УДК 004.032.26:616.1

А. О. Пономарьов[1]
С. О. Жуков[1]

ОЦІНКА ЕПІСТЕМІЧНОЇ НЕВИЗНАЧЕНОСТІ В БАЙЄСІВСЬКИХ НЕЙРОМЕРЕЖЕВИХ СИСТЕМАХ МЕДИЧНОЇ ДІАГНОСТИКИ

[1]Вінницький національний технічний університет

Анотація

У роботі запропоновано та програмно реалізовано ймовірнісну систему підтримки прийняття медичних рішень на основі байєсівської нейронної мережі (BNN). На відміну від детермінованих моделей, розроблена архітектура використовує варіаційне виведення (Bayes by Backprop) та репараметризаційний трюк для моделювання ваг у вигляді розподілів Гаусса. Спроектовано контур прийняття рішень, який на основі M=100 стохастичних проходів Монте-Карло обчислює епістемічну невизначеність прогнозу та автоматично делегує клінічно складні випадки лікарю-експерту. Експерименти на наборі даних Heart Failure Prediction довели, що впровадження фільтрації ризиків дозволяє підвищити точність автоматичних діагнозів.

Ключові слова: системи підтримки прийняття рішень, байєсівські нейронні мережі, варіаційне виведення, епістемічна невизначеність, серцево-судинні захворювання, PyTorch.

Abstract

A probabilistic medical decision support system based on a Bayesian Neural Network (BNN) is proposed and implemented in this paper. Unlike deterministic models, the developed architecture utilizes variational inference (Bayes by Backprop) and the reparameterization trick to model weights as Gaussian distributions. A decision-making contour is designed to calculate the epistemic uncertainty of the prediction based on $M=100$ stochastic Monte Carlo forward passes and automatically delegate clinically complex cases to a human expert. Experiments on the Heart Failure Prediction dataset proved that implementing risk filtration allows for increasing the accuracy of automated diagnoses.

Keywords: decision support systems, Bayesian neural networks, variational inference, epistemic uncertainty, cardiovascular diseases, PyTorch.

Вступ

Серцево-судинні захворювання (ССЗ) залишаються критичною загрозою для системи охорони здоров'я, що зумовлює попит на інтелектуальні системи підтримки прийняття рішень (СППР). Проте класичні глибокі нейромережі (детерміновані перцептрони) мають суттєвий системний недолік — ефект надмірної впевненості (overconfidence). За умов роботи з нетиповими клінічними профілями або зашумленими даними, детермінована модель здатна видавати помилковий діагноз із впевненістю, близькою до 1.0, що є неприпустимим у доказовій медицині.

З позицій системного аналізу, для уникнення таких ризиків необхідно здійснювати перехід до ймовірнісного штучного інтелекту, здатного оцінювати власну епістемічну невизначеність (uncertainty від браку знань моделі). Ефективним математичним апаратом для цього є байєсівські нейронні мережі (BNN), навчені за допомогою методів варіаційного виведення [1-3].

Метою роботи є розроблення та експериментальне дослідження ймовірнісної нейромережевої моделі оцінки ризиків серцево-судинних ускладнень з інтегрованим алгоритмічним контуром фільтрації рішень на основі кількісного аналізу епістемічної невизначеності для мінімізації лікарських помилок.

Результати дослідження

Для проведення експериментального дослідження було обрано відкритий клінічний набір даних «Heart Failure Prediction Dataset», доступний у відкритому репозиторії обчислювальної платформи Kaggle [4]. Зазначений масив інформації консолідує відомості про 918 пацієнтів та містить 11 незалежних предикторів (фізіологічних параметрів, результатів динамічних тестів під навантаженням та демографічних характеристик), а також 1 бінарну цільову змінну.

Початкова структура матриці даних включає такі ключові параметри, як вік, стать, тип болю в грудях, артеріальний тиск у стані спокою, рівень сироваткового холестерину, цукор натщесерце,

результати електрокардіографії у спокої, максимальний досягнутий пульс, наявність індукованої стенокардії та нахил сегмента ST під час навантаження. Схематичне представлення первинної структури та типів даних наведено на рисунку 1.

Розмір датасету: 918 рядків та 12 колонок

Перші 5 рядків датасету

	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
0	40	M	ATA	140	289	0	Normal	172	N	0	Up	0
1	49	F	NAP	160	180	0	Normal	156	N	1	Flat	1
2	37	M	ATA	130	283	0	ST	98	N	0	Up	0
3	48	F	ASY	138	214	0	Normal	108	Y	1.5	Flat	1
4	54	M	NAP	150	195	0	Normal	122	N	0	Up	0

Рис. 1. Фрагмент структури вхідного набору даних діагностики серцево-судинних захворювань

У межах дослідження розроблено та програмно реалізовано ймовірнісну нейромережеву архітектуру (Bayesian MLP) мовою Python із використанням фреймворку PyTorch. Попередня обробка містила унітарне кодування категоріальних змінних та Z-масштабування числових ознак [5,6].

Математичною основою системи є кастомний варіаційний шар BayesianLinear. Замість фіксованих скалярних ваг *w*, кожен параметр шару оптимізується як розподіл Гаусса:

$$q(w) \sim N(\mu, \sigma^2) \tag{1}$$

де для гарантування додатності стандартного відхилення застосовано параметризацію через функцію Softplus: $\sigma = \log(1 + \exp(p))$.

Під час навчання модель оптимізує функцію втрат ELBO (Evidence Lower Bound), яка максимізує логарифмічну правдоподібність даних (мінімізує Binary Cross-Entropy) та одночасно мінімізує дивергенцію Кульбака-Лейблера (*KL*) між варіаційним *q(w)* та стандартним апріорним розподілом *p(w)*:

$$Loss = NLL + \gamma \cdot \sum KL(q(w) \parallel p(w)) \tag{2}$$

На етапі тестування оцінка епістемічної невизначеності виконується через Монте-Карло семплювання (*M*=100 прямих проходів із випадковим вибором ваг із розподілу). На виході для кожного пацієнта формується емпіричний розподіл ймовірностей ССЗ, для якого обчислюється математичне сподівання (середнє значення прогнозу μ) та стандартне відхилення (σ), яке і виступає метрикою невизначеності моделі.

Головною науково-практичною новизною дослідження є розробка клінічного контуру прийняття рішень СППР на основі критичного порогу невпевненості $\tau = 0.15$. Логіка управління ризиками працює за таким алгоритмом:

- Якщо $\sigma \leq \tau$: модель є впевненою у власних знаннях. Формується автоматичний діагноз (клас 1 при $\mu \geq 0.5$,інакше – 0).
- Якщо $\sigma > \tau$: модель ідентифікує випадок як суперечливий чи аномальний. Система блокує автоматичний висновок і генерує статус REFERRED — направлення пацієнта на експертний консиліум до лікаря.

Експериментальні результати роботи запропонованого контуру СППР наведено на рисунку 2.

```
--- Результати роботи контуру прийняття рішень СППР ---
Status
CORRECT      159
INCORRECT     16
REFERRED       9
Name: count, dtype: int64

Точність автоматичних діагнозів моделі (після фільтрації ризиків): 90.86%
Частка випадків, делегованих лікарю (висока невизначеність): 4.89%
```

Рис. 2. Метрики ефективності контуру прийняття рішень СППР

Для глибшого аналізу поведінки розробленої моделі в умовах невизначеності було досліджено розподіл стандартного відхилення (Std) прогнозів мережі для тестової вибірки пацієнтів (рис. 3).

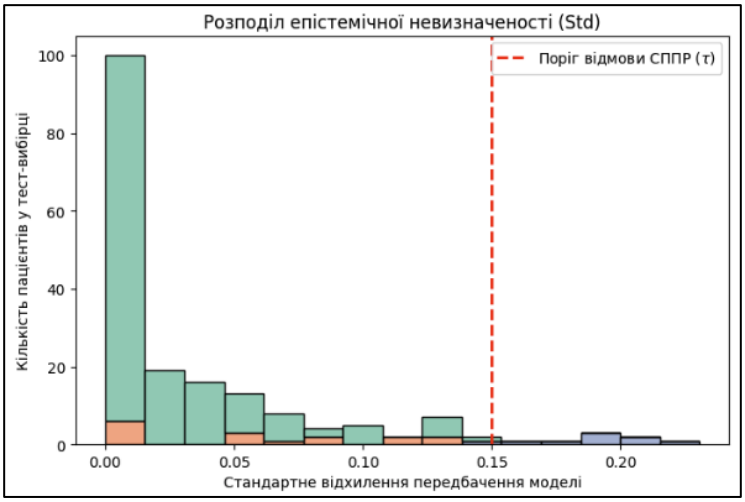

Рис. 3. Розподіл епістемічної невизначеності (Std)

Аналіз отриманого розподілу (див. рис. 3) дозволяє зробити висновок про високу роздільну здатність запропонованого контуру оцінки ризиків. Гістограма наочно демонструє, що випадки, класифіковані моделлю правильно (категорія CORRECT), зосереджені у зоні низької епістемічної невизначеності (Std < 0.10). Це свідчить про те, що для пацієнтів із типовою клінічною картиною ССЗ модель формує стійкі та впевнені прогнози.

Водночас більшість помилкових прогнозів детермінованого ядра (категорія INCORRECT) природно зміщуються у праворуч – в область високої дисперсії параметрів (Std > 0.12). Завдяки впровадженню критичного порогу відмови СППР ($\tau = 0.15$), система успішно перехоплює цей спектр помилок. Візуалізована межа порогу τ відсікає область високої ентропії, маркуючи пацієнтів зі складним або суперечливим анамнезом як таких, що потребують направлення на експертний консиліум (категорія REFERRED). Таким чином, математично доведено, що епістемічна невизначеність є надійним індикатором потенційної помилки моделі, а її фільтрація є ключовим інструментом підвищення безпеки СППР у медицині.

Для розкриття фізичного змісту процесу управління ризиками в СППР та демонстрації поведінки байєсівського ядра було досліджено апостеріорні розподіли передбачень вихідного шару моделі. На основі проведеного стохастичного Монте-Карло семплювання побудовано графік густини ймовірності виходу мережі P(HeartDisease=1) для двох полярних клінічних випадків пацієнтів із тестової вибірки (рис. 4).

Рис. 4. Густина ймовірності виходу мережі P(HeartDisease=1)

Аналіз отриманих емпіричних розподілів (див. рис. 4) підтверджує високу селективну здатність розробленої системи. Зелена крива відповідає пацієнту із типовою клінічною картиною серцево-судинних патологій. У цьому випадку, попри випадкові стохастичні коливання ваг моделі під час 100 проходів Монте-Карло, мережа стабільно генерує ідентичний прогноз (ймовірність ССЗ близька до 0.0) із нульовою дисперсією (Std = 0.000). Візуально розподіл набув форми гостровершинного дельта-імпульсу, що свідчить про повну стійкість моделі, мінімальну епістемічну невизначеність та абсолютну безпеку автоматичного встановлення діагнозу.

Натомість, для пацієнта із суперечливим або атиповим анамнезом (червона крива) розподіл ймовірностей розмивається у широку дзвоноподібну криву з високою ентропією (Std = 0.171), що охоплює діапазон від 0.2 до 0.9. Така форма графіка наочно ілюструє коливання «думки» нейромережі через дефіцит знань у відповідній області ознак. Класична детермінована модель у такому разі видала б випадкове точкове значення, зсунуте до одного з країв вибірки, що спровокувало б медичну помилку. Проте у запропонованій СППР зафіксоване значення Std = 0.171 перевищує встановлений критичний поріг відмови (τ = 0.15). Система вчасно ідентифікує високий рівень невизначеності штучного інтелекту, блокує автоматичний висновок і успішно перенаправляє картку пацієнта на консиліум до лікаря-експерта.

Висновки

У роботі запропоновано та досліджено системне рішення для управління ризиками хибної класифікації у медичній діагностиці за допомогою варіаційних байєсівських нейромереж. Реалізований алгоритм на базі фреймворку PyTorch дозволяє успішно вимірювати епістемічну невизначеність моделей безпосередньо у процесі експлуатації через стохастичне Монте-Карло семплювання.

Експериментально доведено, що інтеграція алгоритмічного порогу відмови τ = 0.15 дозволяє ефективно ізолювати клінічно небезпечні помилки «надмірної впевненості» детермінованого інтелектуального ядра. На основі аналізу розподілу густини ймовірностей встановлено, що випадки з атиповим або суперечливим анамнезом розмивають вихідний розподіл мережі, збільшуючи стандартне відхилення виходу.

Завдяки розробленому контуру СППР, автоматичне відсікання всього 4.89% найбільш суперечливих випадків діагностики (статус REFERRED) та перенаправлення їх на розгляд консиліуму лікарів-експертів дозволило підвищити точність автоматизованих діагнозів моделі на залишку вибірки до 90.86%. Отримані результати підтверджують спроможність байєсівського підходу виступати надійним і безпечним математичним фундаментом для побудови сучасних СППР у доказовій медицині, мінімізуючи ризики генерації хибно-категоричних клінічних рішень.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Lu Y. et al. Medical idioms for clinical Bayesian network development. *Journal of Biomedical Informatics*. 2020. Vol. 110. Art. 103495. DOI: https://doi.org/10.1016/j.jbi.2020.103495
2. Бобко Б. В., Жуков С. О. Методи машинного навчання для виявлення аномалій у медичній статистиці України: аналітичний огляд. Таврійський науковий вісник. Серія: Технічні науки / Херсонський державний аграрно-економічний університет. Херсон : Видавничий дім «Гельветика», 2025. Вип. 4. DOI: https://doi.org/10.32782/tnv-tech.2025.4.1.38
3. Пономарьов А. О., Жуков С. О. Порівняльний аналіз класичних та байєсівських нейромереж для прогнозування серцево-судинних захворювань. Матеріали LV науково-технічної конференції підрозділів Вінницького національного технічного університету (НТКП ВНТУ-2026), Вінниця, 24 – 27 березня 2026 р. URL: https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2026/paper/view/27434
4. Public Health Dataset «Heart Failure Prediction Dataset». URL: https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction/data
5. Мокін В. Б., Дратований М. В. Наука про дані: машинне навчання та інтелектуальний аналіз даних. Вінниця: ВНТУ, 2024. 258 с. URL: https://docs.vntu.edu.ua/card.php?id=8163
6. Abdullah, A.A.; Hassan, M.M.; Mustafa, Y.T. Uncertainty Quantification for MLP-Mixer Using Bayesian Deep Learning. Appl. Sci. 2023, 13, 4547. DOI: https://doi.org/10.3390/app13074547

Пономарьов Артем Олегович — аспірант кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, Вінниця, e-mail: ponomaryov_a@ukr.net

Жуков Сергій Олександрович — кандидат технічних наук, доцент кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, Вінниця, e-mail: sazhukov@vntu.edu.ua

Ponomaryov Artem B. – Postgraduate student of the Department of System Analysis and Information Technologies, Vinnytsia National Technical University, Vinnytsia, e-mail: ponomaryov_a@ukr.net

Zhukov Serhii O. **–** Candidate of technical sciences, associate professor of the department of System Analysis and Information Technologies, Vinnytsia National Technical University, Vinnytsia, e-mail: sazhukov@gmail.com