці, π – її апіорна частка позитивного класу, а π' – цільова поширеність.

Принципово, що природну поширеність оцінено не за мітками тестової вибірки, а на незалежній калібрувальній підвибірці етапу розробки ($\pi' \approx 6,7\%$), тож тестова вибірка залишається повністю незалежною. Корекція є строго монотонним перетворенням, тому ROC-AUC, PR-AUC і рішення за відповідно перерахованим порогом не змінюються, а змінюються лише метрики калібрування, наведені у стовпцях «кор.» Таблиці 1: ECE знижується до 0,046 (А), 0,010 (В) і 0,004 (С), оцінка Брайера – до 0,065 / 0,050 / 0,048, лог-втрати – до 0,224 / 0,168 / 0,163. За 95 % bootstrap-довірчими інтервалами ($B = 1000$) скоригована ECE моделі С становить [0,002; 0,006], оцінка Брайера – [0,047; 0,050], лог-втрати – [0,158; 0,168], причому інтервали моделі С не перекриваються з інтервалами наївного Байєса за жодною з метрик калібрування.

Аналіз чутливості до значення природної поширеності підтверджує стійкість висновку: скоригована ECE моделі С не перевищує 0,05 у діапазоні поширеності від 3 % до 15 % (0,035 за 3 %; 0,003 за 7,3 %; 0,046 за 15 %). Попарне порівняння на тих самих рядках (paired bootstrap) підтверджує статистичну значущість переваги моделі С над моделлю В за якістю

ймовірностей, тоді як різниця за ROC AUC не є статистично значущою, що узгоджується з тезою про близьку розпізнавальну здатність усіх трьох моделей (Таблиця 2).

Таблиця 2. Метрики моделей на збалансованій контрольній вибірці (holdout 20 %, $N \approx 3\,000$); пороги всіх моделей визначено за єдиним вартісним протоколом на навчальних даних

Модель	ROC AUC	F1	Sensitivity	Specificity	Brier	Поріг τ
А: наївний Байєс	0.9045	0.7959	0.9325	0.7261	0.1283	0.2335
В: покращений TAN	0.9111	0.8078	0.9267	0.7550	0.1151	0.1771
С: обмежений МПС	0.9117	0.8163	0.9075	0.7894	0.1142	0.2234

Джерело: розроблено авторами

Рис. 5 наочно демонструє цю закономірність на природній вибірці за шістьма показниками: три верхні графіки показують, що дискримінація практично не змінюється, тоді як PR-AUC зростає, а три нижні підтверджують, що модель С дає найкраще калібрування. За єдиного вартісного протоколу (усі моделі використовують вартісно-оптимальний поріг, визначений на навчальних даних) середня клінічна вартість на пацієнта за співвідношення штрафів 3:1 становить 0,266 для А, 0,243 для В та 0,217 для С, тобто найменшою є вартість структурно обмеженої моделі С. Це важлива відмінність від наївного зіставлення за фіксованого порога 0,5: коли наївний Байєс оцінюють за єдиним протоколом, а не за нейтральним порогом, його позірна «дешевизна» зникає.

Водночас, як показано далі в аналізі чутливості (Таблиця 4), домінування за вартістю не є універсальним і залежить від заданого співвідношення cFN : cFP, тому цей показник слід інтерпретувати разом із PR-AUC та метриками калібрування.

Узгодженість прогнозованих імовірностей із фактичними частотами ілюструють калібрувальні криві (Рис. 6). Крива відкладає прогнозовану ймовірність проти спостережуваної частоти; діагональ відповідає ідеальному калібруванню.

На лівій панелі Рис. 6 показано, сірі ймовірності моделей, навчених на збалансованій вибірці: усі три криві, зокрема й модель С, систематично лежать нижче діагоналі на всьому діапазоні прогнозів понад 0,2, тобто на природній превалентності усі моделі переоцінюють ризик; це узгоджується з великими значеннями очікуваної калібрувальної похибки (ECE = 0,186 для А, 0,164 для В та 0,152 для С) і зумовлено зсувом апіорної поширеності через балансування навчальної вибірки.

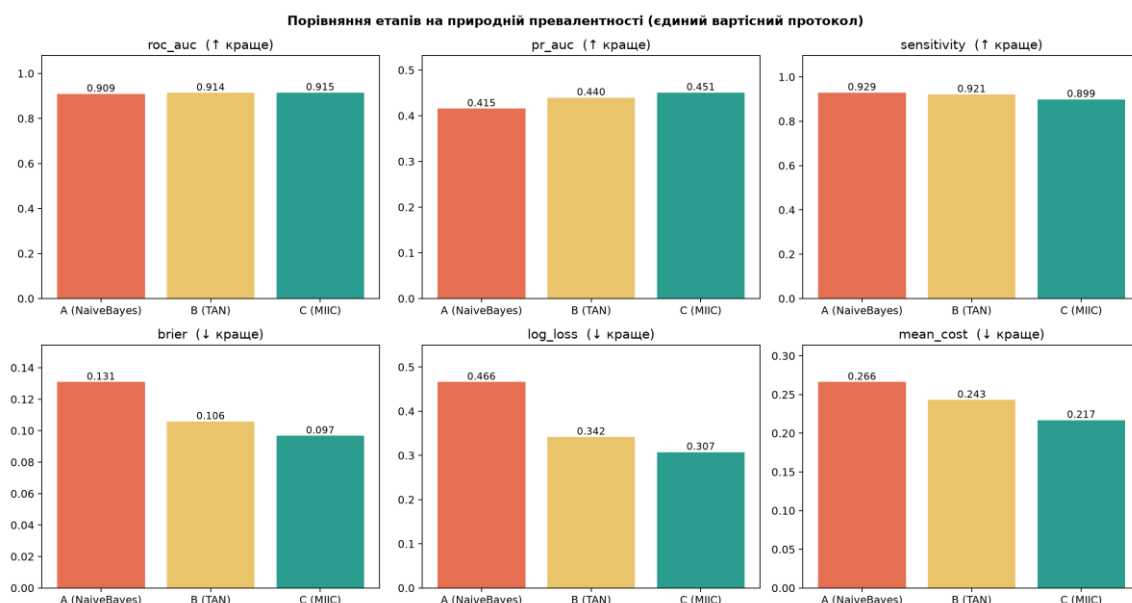


Рис. 5. Порівняння етапів А → В → С на природній превалентності за ключовими метриками

Джерело: розроблено авторами

На правій панелі наведено ті самі моделі після корекції апіорної ймовірності до незалежно оціненої природної поширеності ($\approx 6,7\%$): криві наближаються до діагоналі, а модель С відтворює її найточніше (ECE $\approx 0,004$; оцінка Брайера 0,048; лог-втрати 0,163). Кожну точку супроводжено 95 % довірчим інтервалом Вілсона для спостережуваної частоти, а нижні панелі показують кількість випадків у кожному біні, тому криву слід інтерпретувати разом із наповненням бінів, а не як самостійний доказ переваги.

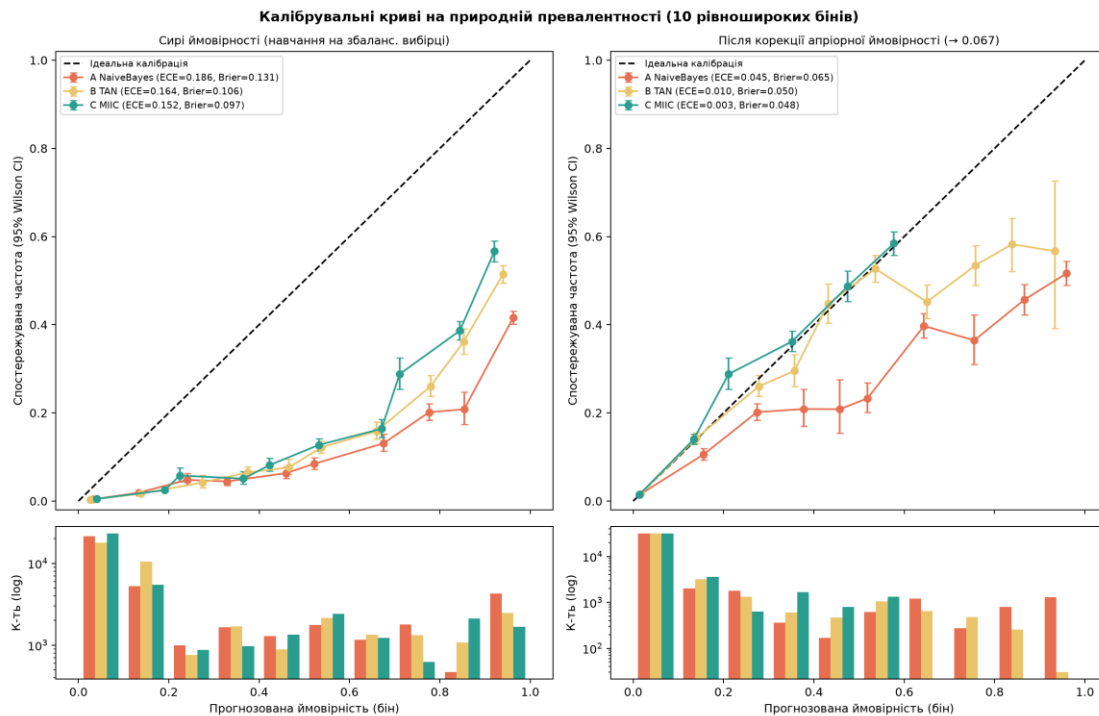


Рис. 6. Калібрувальні криві моделей А, В, С на природній превалентності: сирі ймовірності (ліворуч) і після корекції апіорної ймовірності (праворуч)

Джерело: розроблено авторами

ROC-криві трьох моделей майже збігаються (близький ранжувальний потенціал даних), проте криві «точність-повнота» (коректний ракурс для рідкісного позитивного класу) показують, що модель С має найвище інтегральне значення PR-AUC, хоча окремі ділянки кривих моделей перетинаються, узгоджуючись із її вищим PR-AUC (Рис. 7).

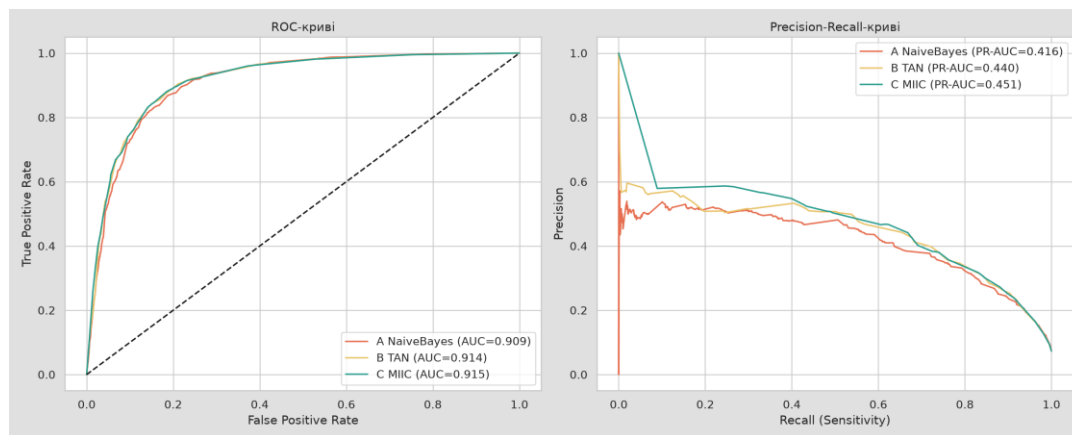


Рис. 7. ROC- та Precision–Recall-криві моделей на природній превалентності

Джерело: розроблено авторами

Окрему цінність має здатність моделі С до імовірнісного «що-якщо» висновку (Рис. 8). Апостеріорна ймовірність летальності зростає від лише 2,9 % для молодого пацієнта без ускладнень до 40,0 % за відсутності свідчень у збалансованій робочій підмножині (це

значення відображає розподіл класів цієї підмножини і не є базовим ризиком природної популяції), 83,8 % за наявності самої пневмонії та 90,2 % для літнього пацієнта з пневмонією і хронічною нирковою недостатністю. Ці значення є модельними ймовірностями на збалансованій підмножині й ілюструють відносну зміну ризику за наявних свідчень, а не абсолютний популяційний ризик.

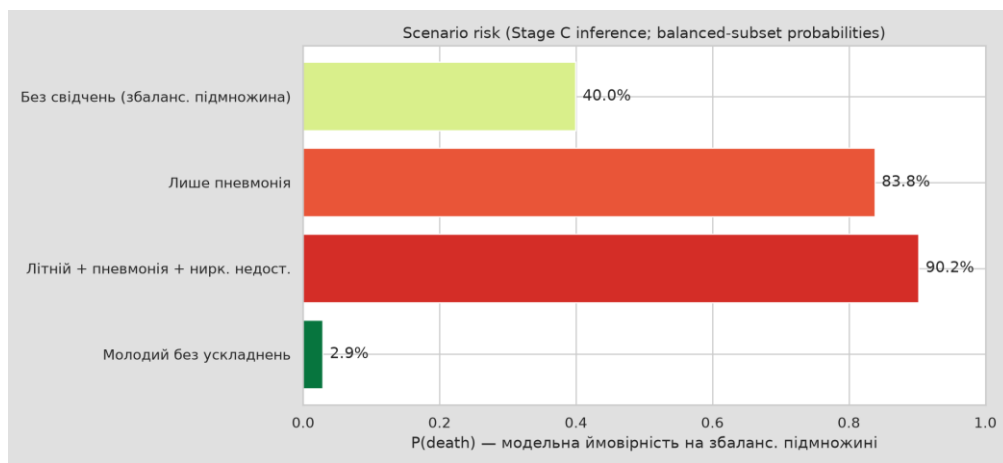


Рис. 8. Байєсівський висновок: імовірність смерті для типових клінічних сценаріїв (модель С)

Джерело: розроблено авторами

Абляційний аналіз (розділення впливу структури, ознак і обмежень). Щоб відокремити вплив структурної складності від відбору ознак, налаштування параметрів, вибору порога та експертних обмежень, проведено абляційне порівняння за однакових умов на природній тестовій вибірці (Таблиця 3).

Порівнювали п'ять конфігурацій: наївний Байєс на 11 і на 6 ознаках, TAN на 6 ознаках, а також МПС на тих самих 6 ознаках без обмежень і з експертними обмеженнями. Пороги для всіх п'яти моделей визначено за єдиним вартісним протоколом (4) виключно на навчальних даних. Скорочення набору ознак з 11 до 6 (наївний Байєс → наївний Байєс на 6 ознаках) майже не змінює якості ймовірностей (оцінка Брайєра 0,1309 → 0,1298; ECE 0,186 → 0,188), тоді як зміна класу структури (наївний Байєс → TAN на тих самих 6 ознаках) дає основний приріст (оцінка Брайєра 0,1298 → 0,1058; лог-втрати 0,453 → 0,342). Це доводить, що перевага складніших моделей зумовлена саме послабленням припущення умовної незалежності, а не відбором ознак. Ключовим є й те, що МПС-мережа з експертними обмеженнями та повністю необмежена МПС-мережа збіглися і за структурою (той самий набір із 12 дуг та той самий перелік батьків цільової змінної), і за всіма метриками, тобто експертні обмеження узгоджуються з даними, а не визначають результат.

Таблиця 3. Абляційне порівняння за однакових умов (природна вибірка, єдиний вартісний протокол)

Конфігурація	Ознак	ROC AUC	PR-AUC	Brier	Log-loss	ECE	ECE (кор.)	Sens	Серед. вартість	τ
Наївний Байєс	11	0.9092	0.4155	0.1309	0.4660	0.1858	0.0455	0.9291	0.266	0.2335
Наївний Байєс	6	0.9105	0.4284	0.1298	0.4526	0.1884	0.0429	0.9267	0.262	0.3263
TAN	6	0.9139	0.4396	0.1058	0.3417	0.1644	0.0103	0.9209	0.243	0.1771
МПС без обмежень	6	0.9149	0.4506	0.0967	0.3075	0.1519	0.0035	0.8987	0.217	0.2234
МПС з обмеженнями	6	0.9149	0.4506	0.0967	0.3075	0.1519	0.0035	0.8987	0.217	0.2234

Джерело: розроблено авторами

Чутливість до співвідношення вартостей і аналіз кривих рішень. Оскільки твердження про клінічну перевагу залежить від заданого співвідношення вартостей помилок, виконано аналіз чутливості для п'яти клінічно вмотивованих співвідношень cFN : cFP (Таблиця 4, Рис. 9); для кожного співвідношення порогови всіх моделей переоптимізовано за єдиним протоколом на навчальних даних. За єдиного вартісного протоколу й обраного співвідношення 3:1 найменшу середню клінічну вартість забезпечує модель С (0,217 проти 0,266 для А та 0,243 для В); попереднє враження про «найдешевшу» базову модель було артефактом порівняння наївного Байєса за фіксованого порога 0,5 з вартісно-оптимізованими В і С. Водночас жодна модель не домінує за усіма співвідношеннями: за 1:1 найдешевшою є проста модель А, за 2:1 та 10:1 – модель В, за 3:1 – модель С, що підтверджує потребу узгодження співвідношення вартостей із клінічними експертами. Додатково проведено аналіз кривих рішень (decision curve analysis, Рис. 10) [16] на скоригованих (каліброваних під природну поширеність) ймовірностях: у діапазоні клінічно доцільних порогових ймовірностей моделі В і С дають додатну чисту клінічну користь, вищу за стратегії «лікувати всіх» та «не лікувати нікого».

Таблиця 4. Чутливість середньої клінічної вартості до співвідношення штрафів cFN : cFP

cFN : cFP	Вартість А	Вартість В	Вартість С	Найкраща модель
1:1	0.132	0.152	0.152	А
2:1	0.242	0.203	0.205	В
3:1	0.266	0.243	0.217	С
5:1	0.290	0.375	0.357	А
10:1	0.412	0.388	0.502	В

Джерело: розроблено авторами

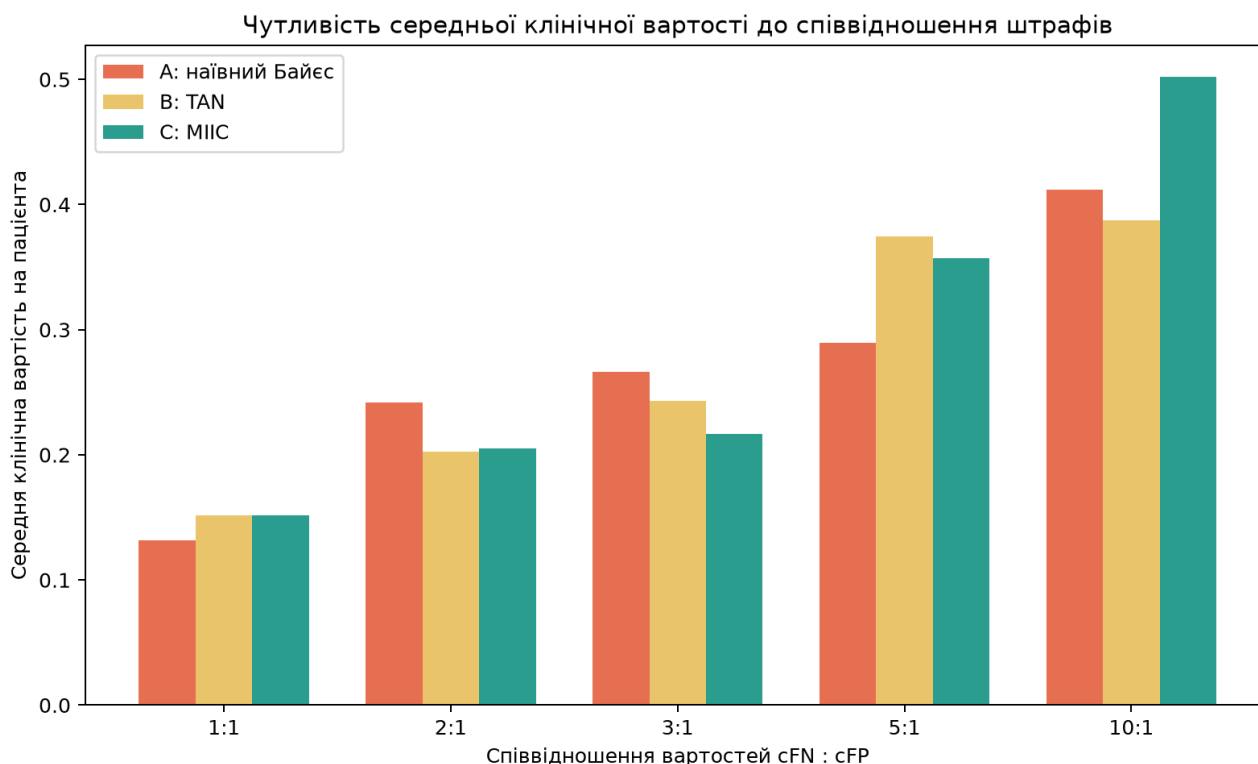


Рис. 9. Чутливість середньої клінічної вартості до співвідношення штрафів cFN:cFP

Джерело: розроблено авторами

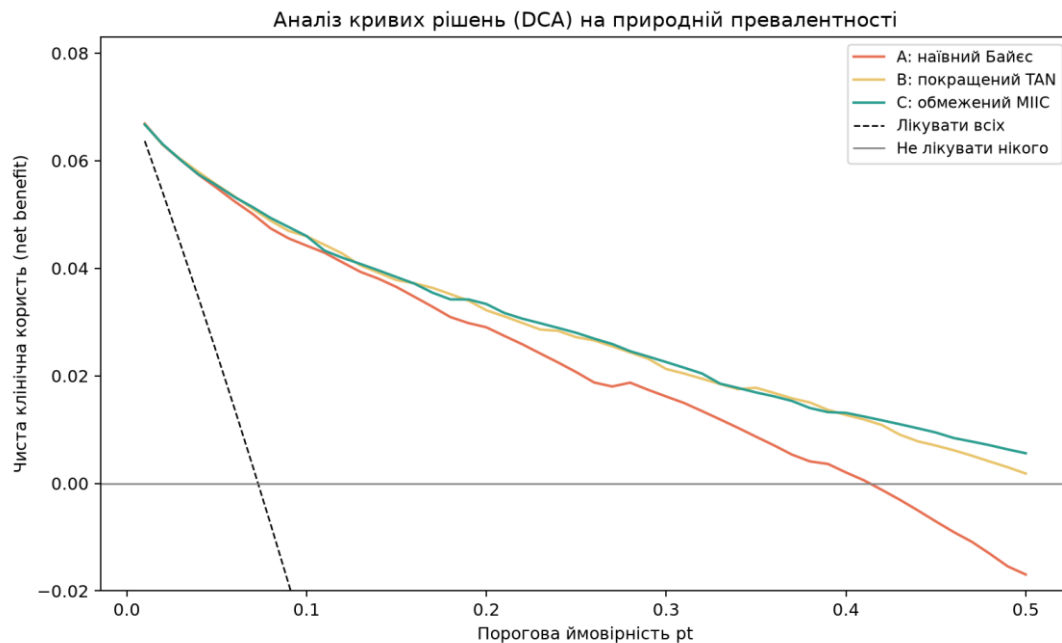


Рис. 10. Аналіз кривих рішень (DCA) моделей А, В, С на природній превалентності

Джерело: розроблено авторами

Стабільність структури: суміжність проти орієнтації. Оскільки за Марківської еквівалентності напрям статистично вивченої дуги не можна ототожнювати з підтвердженим напрямом клінічного впливу, стабільність структури оцінено окремо за двома складовими (bootstrap, $B = 500$, Рис. 11): частотою суміжності (наявністю ребра незалежно від орієнтації) та умовною частотою орієнтації (часткою того самого напрямку серед перевибірок, у яких ребро присутнє).

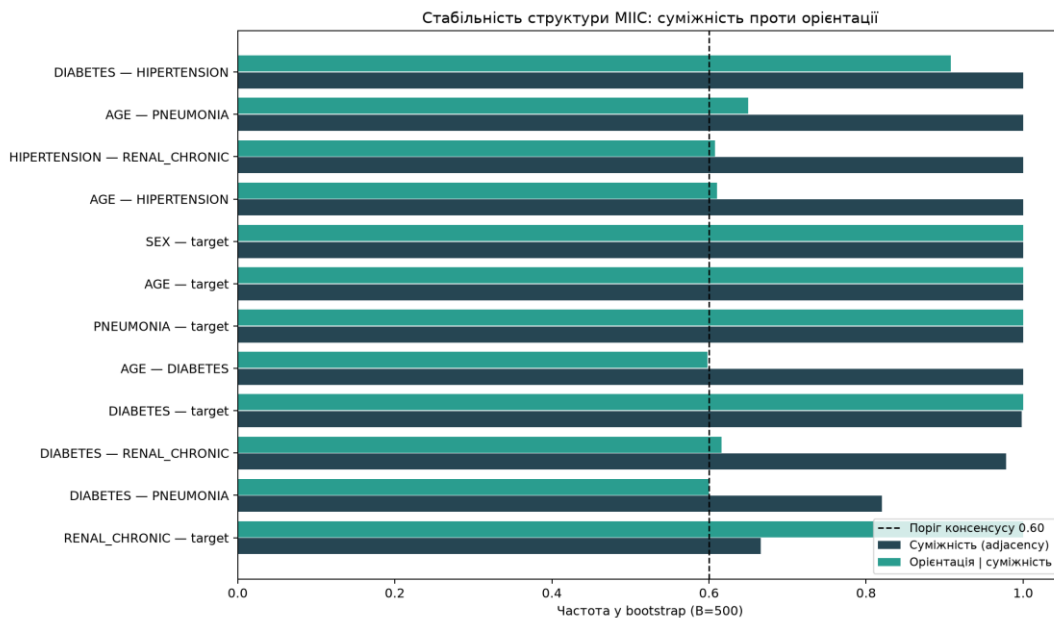


Рис. 11. Стабільність структури МПС: частота суміжності проти умовної частоти орієнтації ($B = 500$)

Джерело: розроблено авторами

Суміжності основних зв'язків є високостабільними (наприклад, SEX — target, DIABETES — target, AGE — HIPERTENSION, DIABETES — HIPERTENSION мають частоту суміжності близько 1,0), тоді як орієнтація частини ребер є неоднозначною: для AGE — PNEUMONIA умовна частота орієнтації становить лише 0,65, для AGE — HIPERTENSION — 0,61, для DIABETES — RENAL_CHRONIC — 0,62, для AGE — DIABETES — 0,60. Це прямо ілюструє

Марківську еквівалентність: суміжність підтверджується даними, а напрям – лише частково. Обов'язкові дуги PNEUMONIA $\rightarrow$ target та AGE $\rightarrow$ target зберігають 100 % суміжність і 100 % орієнтацію навіть у повністю необмеженому bootstrap-аналізі, що є незалежним емпіричним підтвердженням цих зв'язків. Експертні обмеження обґрунтовано клінічними джерелами: провідна прогностична роль пневмонії та віку при COVID-19 узгоджується з оглядовими та прикладними працями [5], [7], а заборону зворотних дуг від цільової змінної до фізіологічних ознак уведено, щоб унеможливити біологічно абсурдні залежності. Через можливу Марківську еквівалентність орієнтацію статистично вивчених дуг трактуємо як структурну гіпотезу, а не як доказ причинно-наслідкового впливу.

Висновки. У роботі досліджено вплив структурної складності байєсівських моделей на прогнозування летальності пацієнтів із COVID-19 шляхом порівняння наївного Байєса, TAN та структурно обмеженої байєсівської мережі, навченої алгоритмом МПС. Абляційне порівняння за однакових умов показало, що приріст якості ймовірного прогнозу зумовлений саме послабленням припущення умовної незалежності (переходом до багатшої структури), а не скороченням набору ознак, а структурно обмежена й повністю необмежена МПС-мережі збігаються за структурою та метриками. Збільшення структурної складності не поліпшує суттєво розпізнавальної здатності: усі три моделі мають близькі значення ROC AUC ($\approx 0,91$). Натомість модель С забезпечила найвище значення PR-AUC (0,451) та найменші значення Brier Score і log-loss.

Показано, що зміна поширеності цільового класу між навчальною та природною клінічною вибірками призводить до систематичного завищення прогнозованого ризику. Корекція апіорної ймовірності за незалежно оціненою поширеністю ($\approx 6,7$ %) істотно поліпшила калібрування моделей: для моделі С скоригована ЕСЕ становить $\approx 0,004$ із 95 % довірчим інтервалом $[0,002; 0,006]$, а аналіз чутливості підтвердив стійкість цього результату для поширеності від 3 % до 15 %. Отже, для систем імовірного прогнозування в умовах дисбалансу класів важливим є не лише розпізнавальна здатність, а й коректне перенесення прогнозованих імовірностей до цільової популяції за незалежно оціненою поширеністю.

У межах проведеного експерименту модель С продемонструвала найкраще поєднання якості ймовірного прогнозу та структурної інтерпретованості, хоча перевага не є універсальною за всіма аспектами: за єдиного вартісного протоколу й співвідношення 3:1 модель С має найменшу середню клінічну вартість, проте за інших співвідношень вартостей найдешевшими можуть бути простіші моделі, а аналіз кривих рішень підтверджує додатну клінічну користь складніших моделей лише в певному діапазоні порогових ймовірностей. Структурний аналіз із розділенням суміжності й орієнтації виявив високу стабільність суміжностей за неоднозначності напрямів окремих дуг; через можливу Марківську еквівалентність орієнтації статистично вивчених дуг трактуються як структурні гіпотези, а не як доказ причинно-наслідкових ефектів.

Практичне значення отриманих результатів полягає в обґрунтуванні комплексного оцінювання байєсівських моделей за розпізнавальною здатністю, якістю ймовірного прогнозу та калібруванням, а також в узгодженні порогів за єдиним вартісним протоколом і незалежному оцінюванні поширеності. Подальші дослідження доцільно спрямувати на зовнішню валідацію на незалежних реєстрах, урахування можливого правого цензурування, розширення набору клінічних ознак та аналіз стійкості результатів до структурних обмежень і параметрів навчання. За 95 % bootstrap-довірчими інтервалами перевага моделі С над В за Brier Score, log-loss та PR-AUC є статистично підтвердженою, тоді як відмінність за ROC AUC не є статистично значущою.

Конфлікт інтересів: Автори декларують, що не мають конфлікту інтересів, зокрема фінансового, особистого, авторського чи будь-якого іншого характеру, який міг би вплинути на дослідження, а також на результати, опубліковані в цій статті.

Фінансування: Дослідження проводилося без фінансової підтримки.

Доступність даних: Використано відкритий набір даних COVID-19 Уряду Мексики (Kaggle: meirnizri/covid19-dataset). Код моделювання доступний у відкритому репозиторії: <https://www.kaggle.com/code/romantick/bn-for-predicting-covid-19>.

mortality. Узагальнені результати, параметри експерименту та сформовані графічні матеріали наведено в основному тексті рукопису.

Використання засобів штучного інтелекту: Під час підготовки рукопису інструменти штучного інтелекту могли використовуватися лише для граматичної та стилістичної перевірки тексту. ШІ не застосовувався для генерації тексту, формул і таблиць, формування наукових положень, отримання результатів або формулювання висновків. Наукові положення, результати та висновки є власним інтелектуальним внеском авторів.

СПИСОК ЛІТЕРАТУРИ

1. Bralatan, R. A. & Zhukov, S. O. “Bayesian modeling for assessing lung cancer risk based on medical data analysis”. *Scientific Works of Vinnytsia National Technical University*. 2025; (3): 47–54. DOI: <https://doi.org/10.31649/2307-5376-2025-3-47-54>.
2. Zhukov, S. O. & Rudzetyvych, O. V. “Optimization of deep neural networks for speech emotion classification using dynamic quantization”. *Scientific Works of Vinnytsia National Technical University*. 2025; (2): 51–65. DOI: <https://doi.org/10.31649/2307-5376-2025-2-51-65>.
3. Kelly, C. J., Karthikesalingam, A., Suleyman M., Corrado, G. & King, D. “Key challenges for delivering clinical impact with artificial intelligence”. *BMC Medicine*. 2019; 17 (1): 195, <https://www.scopus.com/pages/publications/85074427658>. DOI: <https://doi.org/10.1186/s12916-019-1426-2>.
4. Nia, N. G., Kaplanoğlu, E. & Nasab, A. “Evaluation of artificial intelligence techniques in disease diagnosis and prediction”. *Discover Artificial Intelligence*. 2023; 3 (1): 5, <https://www.scopus.com/pages/publications/85187368062>. DOI: <https://doi.org/10.1007/s44163-023-00049-5>.
5. Polotskaya, K., Muñoz-Valencia, C. S., Rabasa, A., et al. “Bayesian networks for the diagnosis and prognosis of diseases: A scoping review”. *Machine Learning and Knowledge Extraction*. 2024; 6 (2): 1243–1262, <https://www.scopus.com/pages/publications/85196904996>. DOI: <https://doi.org/10.3390/make6020058>.
6. Corrales, D., Santos-Lozano, A., López-Ortiz, S., Lucía, A. & Ríos Insua, D. “Colorectal cancer risk mapping through Bayesian networks”. *Computer Methods and Programs in Biomedicine*. 2024; 257: 108407, <https://www.scopus.com/pages/publications/85203618974>. DOI: <https://doi.org/10.1016/j.cmpb.2024.108407>.
7. Lombaers, M., et al. “ENDORISK-2: A personalized Bayesian network for preoperative risk stratification in endometrial cancer”. *International Journal of Gynecological Cancer*. 2025; 35 (2): 100249. DOI: <https://doi.org/10.1016/j.ijgc.2024.100249>.
8. Wood, A., Shpilrain, V., Najarian, K. & Kahrobaei, D. “Private naive Bayes classification of personal biomedical data: Application in cancer data analysis”. *Computers in Biology and Medicine*. 2019; 105: 144–150, <https://www.scopus.com/pages/publications/85059820470>. DOI: <https://doi.org/10.1016/j.compbimed.2018.11.018>.
9. Taheri, S. & Mammadov, M. “Learning the naive Bayes classifier with optimization models”. *International Journal of Applied Mathematics and Computer Science*. 2013; 23 (4): 787–795. <https://www.scopus.com/pages/publications/84893722466>. DOI: <https://doi.org/10.2478/amcs-2013-0059>.
10. Cheng, J. & Greiner, R. “Learning Bayesian networks from data: An information-theoretic approach”. *Artificial Intelligence*. 2001; 137 (1–2): 43–90. DOI: [https://doi.org/10.1016/S0004-3702\(02\)00191-1](https://doi.org/10.1016/S0004-3702(02)00191-1).
11. Scanagatta, M., Salmerón, A. & Stella, F. “A survey on Bayesian network structure learning from data”. *Progress in Artificial Intelligence*. 2019; 8 (4): 425–439. DOI: <https://doi.org/10.1007/s13748-019-00194-y>.
12. Nizri, M. “COVID-19 Dataset (Mexican Government COVID-19 registry)”. *Kaggle*. 2021. – Available from: <https://www.kaggle.com/datasets/meirnazri/covid19-dataset>. – [Accessed: 2026].
13. Bralatan, R. A. “BN for Predicting COVID-19 Mortality”. *Kaggle*. 2025. – Available from: <https://www.kaggle.com/code/romantick/bn-for-predicting-covid-19-mortality>. – [Accessed: 2026].
14. Verny, L., Sella, N., Affeldt, S., Singh, P. P. & Isambert, H. “Learning causal networks with latent variables from multivariate information in genomic data”. *PLoS Computational Biology*. 2017; 13 (10): e1005662. DOI: <https://doi.org/10.1371/journal.pcbi.1005662>.

15. Friedman, N., Geiger, D. & Goldszmidt, M. “Bayesian network classifiers”. *Machine Learning*. 1997; 29 (2–3): 131–163. DOI: <https://doi.org/10.1023/A:1007465528199>.
16. Vickers, A. J. & Elkin, E. B. “Decision curve analysis: a novel method for evaluating prediction models”. *Medical Decision Making*. 2006; 26 (6): 565–574. DOI: <https://doi.org/10.1177/0272989X06295361>.
17. Saerens, M., Latinne, P. & Decaestecker, C. “Adjusting the outputs of a classifier to new a priori probabilities: A simple procedure”. *Neural Computation*. 2002; 14 (1): 21–41. DOI: <https://doi.org/10.1162/089976602753284446>.
18. Domingos, P. “MetaCost: A general method for making classifiers cost-sensitive”. *Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. 1999. 155–164. DOI: <https://doi.org/10.1145/312129.312220>.

Отримано 10.07.2026

Отримано після перегляду 26.08.2026

Прийнято 03.09.2026

DOI: <https://doi.org/10.15276/ict.03.2026.22>

UDC 004.67:616.98

Comparison of bayesian networks of different structural complexity for predicting COVID-19 mortality

Roman A. Bralatan¹⁾

ORCID: <https://orcid.org/0009-0002-3439-5234>; bralatan.roman@gmail.com

Serhii O. Zhukov¹⁾

ORCID: <https://orcid.org/0009-0002-3443-6139>; sazhukov@gmail.com. Scopus Author ID: 57191729653

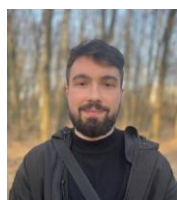
¹⁾ Vinnytsia National Technical University, 95, Khmelnytske Shose. Vinnytsia, 21021, Ukraine

ABSTRACT

This paper considers a methodology for developing, tuning, and comparing three probabilistic models – Bayesian networks of different classes with increasing structural complexity – for the task of predicting the mortality of patients with coronavirus disease. The compared models are the naive Bayes classifier, the tree-augmented naive Bayes classifier, and a structurally constrained Bayesian network whose structure is recovered from the data by a structure-learning algorithm under expert clinical constraints. To separate the effect of structural complexity from feature selection, parameter tuning, and threshold choice, a comparison under matched conditions with successive simplification and enrichment of the model structure is carried out. Particular attention is paid to distinguishing two aspects of model quality – the discriminative ability and the quality (reliability) of the predicted probabilities – because for a clinical decision-support system it is the trustworthiness of the risk estimate, rather than the mere ranking of patients by their level of risk, that is decisive. The study is conducted on an open registry of patients with corona virus disease using clinical features, with evaluation on a sample that reproduces the natural prevalence of fatal outcomes. The paper also considers the transfer of the predicted probabilities to the natural population through a prior-probability correction, as well as the sensitivity of clinical decisions to the ratio of error costs and an analysis of the net clinical benefit. The aim of the work is to substantiate the suitability of a structurally constrained Bayesian network as an interpretable risk-assessment tool in clinical decision-support systems.

Keywords: Bayesian networks; Naive Bayes; TAN; MIIC; PyAgrim; COVID-19; mortality prediction; probability calibration; interpretable machine learning; structure learning; decision support systems

ІНФОРМАЦІЯ ПРО АВТОРІВ



Бралатан Роман Анатолійович – аспірант кафедри Системного аналізу та інформаційних технологій. Вінницький національний технічний університет, Хмельницьке шосе, 95. Вінниця, 21021, Україна
Галузь досліджень: системний аналіз

Roman A. Bralatan – PhD Student, Department of System Analysis and Information Technologies. Vinnytsia National Technical University, 95, Khmelnytske Shose. Vinnytsia, 21021, Ukraine
ORCID: <https://orcid.org/0009-0002-3439-5234>; bralatan.roman@gmail.com
Research field: system analysis



Жуков Сергій Олександрович – кандидат технічних наук, доцент кафедри Системного аналізу та інформаційних технологій. Вінницький національний технічний університет, Хмельницьке шосе, 95. Вінниця, 21021, Україна
Галузь досліджень: системний аналіз

Serhii O. Zhukov – PhD (Engineering), Associate Professor, Department of System Analysis and Information Technologies. Vinnytsia National Technical University, 95, Khmelnytske Shose. Vinnytsia, 21021, Ukraine
ORCID: <https://orcid.org/0009-0002-3443-6139>; sazhukov@gmail.com. Scopus Author ID: 57191729653
Research field: system analysis