**DOI: https://doi.org/10.15276/ict.03.2026.27**
**УДК 004.85:004.032.26:616.1**

# Порівняльне оцінювання байєсівських та нейромережевих моделей прогнозування серцево-судинних захворювань з урахуванням калібрування та невизначеності

**Пономарьов Артем Олегович**[1)]
ORCID: https://orcid.org/0009-0001-2112-6759; artponexcelsior@gmail.com
**Козачко Олексій Миколайович**[1)]
ORCID: https://orcid.org/0000-0003-3033-9745; lekoz80@gmail.com. Scopus Author ID: 57200139908
[1)] Вінницький національний технічний університет, Хмельницьке шосе, 95. Вінниця, 21021, Україна

## АНОТАЦІЯ

У роботі досліджено проблему надійності ймовірнісного прогнозування серцево-судинних захворювань у системах підтримки прийняття рішень. На відміну від підходу, за якого якість моделі оцінюють переважно за однією метрикою класифікації, розглянуто сукупність взаємопов'язаних властивостей прогнозної системи: дискримінаційну здатність, відповідність прогнозованих імовірностей фактичним частотам, оцінку невизначеності, здатність виявляти помилкові прогнози та можливість відмови від автоматичного рішення для складних випадків. Експерименти виконано на відкритому наборі Cardiovascular Disease Dataset. Після консервативного очищення проаналізовано понад шістдесят вісім тисяч спостережень; із вихідних ознак сформовано двадцять одну вхідну змінну після кодування та масштабування. У єдиній схемі порівняно наївний байєсівський класифікатор, байєсівську мережу, логістичну регресію, багатошаровий перцептрон, стохастичне виключення нейронів за методом Монте-Карло, наближення Лапласа для останнього шару, байєсівський останній шар із вибіркою Гамільтона–Монте-Карло та глибокий ансамбль. Показано, що на повній тестовій вибірці розходження моделей за здатністю ранжувати спостереження є невеликим: найбільше значення площі під характеристичною кривою «чутливість–специфічність» отримав глибокий ансамбль, тоді як Neural-Laplace мав найменшу очікувану похибку калібрування серед досліджених моделей. Для моделей із розподілом прогнозів оцінка невизначеності виявляла помилкові прогнози краще за випадкове ранжування; найвище значення відповідної площі під характеристичною кривою отримано для Neural-Laplace. Селективне прогнозування показало, що вилучення випадків із найбільшою оціненою невизначеністю із автоматичної обробки зменшує ризик на решті вибірки. Експеримент зі зменшенням навчальної вибірки показав, що при використанні десяти відсотків навчальних даних різниця між Neural-Laplace та багатошаровим перцептроном за показниками калібрування була найбільш вираженою: очікувана похибка калібрування Neural-Laplace виявилася у понад сім разів меншою. Науковий результат полягає у встановленні в межах єдиної експериментальної схеми того, що байєсівська складова не забезпечує автоматичного приросту дискримінаційної здатності, але може змінювати профіль якості прогнозної системи за калібруванням і невизначеністю: Neural-Laplace мала найнижчу очікувану похибку калібрування та найкращу здатність виявляти помилкові прогнози за оцінкою невизначеності. Практична цінність полягає у використанні такого профілю характеристик для побудови селективного прогнозування в системах підтримки прийняття рішень.

**Ключові слова:** серцево-судинні захворювання; системний аналіз; прогнозування ризику; байєсівське моделювання; нейронні мережі; калібрування ймовірностей; невизначеність прогнозу; селективне прогнозування

***Для цитування:*** **Пономарьов А. О., Козачко О.М. "Порівняльне оцінювання байєсівських та нейромережевих моделей прогнозування серцево-судинних захворювань з урахуванням калібрування та невизначеності".** ***Інформатика. Культура. Техніка.*** **2026; Том 3 № 1 (3): 321–335. DOI: https://doi.org/10.15276/ict.03.2026.27**

**Актальність**. Для інформаційної системи підтримки прийняття рішень важливо відрізняти два питання: наскільки добре модель розділяє два класи та наскільки надійними є числові ймовірності, які вона подає користувачеві. У першому випадку достатньо оцінити здатність моделі правильно впорядковувати об'єкти за ризиком. У другому потрібно встановити, чи відповідає, наприклад, прогноз 0,8 приблизно вісімдесятивідсотковій частоті події серед подібних випадків. Для практичного аналізу даних ця різниця принципова, оскільки одна й та сама модель може мати добру класифікаційну здатність і водночас бути надмірно впевненою у власних помилках. Проблематика якості та стійкості моделей машинного навчання у ВНТУ вже розглядається в контексті виявлення аномальних спостережень у медичній статистиці та оптимізації нейромережевих архітектур [1], [2].

У медичній задачі прогнозування серцево-судинного захворювання така відмінність посилюється тим, що рішення системи може бути не бінарним «прийняти/відхилити»,

© Пономарьов А., Козачко О., 2026

Стаття опублікована у відкритому доступі відповідно до умов ліцензії CC BY (https://creativecommons.org/licenses/by/4.0/deed.uk)

а багаторівневим: автоматично сформувати рекомендаційний сигнал, передати випадок на додаткове обстеження або утриматися від автоматичної класифікації. Отже, система має не тільки оцінювати ймовірність цільового стану, а й передавати інформацію про власну невпевненість. Це переводить задачу з простого порівняння класифікаторів до задачі системного аналізу, у якій одночасно досліджуються точність розпізнавання, властивості ймовірнісного оцінювання та спосіб взаємодії алгоритму з особою, що приймає рішення.

Особливий інтерес становить контрольоване порівняння кількох способів формування прогнозу на одній і тій самій нейромережевій основі. У такому разі різниця між детермінованим багатошаровим перцептроном, стохастичним виключенням нейронів, наближенням Лапласа, вибіркою Гамільтона-Монте-Карло та ансамблем визначається передусім способом представлення параметричної невизначеності, а не зміною всього екстрактора ознак. Саме такий контрольований експеримент дає змогу поставити більш змістовне питання: чи створює байєсівська постановка додаткову корисність поверх тієї інформації, яку вже отримує базова нейронна мережа?

**Аналіз останніх досліджень і публікацій.** Проблема калібрування сучасних нейронних мереж стала окремим напрямом досліджень після встановлення того, що висока точність не гарантує адекватної відповідності між упевненістю моделі та реальною частотою правильних рішень [3]. Калібрування розглядається як властивість імовірнісного прогнозу, а не як синонім точності. Це важливе розмежування для задач, де вихід моделі використовується як оцінка ризику.

Стохастичне виключення нейронів може бути інтерпретоване як наближений байєсівський механізм оцінювання невизначеності, що дозволяє одержувати не одну, а сукупність прогнозів для одного спостереження [4]. Інший напрям пов’язаний із глибокими ансамблями: незалежне навчання кількох моделей і агрегація їхніх прогнозів дає простий спосіб побудови розподілу прогнозів та часто забезпечує конкурентну якість оцінки невизначеності [5].

Для нейромережевих моделей із великою кількістю параметрів повне байєсівське виведення є практично складним. Тому використовуються локальні наближення до апостеріорного розподілу. Наближення Лапласа є одним із таких підходів і може бути застосоване лише до останнього шару нейронної мережі, залишаючи вже навчене нелінійне перетворення ознак незмінним [6]. Метод Гамільтона–Монте-Карло, своєю чергою, дає можливість безпосередньо здійснювати вибірку з цільового апостеріорного розподілу параметрів, хоча його обчислювальна вартість є вищою [7].

Окремо розвивається напрям селективної класифікації, у якому система може відмовлятися від автоматичного прогнозу. Відмова дозволяє зменшити ризик серед тих випадків, які передаються алгоритму для остаточної класифікації, ціною зменшення охоплення [8]. Такий підхід узгоджується з концепцією людино-орієнтованих клінічних систем підтримки прийняття рішень, у яких алгоритм доповнює, а не повністю замінює професійне рішення [9].

Для кількісного оцінювання невизначеності застосовують також конформні методи, які перетворюють точковий прогноз на множину можливих класів із наперед заданим рівнем покриття [10]. Водночас поведінка невизначеності може змінюватися за зміни розподілу даних, тому оцінювання лише на одній тестовій вибірці не дає підстав говорити про універсальну надійність моделі [11]. Для клінічних систем важливість індивідуальної оцінки прогнозної невизначеності також підкреслюється в сучасних дослідженнях [12].

У зв’язку з цим комплексні порівняння, де одночасно аналізують розпізнавальну здатність, калібрування, невизначеність і поведінку за обмеженої навчальної вибірки, мають окрему методологічну цінність.

**Мета дослідження.** Метою роботи є порівняльне системне оцінювання байєсівських та нейромережевих методів прогнозування серцево-судинного захворювання за умови, що якість розглядається як багатовимірна характеристика. Для досягнення мети поставлено такі завдання: формалізувати прогнозну систему як об’єкт системного аналізу; сформувати єдину

експериментальну схему з розділенням навчання, налаштування та фінальної перевірки; зіставити моделі за дискримінаційною здатністю та якістю прогнозованих імовірностей; дослідити невизначеність та її здатність розрізняти правильні і помилкові рішення; оцінити селективне прогнозування; перевірити, як змінюється співвідношення властивостей моделей при зменшенні навчальної вибірки; та визначити, як змінюється профіль властивостей моделей при зменшенні навчальної вибірки, а також оцінити стабільність ранжування ознак.

**Матеріали та методи.** У роботі прогнозування розглядається як функціонування інформаційної системи, яка отримує вектор характеристик об'єкта спостереження, оцінює ймовірність цільової події, а для стохастичних моделей додатково формує характеристику невизначеності.

Формально систему можна записати як:

$$\mathrm{S} = \langle \mathrm{X}, \mathrm{Y}, \mathrm{F}\theta, \mathrm{U}, \delta, \mathrm{g} \rangle, \tag{1}$$

де $X \subset R^{\mathrm{d}}$ – простір вхідних характеристик; $Y = \{0,1\}$ – множина станів; $F\theta(x) \in [0,1]$ – імовірнісний прогноз; $U(x)$ – міра невизначеності; $\delta$ – правило прийняття бінарного рішення; g – правило відмови від автоматичної класифікації.

У такому поданні якість системи не зводиться до одного числа. Вона визначається вектором критеріїв:

$$Q(M) = [D(M), P(M), C(M), U(M), R_{sel}(M), E(M)], \tag{2}$$

де $M$ – досліджувана модель; $D$ – критерії дискримінаційної здатності; $P$ – якість прогнозованих імовірностей; $C$ – калібрування; $U$ – корисність оцінки невизначеності; $R_{sel}$ – характеристики селективного прогнозування; $E(M)$ – показники стабільності пояснюваності.

Таким чином, задача порівняння полягає не у пошуку моделі, яка максимізує один показник, а у визначенні відносного профілю властивостей *Q(M)* та встановленні того, за яких умов одна модель має перевагу за окремими групами критеріїв. Така постановка відповідає суті системного аналізу: потрібно дослідити не тільки локальну ефективність алгоритму, а й взаємозв'язок його властивостей у межах цілісної системи прийняття рішень.

Оскільки наведені критерії мають різну природу та не можуть бути зведені до одного показника без додаткових припущень щодо їхніх ваг, у роботі застосовано компаративний аналіз профілів властивостей моделей. Порівняння проводиться окремо за групами критеріїв, після чого формується узагальнений висновок про характер переваг і компромісів між дискримінаційною здатністю, якістю ймовірнісного прогнозу, оцінкою невизначеності та селективним прогнозуванням. Такий підхід відповідає системному аналізу прогнозної системи без введення довільної інтегральної функції якості.

Для обраного набору даних маємо навчальну множину:

$$D_{tr} = \{(xi, yi)\} i = 1 \ldots N. \tag{3}$$

У детермінованому випадку оцінюється один набір параметрів $\theta$, тоді як у байєсівській постановці розглядається розподіл параметрів.

Для базової мережі використано мінімізацію регуляризованої функції логарифмічних втрат:

$$\hat{\theta} = \arg\min_{\theta} \left\{ -\left(\frac{1}{N}\right) \sum_{i=1}^{N} \left[ y_i \ln p_\theta(x_i) + (1 - y_i) \ln\left(1 - p_\theta(x_i)\right) \right] + \lambda \left|\left|\theta\right|\right|_2^2 \right\}. \tag{4}$$

Тут $p_\theta(x_i)$ – прогнозована моделлю ймовірність належності $i$-го спостереження до позитивного класу, $y_i \in \{0,1\}$, $N$ – обсяг навчальної вибірки, а $\lambda$ – коефіцієнт L2-регуляризації.

Байєсівська постановка замінює одну точкову оцінку розподілом параметрів:

$$p(\theta|D) = p(D|\theta)p(\theta) / \int p(D|\theta')p(\theta')d\theta'. \tag{5}$$

Тоді прогноз для нового вектора $x^*$ визначається маргіналізацією за параметрами:

$$p(y^* = 1|x^*, D) = \int p(y^* = 1|x^*, \theta)p(\theta|D)d\theta. \quad (6)$$

Для глибокої мережі цей інтеграл не обчислюється аналітично, тому в експерименті використано три різні способи його наближення або заміщення розподілом прогнозів.

Для стохастичного виключення нейронів середній прогноз задається як:

$$\bar{p}(x) \approx \left(\frac{1}{T}\right) \sum_{t=1}^{T} p_{\theta_t}(x), \quad (7)$$

де $T = 60$ – кількість стохастичних проходів.

У випадку стохастичного виключення нейронів значення отримуються за послідовних стохастичних проходів тієї самої навченої мережі при різних реалізаціях маски виключення нейронів; параметри мережі при цьому не перенавчаються.

Для наближення Лапласа апостеріорний розподіл останнього шару апроксимується нормальним розподілом:

$$p(w|D) \approx N(w_{MAP}, H^{-1}), \quad (8)$$

де $H$ – матриця Гессе негативного логарифма апостеріорної щільності параметрів у точці MAP; у практичній реалізації використано відповідне наближення кривизни, передбачене методом Лапласа. У реалізації використовувалося узагальнене наближення Гаусса–Ньютона/логістичної кривизни та 120 вибірок з наближеного апостеріорного розподілу. Для байєсівського останнього шару з методом Гамільтона–Монте-Карло використовувалися 60 вибірок після 60 ітерацій прогріву (burn-in).

Для стохастичних моделей як індикатор невизначеності використано вибіркове стандартне відхилення множини прогнозованих імовірностей.

Нехай $\mu(x)$ – середня прогнозована ймовірність, тоді:

$$U(x) = \sqrt{\left(\frac{1}{T-1}\right) \sum_{t=1}^{T} \left(p_t(x) - \mu(x)\right)^2}. \quad (9)$$

Саме *U(x)* у подальшому використовується як порядок, за яким випадки сортуються для аналізу виявлення помилок та селективного прогнозування. Отримана величина характеризує варіативність прогнозів конкретної моделі і в цій роботі не ототожнюється з повним розкладанням прогнозної невизначеності на епістемічну та алеаторну складові.

**Дані та попередня обробка.** Експерименти проведено на відкритому наборі «Cardiovascular Disease Dataset» [13]. Початково він містить 70 000 спостережень та 16 колонок. Цільова змінна cardio визначає наявність серцево-судинного захворювання. Додатково сформовано вік у роках, індекс маси тіла та пульсовий тиск. Ідентифікатор id до набору предикторів не включався.

Основним сценарієм було консервативне очищення. З нього вилучено 1 394 записи з фізіологічно некоректними або сумнівними комбінаціями віку, зросту, маси тіла та артеріального тиску. До фізіологічно сумнівних відносилися записи, що не відповідали встановленим у програмній реалізації межам для віку, зросту, маси тіла та артеріального тиску, а також записи з некоректним співвідношенням систолічного та діастолічного тиску. Конкретні правила фільтрації наведено в експериментальному ноутбуці [14]. Після цього залишилося 68 606 спостережень. Цільова змінна лишилася близькою до збалансованої: частка позитивного класу становила 49,4694 %, а негативного – 50,5306 %. Для перевірки чутливості використовувався суворіший сценарій очищення, який залишав 68 329 записів, але не застосовувався для вибору основної моделі.

Після очищення вхідні дані містили 13 логічно осмислених ознак: вік у роках, зріст, масу тіла, індекс маси тіла, систолічний і діастолічний артеріальний тиск, пульсовий тиск, стать, рівень холестерину, рівень глюкози, факт куріння, вживання алкоголю та фізичну активність. Безперервні ознаки масштабувалися за допомогою StandardScaler, категоріальні – кодувалися методом one-hot. У результаті формувалася матриця з 21 обробленої ознаки.

Дані розділено стратифіковано у співвідношенні 70:15:15. Навчальна частина містила 48 024 записи, валідаційна – 10 291, тестова – 10 291. Параметри масштабування та кодування оцінювалися виключно на навчальній частині й після цього застосовувалися до валідаційної і тестової частин. Порогові значення класифікації визначалися лише за валідаційною вибіркою за критерієм Юдена. Тестова вибірка використовувалася один раз для фінального оцінювання. Фіксоване зерно випадкових чисел дорівнювало 42.

**Дизайн експерименту та критерії оцінювання.** Усього порівняно вісім моделей: наївний байєсівський класифікатор, байєсівську мережу у формі Tree-Augmented Naive Bayes, логістичну регресію, багатошаровий перцептрон, метод стохастичного виключення нейронів (Monte Carlo Dropout або MC Dropout), нейромережева модель із наближенням Лапласа (Neural-Laplace), байєсівська модель останнього шару (Bayesian Last Layer) із вибіркою методом Гамільтона–Монте-Карло (Hamiltonian Monte Carlo або HMC) та глибокий ансамбль нейронних мереж (Deep Ensemble). Для MLP, MC Dropout, Neural-Laplace та HMC використано спільну нейромережеву основу з двома прихованими шарами (32 і 16 нейронів), dropout 0,15, оптимізатором AdamW, швидкістю навчання 0,001, L2-регуляризацією 0,0001, розміром пакета 1024 та ранньою зупинкою після шести епох без покращення валідаційної втрати; максимальна кількість епох – 24. MC Dropout використовував 60 стохастичних проходів, Neural-Laplace – 120 вибірок апостеріорного розподілу, HMC – 60 вибірок після 60 ітерацій прогріву та вісім кроків leapfrog. Глибокий ансамбль складався з п'яти моделей із початковими значеннями 42–46. Основне порівняння виконувалося на одному зафіксованому розділенні з початковим значенням 42, а експеримент із різним обсягом навчальної вибірки – для трьох початкових значень; статистичні відмінності оцінювали парною бутстреп-перевіркою.

Таке компонування має методичну перевагу: порівняння змінює не всю архітектуру системи, а спосіб формування розподілу прогнозів. Тому спостережувана різниця може бути інтерпретована передусім як наслідок відмінностей у способі представлення невизначеності та агрегування прогнозів. Байєсівська мережа і наївний Байєс розглядаються як окремі ймовірнісні базові лінії, а не як частина цієї контрольованої нейромережевої четвірки.

Для дискримінаційної здатності використано площу під ROC-кривою та площу під кривою точність–повнота, а для порогових прогнозів – F1.

Якість самих імовірностей оцінено оцінкою Брайєра та логарифмічними втратами:

$$BS = \left(\frac{1}{N}\right) \Sigma_{\text{i}} (p_{\text{i}} - y_{\text{i}})^2, \tag{10}$$

$$LL = -\left(\frac{1}{N}\right) \sum_i [y_{\text{i}} \ln p_{\text{i}} + (1 - y_{\text{i}}) \ln(1 - p_{\text{i}})]. \tag{11}$$

Очікувана похибка калібрування розраховувалася за розбиттям прогнозів на $M$ інтервалів:

$$ECE = \sum_m \left(\frac{|B_{\text{m}}|}{N}\right) |acc(B_{\text{m}}) - conf(B_{\text{m}})|, \tag{12}$$

де $N$ – загальна кількість тестових спостережень; $B_m$ – множина спостережень, прогнозовані ймовірності яких належать до $m$-го інтервалу; $M$ – загальна кількість непорожніх або заданих інтервалів; $acc(B_m)$ – фактична частка позитивних випадків у $B_m$; $conf(B_m)$ – середня прогнозована ймовірність у цьому інтервалі.

Для обчислення очікуваної похибки калібрування (Expected Calibration Error, ECE) діапазон прогнозованих імовірностей [0,1] поділяли на 15 рівних інтервалів. Для кожного непорожнього інтервалу $B_m$ обчислювали середню прогнозовану ймовірність та фактичну частоту позитивного класу. Підсумкове значення ECE визначали як зважену за кількістю спостережень суму абсолютних відхилень між цими величинами.

Для калібрувального аналізу додатково розраховано дільний член і нахил калібрування. Вільний член, близький до нуля, характеризує відсутність систематичного зсуву, тоді як

нахил, близький до одиниці, вказує на відповідний масштаб упевненості. Ці показники не замінюють ECE, а доповнюють його.

Для статистичного порівняння моделей застосовано парну бутстреп-перевибірку однакових тестових спостережень із 1000 повтореннями. Для кожного порівняння оцінювали різницю відповідної метрики між моделями та її 95% довірчий інтервал; отримані p-значення коригували за методом Холма для множинних порівнянь. Статистично значущими вважали відмінності при скоригованому $p < 0,05$. Основну увагу приділено порівнянням нейромережевих моделей зі спільною архітектурною основою, оскільки саме вони дають змогу ізолювати вплив способу формування прогнозного розподілу.

**Результати та їх обговорення.** На повній тестовій вибірці всі чотири моделі, побудовані на спільній нейромережевій основі, показали майже однакову здатність розділяти класи. Найвище ROC-AUC мав глибокий ансамбль – 0,8097. Для MLP цей показник становив 0,8090, для Neural-Laplace – 0,8089, для MC Dropout – 0,8089, а для байєсівського останнього шару – 0,8088. Отже, зміна способу представлення невизначеності не дала суттєвого зрушення у ранжуванні об'єктів.

Цей результат важливий як емпіричне розмежування двох рівнів якості. Близькі значення ROC-AUC означають, що за використаного набору ознак та заданого експериментального протоколу зміна способу формування прогнозного розподілу майже не вплинула на ранжування тестових спостережень. Отже, додаткове ускладнення імовірнісної частини системи в цьому експерименті не супроводжувалося суттєвим зростанням ROC-AUC, а основні відмінності проявлялися у якості прогнозованих імовірностей та властивостях оцінки невизначеності.

За PR-AUC результати також були близькими: від 0,7929 для HMC до 0,7948 для MC Dropout. Найвище F1 отримав глибокий ансамбль – 0,7288, а серед байєсівських підходів найвище значення мав Neural-Laplace – 0,7275. Для логістичної регресії, байєсівської мережі та наївного Байєса відповідні показники були нижчими, що відображає різницю не тільки в способі подання невизначеності, а й у класі базової моделі.

Основні результати на повній тестовій вибірці наведено в Таблиці 1.

*Таблиця 1.* **Основні результати на повній тестовій вибірці**

| Модель | ROC AUC | PR-AUC | F1 | Brier | Log-loss | ECE |
|---|---|---|---|---|---|---|
| Наївний Байєс | 0,7648 | 0,7361 | 0,7049 | 0,2602 | 1,1670 | 0,2357 |
| Байєсівська мережа | 0,7931 | 0,7755 | 0,7133 | 0,1839 | 0,5498 | 0,0213 |
| Логістична регресія | 0,8003 | 0,7830 | 0,7231 | 0,1832 | 0,5507 | 0,0394 |
| MLP | 0,8090 | 0,7942 | 0,7241 | 0,1775 | 0,5335 | 0,0124 |
| MC Dropout | 0,8089 | 0,7948 | 0,7234 | 0,1775 | 0,5337 | 0,0132 |
| Neural-Laplace | 0,8089 | 0,7937 | 0,7275 | 0,1774 | 0,5332 | 0,0116 |
| HMC, останній шар | 0,8088 | 0,7929 | 0,7233 | 0,1774 | 0,5332 | 0,0122 |
| Глибокий ансамбль | 0,8097 | 0,7944 | 0,7288 | 0,1773 | 0,5331 | 0,0187 |

***Джерело:*** **розроблено авторами**

Як видно з Таблиці 1, найменше значення оцінки Брайєра (Brier Score) серед усіх моделей отримав глибокий ансамбль (0,1773), а найменше ECE – Neural-Laplace (0,0116). Отже, у межах сукупності використаних критеріїв жодна з моделей не мала одночасної переваги за всіма характеристиками прогнозної системи. Neural-Laplace мала трохи вищий Brier Score, ніж ансамбль, але краще поводилася за ECE. Це ще раз показує, що різні критерії описують різні сторони прогнозної системи.

Вільний член калібрувальної моделі Neural-Laplace становив -0,006720, нахил – 1,045638. Для HMC ці показники становили -0,002838 та 1,040203, для MLP – 0,042738 та 1,046951. Відхилення не є великими, але їх напрям підтверджує висновок, отриманий за ECE: Neural-Laplace мала найменше значення ECE серед основних нейромережевих

варіантів, що вказує на найменше середнє відхилення між прогнозованою впевненістю та емпіричною частотою в межах використаної схеми оцінювання.

Калібрувальні криві моделей на тестовій вибірці показані на Рис.1.

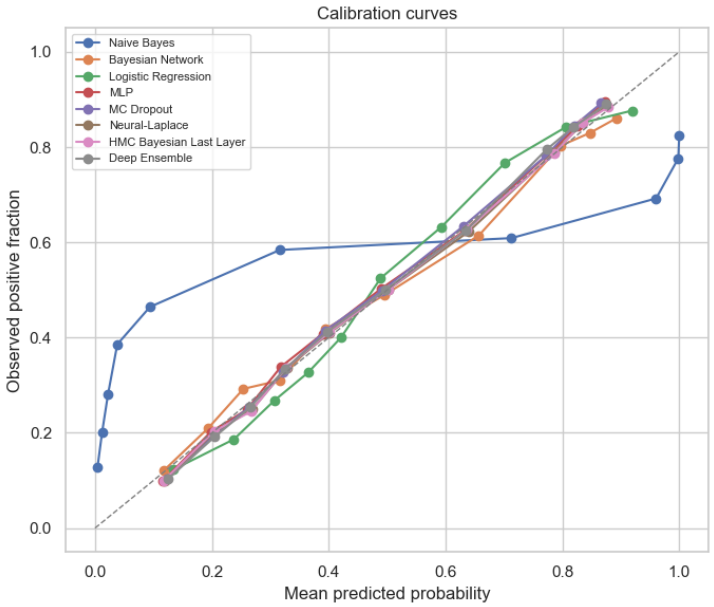


*Рис. 1.* **Калібрувальні криві моделей на тестовій вибірці**
***Джерело:*** **розроблено авторами**

Як видно з Рис. 1, більшість нейромережевих моделей формує калібрувальні криві, близькі за характером до діагоналі ідеального калібрування. Відхилення наївного Байєса є істотно більшими, тоді як для Neural-Laplace крива загалом ближча до діагоналі, що узгоджується з її найнижчим значенням ECE серед досліджених моделей. Водночас відмінності за графічним поданням не слід трактувати як статистично підтверджену перевагу без відповідного порівняння довірчих інтервалів.

Парна бутстреп-перевірка з 1000 повторень на тих самих тестових спостереженнях використовувалася для оцінювання довірчих інтервалів і статистичної значущості відмінностей між моделями. Для Neural-Laplace відносно MLP середня різниця ROC-AUC становила −0,000047, 95 % довірчий інтервал [−0,000530; 0,000434], скориговане значення p = 1,000. Для HMC різниця становила −0,000165 з інтервалом [−0,000751; 0,000432]. Для глибокого ансамблю спостерігалося підвищення ROC-AUC на 0,000690, але після корекції Холма p = 0,308. Отже, на повній тестовій вибірці отримані відмінності ROC-AUC між MLP та Neural-Laplace, HMC і Deep Ensemble не досягли статистичної значущості після корекції множинних порівнянь. Це не свідчить про математичну ідентичність моделей, а лише означає, що в межах цього тестового набору не отримано достатніх статистичних підстав стверджувати про відмінність їхньої дискримінаційної здатності.

Для Brier Score та ECE картина також не дає підстав говорити про загальну статистично значущу перевагу всіх байєсівських моделей. Для Neural-Laplace різниця Brier відносно MLP мала позитивний напрям для метрики «менше – краще», але 95% інтервал охоплював нуль. Аналогічно різниця ECE була близькою до нуля. Отже, коректніше говорити про описову перевагу Neural-Laplace за ECE, а не про доведену статистичну перевагу. Водночас це спостереження має практичний інтерес, оскільки найменше ECE поєднувалося з найкращою здатністю оцінки невизначеності виявляти помилки.

Для Neural-Laplace різниця ECE порівняно з MLP становила −0,000023 (95 % ДІ [−0,005993; 0,005255], скориговане p=1,000). Отже, зменшення ECE має описовий характер і не є статистично підтвердженою перевагою.

Для чотирьох стохастичних моделей було отримано набори прогнозів для кожного тестового спостереження. Середня невизначеність, виміряна стандартним відхиленням

імовірностей, була найбільшою у MC Dropout – 0,042365, тоді як для Neural-Laplace вона становила 0,007471. Величина сама по собі не визначає якість. Важливіше, чи змінюється вона систематично при переході від правильного до помилкового прогнозу.

Саме тут спостерігалася найбільш виражена різниця між методами. Для Neural-Laplace середня невизначеність була 0,007130 у правильних прогнозах і 0,008441 у помилкових. Площа під ROC-кривою розпізнавання помилок становила 0,625381, а PR-AUC – 0,351454. Для HMC ці значення дорівнювали 0,616135 і 0,346816. Для глибокого ансамблю – 0,556264 та 0,314051. Для MC Dropout ROC-AUC становила лише 0,446303. Отже, сама наявність великого розкиду прогнозів не гарантує, що невизначеність буде корисною для пошуку помилок; важливо, наскільки цей розкид структурований відносно правильності рішення.

Для Neural-Laplace тест Манна–Уїтні дав $p = 3{,}08 \cdot 10^{-83}$, а рангово-бісеріальний ефект становив 0,250762 з 95 % інтервалом [0,226405; 0,274353]. Тест Манна–Уїтні виявив статистично значущу відмінність між розподілами оцінки невизначеності для правильних та помилкових прогнозів. Водночас статистична значущість цієї відмінності не означає, що невизначеність безпомилково визначає правильність прогнозу; для оцінювання практичної придатності такого ранжування додатково використано ROC-AUC та PR-AUC виявлення помилок.

Розподіл оцінки невизначеності для правильних і помилкових прогнозів наведено на Рис. 2.

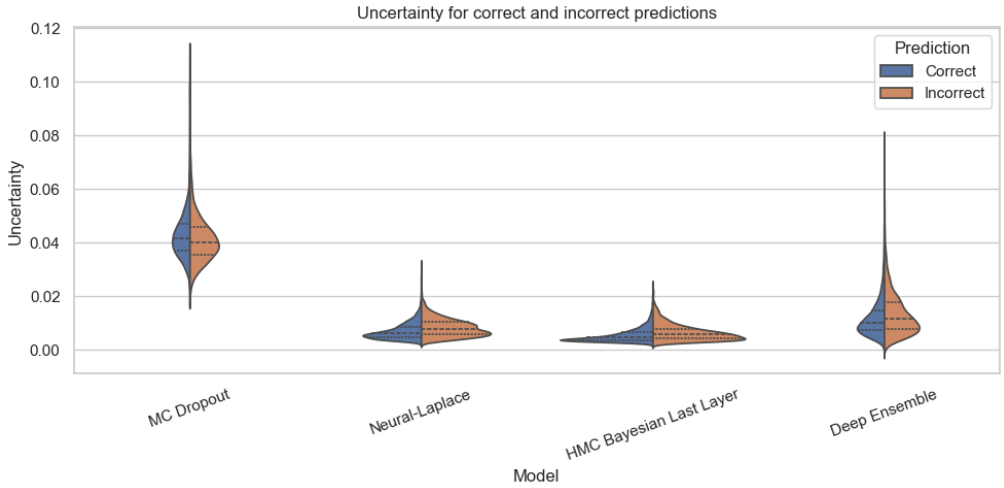


*Рис. 2.* **Розподіл оцінки невизначеності для правильних і помилкових прогнозів**
***Джерело:*** **розроблено авторами**

Як видно з Рис. 2, для Neural-Laplace, HMC та Deep Ensemble розподіли невизначеності для помилкових прогнозів зміщені у бік більших значень порівняно з правильними прогнозами. Для MC Dropout така відмінність виражена слабше, що узгоджується з його низькою здатністю до виявлення помилок за ROC-AUC.

Безпосереднє порівняння здатності невизначеності ранжувати помилкові прогнози наведено на Рис. 3.

На Рис. 3 крива Neural-Laplace має найвищу площу під ROC-кривою серед досліджених методів, тоді як для MC Dropout значення AUC нижче 0,5. Отже, корисність оцінки невизначеності визначається не її абсолютною величиною, а тим, наскільки вона пов'язана зі статусом правильності прогнозу.

Оцінку невизначеності використано як основу для селективного прогнозування:

$$g_\tau(x) = 1\{U(x) \le \tau\}. \tag{13}$$

Параметр $\tau$ визначає максимально допустимий рівень оціненої невизначеності для автоматичного прогнозування; спостереження з $U(x) \le \tau$ передаються на автоматичну класифікацію, а випадки з $U(x) > \tau$ – до режиму відмови від автоматичного рішення.

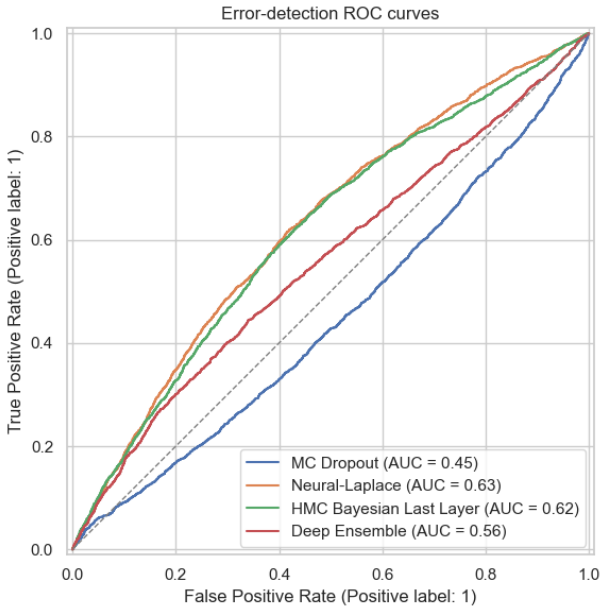


*Рис. 3.* **Здатність невизначеності виявляти помилкові прогнози**
***Джерело:*** **розроблено авторами**

Поріг τ змінювали в межах діапазону отриманих оцінок невизначеності, що давало змогу формувати криву залежності селективного ризику від охоплення. Для ілюстрації практичного режиму використано область охоплення близько 50 %.

Нехай $g_\tau(x) = 1$ означає, що спостереження передається системою на автоматичну класифікацію, а $g_\tau(x) = 0$ – що система відмовляється від автоматичного висновку.

Тоді охоплення визначається як:

$$C(\tau) = E[g_\tau(X)], \tag{14}$$

а селективний ризик – як:

$$R(\tau) = \frac{E[L(f(X),Y)g_\tau(X)]}{C(\tau)}. \tag{15}$$

Практично це означає, що підвищення вимог до впевненості прогнозу через зменшення τ зменшує охоплення, але дозволяє передавати найбільш невизначені випадки на додатковий розгляд фахівця. Для Neural-Laplace при охопленні приблизно 50 % ризик зменшувався до 0,186553, тоді як на повній тестовій вибірці становив 0,260130. Відповідно, селективна точність для решти випадків зростала до 81,3447 %. Площа під кривою «ризик–охоплення» становила 0,184833; менше значення відповідає кращому компромісу між охопленням і ризиком. Отже, механізм відмови може бути використаний як проміжний режим роботи СППР: система автоматично обробляє випадки з нижчою оціненою невизначеністю, а складні випадки передає на додатковий розгляд.

Подібна картина спостерігалася і для HMC, але його AURC дорівнювала 0,191272. Для глибокого ансамблю – 0,226440, для MC Dropout – 0,280662. Таким чином, Neural-Laplace не тільки мала найкраще в межах цієї групи виявлення помилок за невизначеністю, а й давала найкраще узгодження невизначеності з механізмом відмови від автоматичного рішення.

Графічне порівняння зміни ризику залежно від частки автоматично оброблених спостережень наведено на Рис. 4.

Рис. 4 показує, що для Neural-Laplace та HMC зменшення охоплення за рахунок вилучення найбільш невпевнених випадків супроводжується меншим ризиком на залишеній для автоматичної обробки підмножині. Найменша площа під кривою «ризик–охоплення» для Neural-Laplace означає кращий компроміс між охопленням і ризиком у межах цього експерименту.

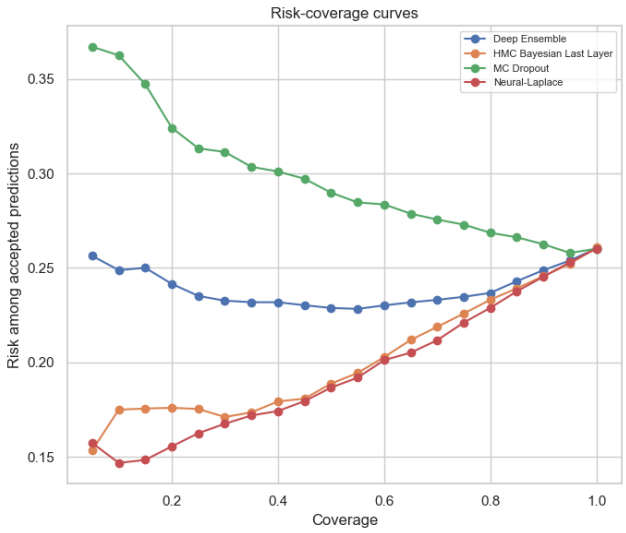


*Рис. 4.* **Криві «ризик–охоплення» для стохастичних моделей**
***Джерело:*** **розроблено авторами**

Щоб дослідити поведінку моделей при різному обсязі навчальних даних, виконано допоміжний експеримент із чотирма обсягами навчальних даних: 10 %, 25 %, 50 % та 100 %. Валідаційна і тестова частини залишалися незмінними, а для кожного обсягу використовувалося три початкові значення генератора випадкових чисел. Отримані значення характеризують поведінку моделей у межах проведених повторень і не трактуються як оцінка повної статистичної стійкості моделей до вибору навчальної підвибірки. Це не замінює повноцінного дослідження навчальних кривих, але дозволяє оцінити напрям зміни характеристик при дефіциті навчальних прикладів.

При використанні 10 % навчальних даних середній ROC-AUC MLP становив 0,782896, тоді як для Neural-Laplace – 0,792630, а для HMC – 0,792506. Водночас різниця за якістю ймовірностей була значно виразнішою: ECE MLP дорівнювало 0,139020, тоді як Neural-Laplace – 0,018656, HMC – 0,022253. Для Brier Score відповідні значення становили 0,216766, 0,185335 та 0,185495. За 25 % даних ECE дорівнювало 0,029755 для MLP і 0,016135 для Neural-Laplace; за 50 % – 0,020392 і 0,017398; за 100 % – 0,022440 і 0,015125.

Отже, головний ефект обмеженої вибірки проявлявся не тільки в точності. У допоміжному експерименті за меншого обсягу навчальних даних Neural-Laplace зберігала нижчі значення ECE порівняно з MLP; найбільша різниця спостерігалася при використанні 10% навчальної вибірки. Саме це підтримує обережну версію основної гіпотези: байєсівський підхід не дає автоматичного приросту дискримінаційної здатності на повному наборі, але в цьому допоміжному експерименті Neural-Laplace демонструвала нижчі значення показників помилки ймовірнісного прогнозування, зокрема ECE, за меншого обсягу навчальних даних.

Зміну ROC-AUC, Brier Score та ECE залежно від частки доступної навчальної вибірки наведено на Рис. 5.

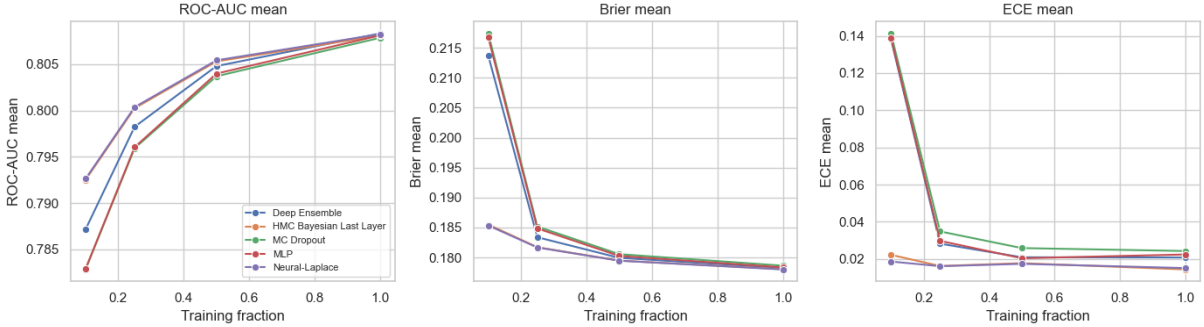


*Рис. 5.* **Залежність основних показників від частки навчальної вибірки**
***Джерело:*** **розроблено авторами**

Рис. 5 показує, що при зменшенні обсягу навчальної вибірки найсильніше погіршувалося калібрування MLP, тоді як Neural-Laplace зберігала нижчі значення ECE на всіх досліджених рівнях. При цьому за 100 % навчальної вибірки відмінності за ROC-AUC між цими двома моделями були мінімальними.

Для оцінювання пояснюваності використано перестановочну важливість ознак. Для кожної ознаки визначалося, наскільки змінюється ROC-AUC після випадкового переставлення її значень. Такий показник оцінює внесок ознаки в прогностичну функцію моделі, але не встановлює причинно-наслідкового зв'язку.

Найвищу середню перестановкову важливість у трьох нейромережевих варіантах мав систолічний артеріальний тиск, далі за рангом розташовувалися вік, діастолічний артеріальний тиск, рівень холестерину та пульсовий тиск. Такий порядок узгоджується з предметною областю: артеріальний тиск, вік та інші метаболічні й поведінкові характеристики належать до ключових чинників, що враховуються під час оцінювання серцево-судинного ризику [15]. Для MLP та Neural-Laplace коефіцієнт рангової кореляції Спірмена між порядками важливості становив 0,989011 ($p = 1{,}75 \cdot 10^{-10}$), для MLP та HMC – 0,994505 ($p = 3{,}91 \cdot 10^{-12}$). Отже, перехід від детермінованого прогнозу до байєсівської оцінки останнього шару не призвів до суттєвої зміни ранжування ознак за перестановковою важливістю.

Середню перестановкову важливість основних ознак для MLP наведено на Рис. 6.

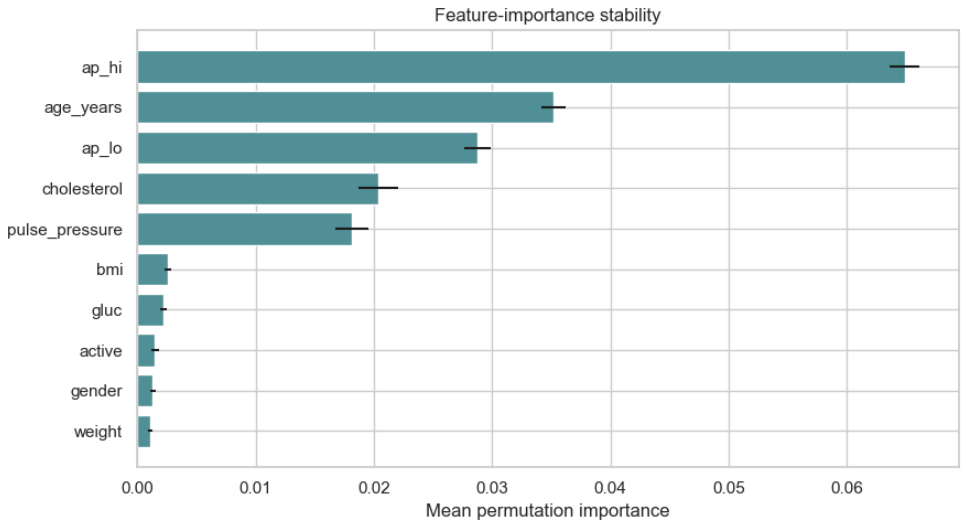


*Рис. 6.* **Середня перестановкова важливість основних ознак для MLP за повторними оцінюваннями**

***Джерело:*** **розроблено авторами**

На Рис. 6 видно домінування систолічного артеріального тиску та віку за перестановковою важливістю. Стабільність рангів ознак між моделями оцінено окремо за коефіцієнтом рангової кореляції Спірмена.

Одержані результати показують, що вплив байєсівської постановки проявляється передусім не у зміні дискримінаційної здатності, а у властивостях ймовірнісного прогнозування та оцінки невизначеності. Найбільш виражено це проявилося для Neural-Laplace, яка мала найнижче ECE та найкращу здатність ранжувати помилкові прогнози за оцінкою невизначеності. За зменшення обсягу навчальних даних відмінності між MLP і Neural-Laplace за калібруванням ставали більш вираженими. Для систем підтримки прийняття рішень це означає, що оцінка невизначеності може бути використана не лише як характеристика моделі, а й як основа правила відмови від автоматичного рішення.

Результати не дають підстав стверджувати, що одна байєсівська модель повинна завжди переважати всі інші. Глибокий ансамбль був кращим за ROC-AUC та Brier Score, Neural-Laplace – за ECE та виявленням помилок за невизначеністю, HMC показав близькі

властивості, а MC Dropout у цьому наборі не забезпечив такої якості виявлення помилок. Тому коректний висновок має стосуватися не перемоги конкретного алгоритму, а межі, за якою додаткова ймовірнісна структура змінює профіль властивостей системи.

Основним обмеженням є використання одного відкритого набору даних та відсутність зовнішньої валідації на незалежних медичних даних, що обмежує оцінювання переносимості отриманих результатів на інші популяції та умови збору даних. Допоміжний експеримент із різним обсягом навчальної вибірки не замінює повноцінних навчальних кривих і незалежної перевірки. Тому отримані результати слід розглядати як експериментальні та методичні, а не як доказ клінічної ефективності або готовності моделі до безпосереднього медичного застосування.

**Висновки.** Проведене дослідження показало, що для системи прогнозування серцево-судинного захворювання якість моделі доцільно розглядати як сукупність взаємопов'язаних характеристик, а не як одну метрику класифікації. У межах єдиного протоколу на 68 606 очищених спостереженнях було порівняно вісім моделей, причому чотири нейромережеві варіанти використовували спільну основу.

Байєсівські підходи не забезпечили статистично підтвердженого приросту дискримінаційної здатності на повній тестовій вибірці. Найвище ROC-AUC та найнижче значення Brier Score отримав глибокий ансамбль, тоді як Neural-Laplace мала найнижче ECE та найкращу здатність виявляти помилкові прогнози за оцінкою невизначеності. Основний результат дослідження полягає у встановленні того, що в межах використаного набору даних та експериментального протоколу додавання байєсівської складової не забезпечує автоматичного покращення дискримінаційної здатності, однак може змінювати профіль інших властивостей прогнозної системи – насамперед калібрування, корисність оцінки невизначеності та поведінку механізму селективного прогнозування.

Одним із найбільш виражених результатів стала здатність Neural-Laplace використовувати оцінку невизначеності для виявлення помилкових прогнозів та організації селективного прогнозування. За охоплення близько половини тестових випадків ризик на автоматично оброблюваній частині вибірки зменшувався порівняно з повним режимом.

Експеримент зі зменшенням навчальної вибірки показав, що при використанні 10% доступних навчальних даних різниця між MLP і Neural-Laplace була особливо вираженою за ECE та Brier Score. При цьому отримані результати слід трактувати як емпіричну тенденцію в межах використаного протоколу, оскільки експеримент із чотирма обсягами вибірки не замінює повноцінного дослідження навчальних кривих. Отже, основна гіпотеза підтверджується частково: байєсівське моделювання не забезпечило універсального підвищення класифікаційної здатності, але для Neural-Laplace було виявлено корисні властивості щодо калібрування та оцінки невизначеності, особливо за обмеженого обсягу навчальних даних.

Науковий результат роботи полягає у встановленні характеру впливу ймовірнісної та байєсівської складових на різні властивості прогнозної системи в межах контрольованої експериментальної схеми. Показано, що за близької дискримінаційної здатності моделей істотні відмінності проявляються насамперед у калібруванні, оцінці невизначеності та селективному прогнозуванні; зокрема, Neural-Laplace мала найнижче ECE та найкращу здатність виявляти помилкові прогнози за невизначеністю. Встановлено, що відсутність статистично підтвердженої переваги байєсівських нейромережевих варіантів за ROC-AUC не виключає їхніх переваг за окремими характеристиками прогнозної системи. У межах проведених експериментів Neural-Laplace мала найнижче ECE та найкращу здатність використовувати оцінку невизначеності для виявлення помилок, а при зменшенні обсягу навчальної вибірки її перевага за якістю ймовірнісного прогнозування порівняно з MLP ставала більш вираженою.

Практична цінність полягає у можливості використання запропонованої схеми оцінювання під час проєктування СППР, у яких автоматичний прогноз доповнюється

оцінкою невизначеності та правилом відмови від автоматичного рішення для випадків із підвищеним ризиком помилки. Остаточна придатність таких систем потребує додаткової перевірки на незалежних медичних наборах даних та в реальних сценаріях застосування.

**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:** Вихідний набір даних є відкритим і розміщений у Kaggle [13]. Експериментальна логіка, параметри моделей, результати та сформовані рисунки зафіксовані в інтерактивному ноутбуці [14]. 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. Бобко, Б. В. та Жуков, С. О. «Методи машинного навчання для виявлення аномалій у медичній статистиці України: аналітичний огляд». *Таврійський науковий вісник. Серія: Технічні науки*. 2025; 1(4): 402–413. DOI: https://doi.org/10.32782/tnv-tech.2025.4.1.38.

2. Жуков, С. О. та Рудзевич, О. В. «Оптимізація глибоких нейронних мереж для класифікації емоційного стану мовлення з використанням динамічного квантування». *Наукові праці Вінницького національного технічного університету.* 2025; 2: 51–65. DOI: https://doi.org/10.31649/2307-5376-2025-2-51-65.

3. Sambyal, A. S., Niyaz, U., Krishnan, N. C. & Bathula, D. R. "Understanding calibration of deep neural networks for medical image classification". *Computer Methods and Programs in Biomedicine*. 2023; 242: 107816, https://www.scopus.com/pages/publications/85172467578. DOI: https://doi.org/10.1016/j.cmpb.2023.107816.

4. Bórquez, S., Pezoa, R., Salinas, L., Torres, C. E., et al. "Uncertainty estimation in the classification of histopathological images with HER2 over expression using Monte Carlo Dropout". *Biomedical Signal Processing and Control*. 2023; 85: 104864, https://www.scopus.com/pages/publications/85150433865.
DOI: https://doi.org/10.1016/j.bspc.2023.104864.

5. Heremans, E. R. M., Seedat, N., Buyse, B., Testelmans, D., van der Schaar, M. & De Vos, M. "U-PASS: An uncertainty-guided deep learning pipeline for automated sleep staging". *Computers in Biology and Medicine*. 2024; 171: 108205, https://www.scopus.com/pages/publications/85185552159
DOI: https://doi.org/10.1016/j.compbiomed.2024.108205.

6. Cinquin, T., Pförtner, M., Fortuin, V., Hennig, P. & Bamler, R. "FSP-Laplace: Function-Space Priors for the Laplace Approximation in Bayesian Deep Learning". *Advances in Neural Information Processing Systems*. 2024. p. 13897–13926, https://www.scopus.com/pages/publications/105000515282. DOI: https://doi.org/10.52202/079017-0446.

7. Langmore, I., Dikovsky, M., Geraedts, S., Norgaard, P. & von Behren, R. "Hamiltonian Monte Carlo in inverse problems: Ill-conditioning and multimodality". *International Journal for Uncertainty Quantification*. 2023; 13 (1): 69–93, https://www.scopus.com/pages/publications/85141715377. DOI: https://doi.org/10.1615/Int.J.UncertaintyQuantification.2022038478.

8. Swaminathan, A., Lopez, I., Wang, W., et al. "Selective prediction for extracting unstructured clinical data". *Journal of the American Medical Informatics Association*. 2024; 31 (1): 188–197, https://www.scopus.com/pages/publications/85181178508.
DOI: https://doi.org/10.1093/jamia/ocad182.

9. Steffny, L., Dahlem, N., Reichl, L., et al. "Design of a Human-in-the-Loop centered AI-based clinical decision support system for professional care planning". *Ebook: HHAI: Augmenting Human Intellect*. 2023; 368: 263 – 273. DOI: https://doi.org/10.3233/FAIA230090.

10. Angelopoulos, A. N. & Bates, S. A. “Gentle introduction to conformal prediction and distribution-free uncertainty quantification”. *Foundations and Trends in Machine Learning*. 2023; 16 (4): 494–591. DOI: https://doi.org/10.1561/2200000094.

11. Ovadia, Y., Fertig, E., Ren, J. et al. “Can you trust your model’s uncertainty? Evaluating predictive uncertainty under dataset shift”. *Advances in Neural Information Processing Systems*. 2019; 32: 13991–14002. DOI: https://doi.org/10.48550/arXiv.1906.02530.

12. Banerji, C., Chakraborty, T., Harbron, C. & MacArthur, B. “Clinical AI tools must convey predictive uncertainty for each individual patient”. *Nature Medicine*. 2023; 29 (12), https://www.scopus.com/pages/publications/ 85173820234. DOI: https://doi.org/10.1038/s41591-023-02562-7.

13. “Cardiovascular Disease Dataset”. *Kaggle Datasets*. 2021. – Available from: https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset. – [Accessed: 06. 2026].

14. “Cardiovascular Bayesian Neural Comparison”. *Kaggle Notebooks*. 2026. – Available from: https://www.kaggle.com/code/serhiizhukov/cardiovascular-bayesian-neural-comparison. – [Accessed: 06. 2026].

15. Visseren, F. L. J., Mach, F., Smulders, Y. M., et al. “ESC Guidelines on cardiovascular disease prevention in clinical practice”. *European Heart Journal*. 2021: 42 (34): 3227–3337. DOI: https://doi.org/10.1093/eurheartj/ehab484.

Отримано 28.07.2026
Отримано після перегляду 29.08.2026
Прийнято 03.09.2026

**DOI: https://doi.org/10.15276/ict.03.2026.27**
**UDC 004.85:004.032.26:616.1**

# Comparative evaluation of bayesian and neural network models for cardiovascular disease prediction considering calibration and uncertainty

**Artem O. Ponomaryov**[1)]
ORCID: https://orcid.org/0009-0001-2112-6759; artponexcelsior@gmail.com
**Oleksii M. Kozachko**[1)]
ORCID: https://orcid.org/0000-0003-3033-9745; lekoz80@gmail.com
[1)] Vinnytsia National Technical University, 95, Khmelnytske Shose. Vinnytsia, 21021, Ukraine

## ABSTRACT

This paper investigates the problem of reliability in probabilistic forecasting of cardiovascular diseases within decision support systems. Unlike the conventional approach, where model quality is primarily evaluated using a single classification metric, this study considers a comprehensive set of interrelated properties of the predictive system: discriminative ability, alignment of predicted probabilities with actual frequencies, uncertainty estimation, error detection capability, and the possibility of rejecting automated decisions for complex cases. The experiments were conducted on the open-source Cardiovascular Disease Dataset. Following conservative data cleaning, over sixty-eight thousand observations were analyzed; twenty-one input variables were generated from the initial features after encoding and scaling. A unified experimental framework was used to compare a Naive Bayes classifier, a Bayesian Network, Logistic Regression, a Multilayer Perceptron, Monte Carlo Dropout, Last-Layer Laplace Approximation (Neural-Laplace), a Hamiltonian Monte Carlo Bayesian Last Layer, and a Deep Ensemble. The results demonstrate that on the full test set, the differences among models in their ability to rank observations are small: the Deep Ensemble achieved the highest area under the receiver operating characteristic curve, whereas the Neural-Laplace model exhibited the lowest expected calibration error among the studied models. For models yielding predictive distributions, uncertainty estimation detected erroneous predictions better than random ranking; the highest area under the receiver operating characteristic curve was obtained by the Neural-Laplace model. Selective prediction revealed that excluding cases with the highest estimated uncertainty from automated processing reduces the risk on the remaining subset. The experiment involving the reduction of the training sample size showed that when using only ten percent of the training data, the discrepancy in calibration metrics between the Neural-Laplace and the Multilayer Perceptron models was most pronounced: the expected calibration error of the Neural-Laplace model was over seven times lower. The scientific result consists in establishing, within a unified experimental framework, that the Bayesian component does not automatically increase discriminative ability but can alter the predictive system's quality profile regarding calibration and uncertainty: the Neural-Laplace

model demonstrated the lowest expected calibration error and the best ability to detect erroneous predictions based on uncertainty estimation. The practical value lies in utilizing such a performance profile to develop selective prediction mechanisms in decision support systems.

**Keywords:** cardiovascular diseases; systems analysis; risk prediction; Bayesian modelling; neural networks; probability calibration; predictive uncertainty; selective prediction

## ІНФОРМАЦІЯ ПРО АВТОРІВ

**Пономарьов Артем Олегович** - аспірант кафедри Системного аналізу та інформаційних технологій. Вінницький національний технічний університет,Хмельницьке шосе, 95. Вінниця, 21021, Україна
***Галузь досліджень*:** системний аналіз

**Artem O. Ponomaryov** - PhD Student, Department of System Analysis and Information Technologies. Vinnytsia National Technical University, 95, Khmelnytske Shose. Vinnytsia, 21021, Ukraine
ORCID: https://orcid.org/0009-0001-2112-6759; artponexcelsior@gmail.com
***Research field:*** system analysis

**Козачко Олексій Миколайович** - кандидат технічних наук, доцент кафедри Системного аналізу та інформаційних технологій. Вінницький національний технічний університет, Хмельницьке шосе, 95. Вінниця, 21021, Україна
***Галузь досліджень*:** системний аналіз, інформаційні технології

**Oleksii M. Kozachko** - 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/0000-0003-3033-9745; lekoz80@gmail.com
***Research field:*** system analysis