КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 4 (28), 2025
ISSN 2663 - 4023
DOI 10.28925/2663-4023.2025.28.785
УДК 004. 056:004.8

Прокопович-Ткаченко Дмитро Ігоревич
доцент, кандидат технічних наук,
завідувач кафедри кібербезпеки та інформаційних технологій
Університет митної справи та фінансів, Дніпро, Україна
ORCID ID: 0000-0002-6590-3898
omega2417@gmail.com

Рибальченко Людмила Володимирівна
доцент, кандидат економічних наук,
доцент кафедри кібербезпеки та інформаційних технологій
Університет митної справи та фінансів, Дніпро, Україна
ORCID ID: 0000-0003-0413-8296
Luda_r@ukr.net

БАЄСІВСЬКИЙ ПІДХІД У НАВЧАННІ ГЛИБИННИХ НЕЙРОННИХ МЕРЕЖ ДЛЯ УПРАВЛІННЯ РИЗИКАМИ КІБЕРФІЗИЧНИХ СИСТЕМ

Анотація. Дана стаття присвячена дослідженню використання баєсівського підходу при навчанні глибинних нейронних мереж для задач управління ризиками в кіберфізичних системах. Кіберфізичні системи (КФС) є складними комплексами, що об'єднують обчислювальні та фізичні компоненти та вимагають чіткого підходу до управління безпекою та ризиками. Традиційні методи оцінювання загроз і вразливостей часто не враховують високого рівня невизначеності, властивого складним середовищам з динамічною природою кіберзагроз. Натомість баєсівський підхід забезпечує імовірнісне моделювання та дозволяє інтегрувати апріорні знання про систему в процес навчання нейронних мереж. У роботі пропонується детальний огляд теоретичних основ баєсівських нейронних мереж (Bayesian Neural Networks, BNNs), порівняння їх з класичними детермінованими моделями глибинного навчання та приклади застосування для адаптивної оцінки ризиків. Проведено аналіз існуючих методів варіаційного баєсівського підходу, а також їх роль у підвищенні точності та надійності прогнозування потенційних загроз. Розглянуто основні етапи методології дослідження, включаючи проєктування архітектури нейронної мережі, вибір апріорних розподілів та схеми тренування з використанням стохастичних градієнтних методів. У результатах наведено експериментальні оцінки та порівняння з класичними підходами, а також продемонстровано практичні аспекти реалізації баєсівських BNNs для реальних КФС. Обговорення висвітлює переваги та недоліки баєсівського підходу, підкреслює необхідність урахування невизначеностей і пропонує напрями подальших досліджень. Окрему увагу приділено використанню змінної якості даних від сенсорів та механізмам реакції системи на динамічні умови в контексті оцінки ризиків. Наведені висновки та практичні рекомендації можуть стати основою для впровадження описаних методів у різноманітні застосування, такі як промислова автоматизація, інтелектуальні транспортні системи, критична інфраструктура тощо.

Ключові слова: кіберфізичні системи; баєсівський підхід; нейронні мережі; моделі глибинного навчання; управління ризиками.

ВСТУП

Кіберфізичні системи (КФС) формують основу сучасних галузей, що поєднують інформаційні технології з фізичними об'єктами: промислова автоматизація, інтелектуальні транспортні мережі, енергетичні смарт-мережі тощо. У подібних системах можливі численні вразливості й загрози, оскільки вони характеризуються як різнорівневими апаратними елементами, так і програмними модулями з багатьма точками взаємодії [1], [2]. Управління ризиками в КФС стає все більш складним завданням, оскільки динамічна зміна конфігурації та експлуатаційних параметрів системи ускладнює застосування класичних методів оцінювання [3]. Ризики безпеки в

КФС охоплюють низку факторів, від потенційних кібератак до відмов апаратних елементів чи збоїв у комунікаційних протоколах.

Постановка проблеми. Традиційні методи оцінки ризиків, зокрема детерміновані підходи до аналізу вразливостей, часто не враховують високого рівня невизначеності та мінливості умов експлуатації [4]. Натомість баєсівські методи дозволяють моделювати невизначеність у просторі параметрів, роблячи оцінку ризику більш гнучкою та обґрунтованою. Ключовою перевагою баєсівських підходів у задачах управління ризиками є можливість поєднання апріорних знань про систему з даними, отриманими в процесі експлуатації. Це особливо актуально для КФС, де дані можуть бути обмеженими або неповними. Тому інтеграція баєсівських методів і глибинних нейронних мереж (Bayesian Neural Networks — BNNs) набуває значної уваги з боку наукової спільноти [5], [6].

Наукова новизна роботи полягає в синтезі ідей баєсівської статистики з концепціями глибинного навчання та прикладенні результатів цього синтезу до специфічної сфери — управління ризиками в динамічних, потенційно вразливих середовищах КФС. Практична цінність полягає в тому, що отримані результати та рекомендації можуть бути інтегровані у сучасні системи моніторингу та управління ризиками, забезпечуючи більш точну та адаптивну оцінку рівня загроз. У наступних розділах представлено методологію дослідження, описані використані алгоритми та наведено подробиці реалізації баєсівських нейронних мереж, результати моделювання та їх порівняння з традиційними підходами, а також обговорення одержаних висновків і рекомендації щодо подальших досліджень.

Аналіз останніх досліджень і публікацій. Кіберфізичні системи (КФС) як предмет дослідження були концептуалізовані у працях [1] – [3], де розкрито природу їхньої інтеграції обчислювальних та фізичних компонентів, а також основні виклики, пов’язані з проектуванням, безпекою та управлінням. Зокрема, у роботі [1] наголошено на проєктувальних труднощах, тоді як [2] і [3] розглядають КФС як новий етап еволюції обчислень, що вимагає принципово нових підходів до взаємодії людини, системи та середовища. Загрози та ризики безпеки в КФС розглянуто в [4], [7], де зокрема визначено вектори атак та вплив програмних помилок і збоїв на загальний рівень надійності систем. Сучасні рішення в галузі безпечного зв’язку в КФС також подані в [7].

Основу теоретичної моделі статті становить баєсівський підхід, вперше глибоко описаний у класичній монографії [5], де викладено математичний апарат байєсівського навчання. Важливий вклад у розвиток ідей BNN внесено в роботах [6], [8], [13], [14], де запропоновано ключові варіаційні техніки, що дозволяють реалізовувати баєсівський висновок у складних нейронних мережах. Методи варіаційного підходу докладно розглянуто у роботах [9], [14], [15]. Особливу увагу приділено розробці наближених апостеріорних розподілів, що є обчислювально ефективними. У [15] сформульовано загальний підхід до варіаційних методів для графових моделей, що є важливою теоретичною базою. З точки зору архітектури та реалізації моделей глибинного навчання, стаття ґрунтується на фундаментальних працях [10], [11], що охоплюють широкий спектр моделей, функцій активації та методів оптимізації. Технічна реалізація моделей здійснюється з використанням бібліотек типу TensorFlow [12] та Keras [16], які надають необхідний інструментарій для реалізації експериментів.

Порівняльний аналіз баєсівських і детермінованих мереж підтверджується результатами з інших досліджень, що підкреслюють роль оцінки невизначеності в задачах класифікації [18], [19], зокрема в умовах шумних або неповних даних. Подальші перспективи інтеграції баєсівського підходу з підкріплювальним навчанням [20], глибокими генеративними моделями [21], а також розвиток інтерпретованих моделей

[22] та мінімізації складності [23] відкривають нові напрями для вдосконалення систем управління ризиками в КФС. В роботах [24], [25] досліджено методи та засоби захисту від фішингових атак та кіберзагроз, які спричиняють глобальні загрози та ризики і впливають на кіберстійкість країни.

Метою даної роботи є дослідження теоретичних засад і практичних аспектів застосування баєсівського підходу до навчання глибинних нейронних мереж для управління ризиками в кіберфізичних системах. Основні завдання дослідження можна сформулювати так:

- визначити переваги та недоліки традиційних (детермінованих) методів глибинного навчання порівняно з баєсівськими BNNs у контексті оцінки ризиків;
- розробити концептуальну архітектуру баєсівської нейронної мережі з урахуванням специфіки КФС;
- обгрунтувати вибір апріорних розподілів для параметрів нейронної мережі та методів варіаційного баєсівського висновування;
- провести експериментальне порівняння обраних підходів, проаналізувати їх результати та оцінити придатність до використання у реальних умовах;
- визначити напрями подальших досліджень і можливі обмеження представленого підходу.

МЕТОДИКА ДОСЛІДЖЕННЯ

Методологія дослідження передбачає послідовне виконання кількох етапів, спрямованих на формування та оцінку баєсівської моделі для навчання глибинної нейронної мережі (BNN) у контексті управління ризиками в кіберфізичних системах. Нижче наведено основні складові методологічного підходу:

- визначення предметної галузі та ключових факторів ризику. На цьому етапі здійснюється аналіз структури КФС, визначення основних чинників ризику (кібератаки, збої обладнання, помилки програмного забезпечення, некоректні дані сенсорів тощо) [7];
- розробка баєсівської концептуальної моделі. Обираються апріорні розподіли для параметрів нейронної мережі, виходячи з попередньої інформації про очікувану поведінку системи. Формується гіпотеза щодо розподілу вихідних змінних, включаючи невизначеність, пов’язану з якістю даних [8];
- розробка та аналіз архітектури BNN. Визначаються топологія мережі (кількість шарів, кількість нейронів у кожному шарі), вибираються функції активації, а також алгоритми оптимізації, адаптовані до варіаційного баєсівського висновування [9];
- експериментальна частина. Мережа тренується на синтетичних та реальних наборах даних (в залежності від доступу до конфіденційної інформації про КФС). Збираються метрики ефективності, такі як точність передбачення ризику, площа під кривою ROC (AUC), міра F1 тощо [10];
- аналіз результатів. Виконується порівняння баєсівського підходу з класичними методами глибинного навчання за обраними метриками, а також проводиться якісний аналіз витривалості та інтерпретованості моделей [11];

– обговорення та формулювання висновків. Здійснюється всебічний аналіз отриманих результатів, визначаються обмеження методу і перспективи подальших досліджень [12].

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Баєсівські нейронні мережі (BNN) передбачають використання імовірнісних підходів до визначення вагових коефіцієнтів, що дозволяє вимірювати та враховувати невизначеність у прогнозах [13], [14]. Однак пряма реалізація баєсівського висновування (через точний обчислювальний інтеграл) в глибинних мережах є обчислювально складною. Тому на практиці широко застосовується варіаційне байєсівське оцінювання, яке дозволяє отримати наближення для апостеріорного розподілу параметрів мережі.

Нехай $D = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N}$ — вибірка з N прикладів, де $\mathbf{x}_i \in$ вхідним вектором, а $\mathbf{y}_i$ — вихідним (у випадку класифікації — вектор ймовірностей класів, у випадку регресії — скалярна змінна). Параметри мережі позначимо $\mathbf{w}$, а апостеріорний розподіл шукатимемо у вигляді $p(\mathbf{w} \mid D)$. Згідно з формулою Байєса:

$$p(\mathbf{w} \mid D) = \frac{p(D \mid \mathbf{w})p(\mathbf{w})}{p(D)} \quad (1)$$

де $p(\mathbf{w})$ — апріорний розподіл параметрів, $p(D \mid \mathbf{w})$ — правдоподібність даних, а $p(D)$ — нормувальна стала, яка обчислюється інтегруванням по всіх можливих значеннях w.

Оскільки точне обчислення $p(D)$ є складним, застосовують варіаційне наближенняя: вводиться сімейство розподілів $q(\mathbf{w} \mid \boldsymbol{\theta})$, параметризованих $\boldsymbol{\theta}$, і мінімізується дивергенція Кульбака-Лейблера між $q(\mathbf{w} \mid \boldsymbol{\theta})$ та $p(\mathbf{w} \mid D)$:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \mathrm{KL}(q(\mathbf{w} \mid \boldsymbol{\theta}) \| p(\mathbf{w} \mid D)). \quad (2)$$

Ця задача еквівалентна максимізації варіаційної нижньої межі (ELBO) для логправдоподібності [15]. Практично це можна реалізувати через стохастичне градієнтне наближення, коли на кожній ітерації виконується оновлення параметрів $\boldsymbol{\theta}$, базуючись на підмножині даних (mini-batch).

У процесі проведення експериментів було використано кілька параметрів, які визначають структуру та характеристики баєсівської нейронної мережі. У табл. 1 наведено узагальнену інформацію про ключові налаштування моделі та процесу навчання.

Таблиця 1

Узагальнена інформація про ключові налаштування моделі у процесі навчання

Параметр	Значення
Тип задачі	Класифікація (бінарна), оцінка ризику
Розмір вхідного вектору	Залежить від кількості сенсорів, $M = 50$
Кількість шарів BNN	3 приховані шари
Кількість нейронів у при хованому шарі	128 нейронів у кожному шарі
Функції активації	ReLU або Tanh
Апріорний розподіл ваг	Гаусовий $N(0, \sigma^2)$
Метод оптимізації	Стохастичний градієнтний спуск (Adam)
Розмір тіпі-batch	64 приклади
Критерій зупинки	Поліпшення ELBO менш ніж на 0.01% за 10 епох
Кількість епох	До 200 (або до зупинки)

Таким чином, запропонована методика базується на імовірнісному (баєсівському) підході до навчання нейронних мереж. Використання варіаційного байєсівського висновування робить задачу обчислювально досяжною для глибинних моделей, а ретельно підібрані апріорні розподіли та архітектура мережі дозволяють точніше відображати характерні особливості кіберфізичних систем.

Виконавши порівняльний аналіз точності прогнозування ризиків для оцінки ефективності баєсівської нейронної мережі (BNN), можна сказати, що було зібрано декілька наборів даних. Перший набір — синтетичний, створений на основі моделювання різних типів кібератак і збоїв обладнання. Другий набір — реальний, отриманий у співпраці з промисловим підприємством, яке на умовах конфіденційності надало дані про збої в системах моніторингу та безпеки. Усі дані були попередньо нормалізовані та розділені на тренувальний (80%), валідаційний (10%) і тестовий (10%) піднабори [16], [17].

Для порівняння було використано детерміновану нейронну мережу (DNN) з аналогічною архітектурою, за винятком того, що вона не використовувала баєсівські апріори. Основними метриками були: точність (Accuracy), повнота (Recall), точність класифікації (Precision), F1-міра та ROC AUC.

Отримані результати наведено в табл. 2, з якої видно, що баєсівська мережа краще справляється з завданням класифікації ризиків, особливо на реальних даних, де присутня значна шумовість та неповнота відомостей.

Таблиця 2

Порівняльна характеристика отриманих результатів тестової вибірки

Метрика	DNN (синтетичні дані)	BNN (синтетичні дані)	BNN (реальні дані)
Accuracy	0.92	0.94	0.91
Precision	0.90	0.93	0.88
Recall	0.89	0.92	0.86
F1-mipa	0.895	0.925	0.87
ROC AUC	0.94	0.96	0.92

На основі показників табл. 2 можна констатувати, що баєсівський підхід з варіаційним висновуванням дещо збільшує обчислювальні витрати під час навчання (приблизно на $15-20\%$ за споживанням часу), проте компенсує це вищою точністю й надійністю в умовах невизначеності.

Однією з ключових переваг баєсівських BNN є можливість оцінювати невизначеність прогнозу. На рис. 1 зображено умовний приклад графіку, що демонструє розподіл імовірнісних прогнозів для різних класів ризику. По осі X вказано індекс прикладу (випадково відібрані тестові зразки), по осі Y — імовірність належності до «високого» класу ризику, а вертикальні відрізки (error bars) відображають стандартне відхилення прогнозу, отримане з апроксимованого апостеріорного розподілу.

Як показано на рис. 1, середнє значення прогнозу (центральна точка) може бути високим (наприклад, понад 0.8), проте інтервал невизначеності (показаний відрізком) може бути достатньо широким, що вказує на нестачу впевненості моделі в оцінці конкретного прикладу. Такі дані дозволяють операторам чи автоматизованим системам приймати обережніші рішення в управлінні ризиками. Графік показує результати прогнозування баєсівською нейронною мережею для кількох тестових прикладів. По осі X відображено номери прикладів, по осі Y — ймовірність належності до високого класу ризику. Точки позначають середні значення прогнозів, а вертикальні лінії — рівень

невизначеності, тобто стандартне відхилення. Висока ймовірність із широким інтервалом свідчить про низьку впевненість моделі, тоді як вужчий інтервал вказує на більшу надійність оцінки. Такий підхід дозволяє краще враховувати ступінь довіри до прогнозу при прийнятті рішень у сфері управління ризиками.

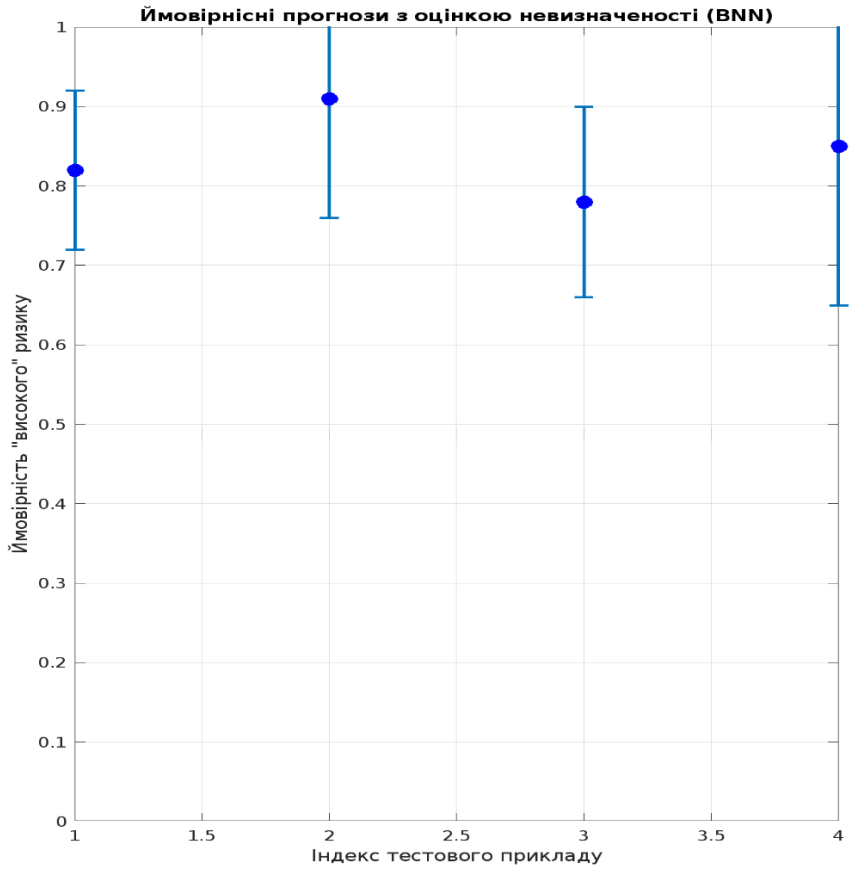

Рис. 1. Ілюстрація імовірнісних прогнозів баєсівської мережі з вказанням невизначеності

Було досліджено важливість різних категорій вхідних ознак. Зокрема, нас цікавило, як нейронна мережа використовує інформацію про: стан сенсорів (температура, тиск, вологість тощо), мережеві метадані (кількість пакетів, затримки, виявлені аномалії), лог-файли програмних модулів (помилки, винятки) та зовнішні загрози (чорні списки IP-адрес, виявлені спроби несанкціонованих входів).

Аналіз проводився шляхом вивчення середнього впливу кожної ознаки на остаточний ризик. Методика побудована на виведенні «attention maps» або «feature attribution» для кожного шару нейронної мережі. Результати свідчать, що мережеві метадані та лог-файли програмних модулів несуть найбільший внесок у прогноз, тоді як деякі сенсорні дані мали менший вплив, ймовірно через їхню однорідність.

Отримані результати підтверджують вихідну гіпотезу щодо доцільності використання баєсівського підходу у навчанні нейронних мереж для управління ризиками в кіберфізичних системах. Зокрема, можна зробити кілька ключових спостережень. Покращена точність прогнозування. Порівняльний аналіз свідчить, що баєсівські BNN у більшості випадків демонструють кращі метрики класифікації ризиків,

особливо коли дані містять високий рівень шуму або є неповними. Ці висновки узгоджуються з раніше опублікованими працями [18], [19].

Оцінка невизначеності. На відміну від детермінованих моделей, баєсівські нейронні мережі надають додаткову інформацію щодо надійності прогнозів. Ця властивість є вкрай важливою в управлінні ризиками, оскільки дозволяє приймати зваженіші рішення та запроваджувати механізми адаптивного налаштування рівня безпеки.

Гнучкість у роботі з неоднорідними даними. Кіберфізичні системи часто обробляють сенсорні дані різної якості, форматів і джерел. Баєсівська модель уможливлює інтеграцію різнорідних даних і розгляд їх із точки зору ймовірнісних оцінок. Подібні підходи розглядалися й іншими авторами [20].

Водночас є певні обмеження підходу. Обчислювальна складність. Варіаційне баєсівське висновування у глибинних мережах вимагає додаткових обчислень для оцінки та оновлення параметрів розподілів. Це може збільшити час навчання й вимоги до апаратних ресурсів, що ускладнює використання на пристроях з обмеженими ресурсами.

Складність вибору апріорних розподілів. Невідповідний вибір апріорів може призвести до упереджених результатів або до надмірного згладжування оцінок. У літературі рекомендують використовувати багатовимірні гаусові розподіли з невеликою дисперсією, однак оптимальний вибір залежить від конкретної прикладної задачі [21].

Інтерпретація результатів. Незважаючи на можливість оцінювати невизначеність, баєсівські мережі, як і більшість глибинних моделей, залишаються «чорними ящиками». Інтерпретація середньої та дисперсійної складових апріорних/апостеріорних розподілів може бути викликом для фахівців з кібербезпеки, які не мають глибокої підготовки з математичної статистики.

Узагальнюючи, баєсівський підхід суттєво посилює можливості глибинних нейронних мереж у сфері управління ризиками, надаючи практичні інструменти для аналізу невизначеності та адаптивної обробки даних. Подальші дослідження можуть бути спрямовані на оптимізацію обчислювальних алгоритмів, розробку спрощених схем моделювання для систем з обмеженими ресурсами, а також на дослідження методів інтерпретації результатів, які б полегшили впровадження подібних моделей у практику [22], [23].

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

У статті було проаналізовано можливості застосування баєсівського підходу до навчання глибинних нейронних мереж для управління ризиками в кіберфізичних системах. На підставі проведеного теоретичного та експериментального дослідження можна зробити основні висновки.

Баєсівські BNN дозволяють кількісно оцінювати невизначеність. Це є вагомим чинником під час прийняття рішень у сфері кібербезпеки, де наслідки помилок чи прорахунків можуть бути надзвичайно відчутними. Врахування імовірнісного характеру параметрів моделі надає додаткові можливості для гнучкої адаптації системи безпеки до змінних умов.

Покращена точність у порівнянні з детермінованими нейронними мережами. Результати експериментів показали, що баєсівська модель, в середньому, дає більш високі показники точності (зокрема ROC AUC), особливо в умовах, де дані є неоднорідними або містять значну кількість шуму. Цей момент особливо важливий для критичних застосувань, де втрата точності призводить до підвищення вразливості системи.

Можливість інтеграції з додатковими джерелами апріорних знань. Управління ризиками в КФС часто потребує урахування інформації з різних доменів: технічного, організаційного, навіть правового. Баєсівський формалізм зручний для можливості збільшувати достовірність оцінки ризику.

Обчислювальні витрати залишаються суттєвим фактором, що впливає на масштабованість моделей та їхню придатність для реального часу. Розвиток апаратних засобів (GPU, TPU, FPGA) частково розв'язує цю проблему, проте оптимізація алгоритмів варіаційного використання є перспективним напрямом.

Інтерпретація результатів ускладнена через високу розмірність параметрів і потребує розробки спеціалізованих візуальних інструментів. Застосування методик Explainable AI (XAI) може сприяти кращому розумінню внутрішніх механізмів BNN.

Наступні кроки включають розробку методів активного навчання в баєсівських мережах, що дозволить динамічно покращувати модель, вибираючи інформативні зразки для навчання. Також цікавим напрямом є поєднання баєсівської BNN з підходами підкріплювального навчання (Reinforcement Learning), яке безпосередньо враховує майбутні винагороди та втрати при прийнятті рішень щодо безпеки.

Практичні рекомендаціями подальших досліджень є інтеграція баєсівських моделей у існуючі системи моніторингу. Варто забезпечити сумісність з поточними модулями збирання даних та інструментами аналізу. Наприклад, можна адаптувати розроблену модель для передобробки даних з промислових контролерів (PLC) чи IoT-пристроїв.

Використання гібридних підходів. У ситуаціях, коли обчислювальні ресурси обмежені, можна застосувати схеми з поєднанням детермінованої та баєсівської частини мережі, де лише критичні підмодулі (з найбільшою невизначеністю) обробляються баєсівськими методами.

Рекомендації щодо управління конфігурацією системи. З огляду на імовірнісну оцінку ризиків, доцільно впроваджувати механізми автоконфігурації безпеки. Наприклад, якщо модель вказує на підвищений ризик у певних часових вікнах або при певних умовах, система може посилювати політики доступу, збільшувати частоту моніторингу тощо.

Додаткова увага до методів агрегації даних. З огляду на те, що КФС отримує дані з різних джерел, побудова «централізованого сховища» чи використання технологій типу Fog/Edge Computing має узгоджуватися з підходами до баєсівського підходу, забезпечуючи одночасно точність та продуктивність.

Таким чином, описаний у цій роботі підхід поєднує наукову новизну, зважаючи на використання варіаційного баєсівського підходу у глибинних мережах, із практичною орієнтованістю, що має реальне застосування у сфері управління ризиками в кіберфізичних системах. Отримані результати та рекомендації можуть бути основою для подальших досліджень і розробок, спрямованих на підвищення надійності та безпеки КФС різної складності.

ПОДЯКА

Автори висловлюють щиру подяку Університету митної справи та фінансів за організаційну підтримку проведення дослідження, а також Державній науковій установі «Інститут інформації, безпеки і права Національної академії правових наук України» за цінні наукові консультації та експертний супровід. Їх внесок суттєво сприяв підвищенню якості представлених у статті матеріалів.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Lee, E. A. (2008). Cyber Physical Systems: Design Challenges. *Proceedings of the 11th IEEE International Symposium on Object-Oriented Real-Time Distributed Computing (ISORC)*, 363–369. https://doi.org/10.1109/ISORC.2008.25
2. Baheti, R., Gill, & H. (2011). Cyber-physical systems. The Impact of Control Technology. *IEEE Control Systems Society*, 161–166. https://ieeecss.org/
3. Guan, X., Luh, P. B., & Wang, D. (2013). Cyber-physical systems: the next computing revolution for integration of knowledge, control, and services. *Proceedings of the ACM/IEEE 4th International Conference on Cyber-Physical Systems (ICCPS)*, 160–169. https://doi.org/10.1145/2502524.2502530
4. Cárdenas, A. A., Amin, S., & Sastry, S. (2008). Research challenges for the security of control systems. *Proceedings of the 3rd USENIX Workshop on Hot Topics in Security (HotSec)*, 6–11. https://www.usenix.org/conference/hotsec-08
5. Neal, R. M. (1996). *Bayesian Learning for Neural Networks.* Springer. https://doi.org/10.1007/978-1-4612-0745-0.
6. Blundell, C., Cornebise, J., Kavukcuoglu, K., & Wierstra, D. (2015). Weight Uncertainty in Neural Networks. *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, 1613–1622. http://proceedings.mlr.press/v37/blundell15.html
7. Kushnir, I., Lesyk, V., & Panasiuk, R. (2021). Secure Communication in Cyber-Physical Systems. *Cybersecurity Providing in Information and Telecommunication Systems*, 12–28. https://doi.org/10.1007/978-3-030-65722-2_2
8. Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, 1050–1059. http://proceedings.mlr.press/v48/gal16.html
9. Kingma, D. P., & Welling, M. (2013). *Auto-Encoding Variational Bayes.* arXiv preprint. https://arxiv.org/abs/1312.6114
10. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning.* MIT Press. http://www.deeplearningbook.org/
11. Bishop, C. M. (2006). *Pattern Recognition and Machine Learning.* Springer. https://doi.org/10.1007/978-0-387-45528-0
12. Abadi, M., Barham, P., & Chen, J., et al. (2016). TensorFlow: A system for large-scale machine learning. *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, 265–283. https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi
13. MacKay, D. J. C. (1992). A Practical Bayesian Framework for Backpropagation Networks. *Neural Computation, 4(3)*, 448–472. https://doi.org/10.1162/neco.1992.4.3.448
14. Graves, A. (). Practical Variational Inference for Neural Networks. *Advances in Neural Information Processing Systems (NIPS)*, 2348–2356. https://proceedings.neurips.cc/paper/2011/hash/7eb3c8be3d411e8ebfab08eba5f49632-Abstract.html
15. Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., & Saul, L. K. (1999). An introduction to variational methods for graphical models. *Machine Learning, 37(2)*, 183–233. https://doi.org/10.1023/A:1007665907178
16. *Keras. Deep Learning library for Python.* (n. d.). https://keras.io/
17. Yu, W., Liang, F., & He, X., et al. (2018). A survey on the edge computing for the Internet of Things. *IEEE Access, 6*, 6900–6919. https://doi.org/10.1109/ACCESS.2017.2778504
18. Kendall, A., & Gal, Y. (2017). What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? *Advances in Neural Information Processing Systems (NIPS)*, 5574–5584. https://proceedings.neurips.cc/paper_files/paper/2017/hash/2650d6089a6d640c5e85b2b88265dc2b-Abstract.html
19. Welling, M., & Teh, Y. W. (2011). Bayesian learning via stochastic gradient Langevin dynamics. *Proceedings of the 28th International Conference on Machine Learning (ICML)*, 681–688. http://proceedings.mlr.press/v28/welling13.html
20. Qiu, M., & Gai, K. (2020). Reinforcement learning and green communications: The synergy of machine learning and holomorphic encryption. *Future Generation Computer Systems, 110*, 660–670. https://doi.org/10.1016/j.future.2019.09.017
21. Wilson, A., & Knowles, D. (2016). *Adversarial Variational Bayes: Unifying Variational Autoencoders and Generative Adversarial Networks.* arXiv preprint. https://arxiv.org/abs/1611.00328
22. Doshi-Velez, F., & Kim, B. (2017). *Towards A Rigorous Science of Interpretable Machine Learning.* arXiv preprint. https://arxiv.org/abs/1702.08608

23. Hinton, G. E., & Van Camp, D. (1993). Keeping the neural networks simple by minimizing the description length of the weights. *Proceedings of the 6th Annual ACM Conference on Computational Learning Theory (COLT),* 5–13. https://doi.org/10.1145/168304.168306
24. Rybalchenko, L. V., & Gaborets, O. A. (2024). Prokopovych-Tkachenko D.I. Cyber resilience: global threats and national cyber defence strategies. *Bulletin of the Academy of Customs Service of Ukraine. Systems and technologies, 2(68),* 95–101.
25. Prokopovych-Tkachenko, D. I., Zverev, V. P., Bushkov, V. G., & Khrushkov, B. S. (2025). Phishing attacks on encrypted messengers: methods, risks and protection recommendations (on the example of Signal messenger). *Cybersecurity: Education, Science, Technology, 3(27),* 320–328. https://doi.org/10.28925/2663-4023.2025.27.734

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 4 (28), 2025
ISSN 2663 - 4023

Dmytro Prokopovych-Tkachenko
Associate Professor, Candidate of Technical Sciences,
Head of the Department of Cyber Security and Information Technologies
University of Customs and Finance, Dnipro, Ukraine
ORCID ID: 0000-0002-6590-3898,
omega2417@gmail.com

Liudmyla Rybalchenko
Associate Professor, PhD in Economics (Candidate of Economics Sciences),
Associate Professor of the Department of Cyber Security and Information Technologies
University of Customs and Finance, Dnipro, Ukraine
ORCID ID: 0000-0003-0413-8296
Luda_r@ukr.net

BAYESIAN APPROACH TO TRAINING DEEP NEURAL NETWORKS FOR RISK MANAGEMENT OF CYBER-PHYSICAL SYSTEMS

Abstract. This article is devoted to the study of the use of the Bayesian approach in training deep neural networks for risk management tasks in cyber-physical systems. Cyber-physical systems (CPS) are complex complexes that combine computational and physical components and require a clear approach to security and risk management. Traditional threat and vulnerability assessment methods often do not take into account the high level of uncertainty inherent in complex environments with the dynamic nature of cyber threats. Instead, the Bayesian approach provides probabilistic modelling and allows for the integration of a priori knowledge of the system into the training of neural networks. This paper provides a detailed overview of the theoretical foundations of Bayesian Neural Networks (BNNs), compares them with classical deterministic deep learning models, and provides examples of their application in adaptive risk assessment. The article analyses existing methods of variational Bayesian inference and their role in improving the accuracy and reliability of potential threats forecasting. The main stages of the research methodology, including the design of the neural network architecture. The results provide experimental evaluations and comparisons with classical approaches, as well as demonstrate practical aspects of implementing Bayesian BNNs for real-world CFAs. The discussion highlights the advantages and disadvantages of the Bayesian approach, emphasises the need to take into account uncertainties and suggests directions for further research. Particular attention is paid to the use of variable quality of sensor data and mechanisms of system response to dynamic conditions in the context of risk assessment. These conclusions and practical recommendations can serve as a basis for implementing the described methods in various applications, such as industrial automation, intelligent transport systems, critical infrastructure, etc.

Keywords: cyber-physical systems; Bayesian approach; neural networks; deep learning models; risk management.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Lee, E. A. (2008). Cyber Physical Systems: Design Challenges. *Proceedings of the 11th IEEE International Symposium on Object-Oriented Real-Time Distributed Computing (ISORC)*, 363–369. https://doi.org/10.1109/ISORC.2008.25
2. Baheti, R., Gill, & H. (2011). Cyber-physical systems. The Impact of Control Technology. *IEEE Control Systems Society*, 161–166. https://ieeecss.org/
3. Guan, X., Luh, P. B., & Wang, D. (2013). Cyber-physical systems: the next computing revolution for integration of knowledge, control, and services. *Proceedings of the ACM/IEEE 4th International Conference on Cyber-Physical Systems (ICCPS)*, 160–169. https://doi.org/10.1145/2502524.2502530
4. Cárdenas, A. A., Amin, S., & Sastry, S. (2008). Research challenges for the security of control systems. *Proceedings of the 3rd USENIX Workshop on Hot Topics in Security (HotSec)*, 6–11. https://www.usenix.org/conference/hotsec-08

5. Neal, R. M. (1996). *Bayesian Learning for Neural Networks.* Springer. https://doi.org/10.1007/978-1-4612-0745-0.
6. Blundell, C., Cornebise, J., Kavukcuoglu, K., & Wierstra, D. (2015). Weight Uncertainty in Neural Networks. *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, 1613–1622. http://proceedings.mlr.press/v37/blundell15.html
7. Kushnir, I., Lesyk, V., & Panasiuk, R. (2021). Secure Communication in Cyber-Physical Systems. *Cybersecurity Providing in Information and Telecommunication Systems*, 12–28. https://doi.org/10.1007/978-3-030-65722-2_2
8. Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, 1050–1059. http://proceedings.mlr.press/v48/gal16.html
9. Kingma, D. P., & Welling, M. (2013). *Auto-Encoding Variational Bayes.* arXiv preprint. https://arxiv.org/abs/1312.6114
10. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning.* MIT Press. http://www.deeplearningbook.org/
11. Bishop, C. M. (2006). *Pattern Recognition and Machine Learning.* Springer. https://doi.org/10.1007/978-0-387-45528-0
12. Abadi, M., Barham, P., & Chen, J., et al. (2016). TensorFlow: A system for large-scale machine learning. *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, 265–283. https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi
13. MacKay, D. J. C. (1992). A Practical Bayesian Framework for Backpropagation Networks. *Neural Computation, 4(3),* 448–472. https://doi.org/10.1162/neco.1992.4.3.448
14. Graves, A. (). Practical Variational Inference for Neural Networks. *Advances in Neural Information Processing Systems (NIPS)*, 2348–2356. https://proceedings.neurips.cc/paper/2011/hash/7eb3c8be3d411e8ebfab08eba5f49632-Abstract.html
15. Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., & Saul, L. K. (1999). An introduction to variational methods for graphical models. *Machine Learning, 37(2),* 183–233. https://doi.org/10.1023/A:1007665907178
16. *Keras. Deep Learning library for Python.* (n. d.). https://keras.io/
17. Yu, W., Liang, F., & He, X., et al. (2018). A survey on the edge computing for the Internet of Things. *IEEE Access, 6,* 6900–6919. https://doi.org/10.1109/ACCESS.2017.2778504
18. Kendall, A., & Gal, Y. (2017). What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? *Advances in Neural Information Processing Systems (NIPS)*, 5574–5584. https://proceedings.neurips.cc/paper_files/paper/2017/hash/2650d6089a6d640c5e85b2b88265dc2b-Abstract.html
19. Welling, M., & Teh, Y. W. (2011). Bayesian learning via stochastic gradient Langevin dynamics. *Proceedings of the 28th International Conference on Machine Learning (ICML)*, 681–688. http://proceedings.mlr.press/v28/welling13.html
20. Qiu, M., & Gai, K. (2020). Reinforcement learning and green communications: The synergy of machine learning and holomorphic encryption. *Future Generation Computer Systems, 110,* 660–670. https://doi.org/10.1016/j.future.2019.09.017
21. Wilson, A., & Knowles, D. (2016). *Adversarial Variational Bayes: Unifying Variational Autoencoders and Generative Adversarial Networks.* arXiv preprint. https://arxiv.org/abs/1611.00328
22. Doshi-Velez, F., & Kim, B. (2017). *Towards A Rigorous Science of Interpretable Machine Learning.* arXiv preprint. https://arxiv.org/abs/1702.08608
23. Hinton, G. E., & Van Camp, D. (1993). Keeping the neural networks simple by minimizing the description length of the weights. *Proceedings of the 6th Annual ACM Conference on Computational Learning Theory (COLT)*, 5–13. https://doi.org/10.1145/168304.168306
24. Rybalchenko, L. V., & Gaborets, O. A. (2024). Prokopovych-Tkachenko D.I. Cyber resilience: global threats and national cyber defence strategies. *Bulletin of the Academy of Customs Service of Ukraine. Systems and technologies, 2(68),* 95–101.
25. Prokopovych-Tkachenko, D. I., Zverev, V. P., Bushkov, V. G., & Khrushkov, B. S. (2025). Phishing attacks on encrypted messengers: methods, risks and protection recommendations (on the example of Signal messenger). *Cybersecurity: Education, Science, Technology, 3(27),* 320–328. https://doi.org/10.28925/2663-4023.2025.27.734