

**Пантелєєв Вадим Олегович**

аспірант кафедри інфокомунікаційної інженерії ім. В. В. Поповського
Харківський національний університет радіоелектроніки, Харків, Україна
ORCID ID: 0009-0008-7824-8782
vadym.pantieliev@nure.ua

Білодід Володимир Григорович

науковий співробітник відділу науки та інновацій
Харківський національний університет Повітряних Сил ім. І. Кожедуба, Харків, Україна
ORCID: 0009-0002-8976-2310
vibos111@gmail.com

**ОПЕРАЦІОНАЛІЗАЦІЯ OSINT-ОЗНАК ТА ГІБРИДНА МОДЕЛЬ ПРОГНОЗУВАННЯ
ІНСАЙДЕРСЬКИХ ІНЦИДЕНТІВ**

Анотація. У статті запропоновано методичний підхід до прогнозування інсайдерських інцидентів. Підхід поєднує операціоналізацію ознак розвідки з відкритих джерел (OSINT) із гібридною моделлю, яка враховує структуру корпоративного графа. Поведінкові передумови внутрішніх інцидентів – невдоволення, фінансовий тиск, зовнішні зв'язки, пошук нового місця роботи – часто з'являються у відкритому цифровому просторі раніше, ніж у внутрішній телеметрії організації. Проте перетворення таких сигналів на валідні, надійні та юридично допустимі ознаки моделі залишається невирішеним завданням. Операціоналізацію подано як ланцюг відображень: латентний конструкт – індикатор – джерело OSINT – вимір із нормалізацією – готова ознака, що об'єднується з внутрішніми показниками доступу й активності. Запропоновано таксономію OSINT-ознак за вісьмома категоріями та чотири критерії їх якості: змістову валідність, надійність вимірювання, часову випереджальність і правомірність збору. Обчислювальний експеримент на синтетичному наборі з двох тисяч профілів показав, що за прийнятих сценарних припущень додавання OSINT-ознак підвищує площу під ROC-кривою з 0,685 до 0,772, поліпшує точність на рідкісному класі та дає помітне випередження сигналу до інциденту. Другу частину дослідження присвячено обмеженням наявних графових підходів. Систематизовано п'ять недоліків; головними з яких є гомофілійне припущення передавання повідомлень, надзгляджування подань на привілейованих вузлах-хабах та недосліджена змагальна стійкість. З їх урахуванням було побудовано гібридну модель. Вона об'єднує OSINT-ознаки, внутрішню телеметрію та структурний сигнал графа комунікацій і доступу до ресурсів, а пропагацію виконує гетерофілійно-стійкими операторами. На гармонізованому за базовими рівнями експерименті структурний сигнал підвищує площу під ROC-кривою до 0,831. На сегменті привілейованих облікових записів глибока гомофілійна згортка деградує, тоді як гетерофілійно-стійкі архітектури утримують якість на рівні 0,80–0,82 і дають найнижчу похибку калібрування. Уперше для цієї задачі кількісно оцінено змагальну стійкість. Маскування відкритого цифрового сліду разом із приховуванням аномальних зв'язків знижує якість прогнозу до рівня самої телеметрії. Запропоноване захисне донавчання з маскованою аутментацією частково відновлює стійкість без істотної втрати чистої якості. У роботі також окреслено етико-правові межі застосування OSINT у кадрових процесах відповідно до засад захисту персональних даних.

Ключові слова: OSINT; інсайдерські загрози; операціоналізація ознак; графові нейронні мережі; події безпеки; інциденти безпеки; прогнозування; захист персональних даних.

ВСТУП

Цифрова присутність співробітників поза корпоративним периметром стрімко зростає. Це змінює саму постановку задачі протидії внутрішнім (інсайдерським) інцидентам. Розвідка з відкритих джерел (OSINT) у поєднанні з машинним навчанням на графах відкриває доступ до випереджальних сигналів ризику, яких не бачать класичні системи моніторингу внутрішньої телеметрії. Щоб скористатися цим



потенціалом, потрібні дві речі: чітка методика перетворення відкритих даних на ознаки моделі та модель, працездатна в умовах гетерофільії й змагального маскуванню. Стаття пропонує таку методику, перевіряє її обчислювальним експериментом і формулює запобіжники відповідального застосування.

Постановка проблеми. Системи протидії внутрішнім загрозам традиційно спираються на внутрішню телеметрію: журнали доступу, активність на робочих станціях, рух даних, кадрові показники. За цим стоїть припущення, що зловмисний намір рано чи пізно проявиться в технічних слідах. Проблема в запізнюванні: коли інсайдер уже копіює чутливі дані, запобігати вже нічому. Галузеві огляди фіксують, що лише близько п'ятої частини організацій вводять у програми виявлення поведінкові та психосоціальні сигнали, хоча саме вони передують технічним діям [1].

Відкриті джерела дають доступ до сигналів, які формуються поза периметром організації: публічних висловлювань, ознак фінансового тиску, зовнішніх зв'язків, оновлень професійних профілів, згадок облікових даних у витоках. Понад половина постачальників OSINT-рішень уже пропонують модулі прогнозу аналітики [2]. Проте відкриті дані надходять неструктурованими, зашумленими та юридично чутливими, тому прямо використовувати їх як ознаки моделі не можна. Невирішеним залишається завдання операціоналізації, тобто строгого переходу від нечітких поведінкових конструктів до вимірюваних, відтворюваних і правомірних ознак. Окремі роботи аналізують сентимент інсайдерів у соціальних медіа [3] та розмови навколо організаційних ризиків [4].

Другий бік проблеми такий. Навіть коректно операціоналізована ознака описує користувача ізольовано. Але інсайдерська діяльність реляційна за природою: вона проявляється в аномальних зв'язках із носіями чутливих ресурсів і в позиції користувача в корпоративному графі. Найбільший ризик створюють привілейовані облікові записи – адміністратори, підрядники, керівники, – які водночас є вузлами-хабами графа. Саме на цьому сегменті класичні графові нейронні мережі втрачають роздільну здатність через надзгладжування подань. Тому пряме перенесення відомих архітектур на задачу прогнозування інсайдерів неефективне.

Аналіз останніх досліджень і публікацій. Прогнозування інсайдерських загроз методами машинного навчання має давню традицію. Ансамблеві класифікатори на даних організаційних журналів типу CERT досягають високої роздільної здатності в передбаченні окремих сценаріїв витоку [5]. Окрема лінія робіт вводить у мультиагентні моделі виявлення людський чинник – довіру, намір, комунікаційний контекст, – що підвищує чутливість за рахунок когнітивного контексту [6]. Українські дослідники розвивають узагальнені моделі прогнозування кібербезпекових аномалій на ансамблях нейромережових компонентів [7] та гібридні стратегії захисту, що поєднують різноманітні джерела сигналів [8].

Паралельно сформувався напрям застосування OSINT у кібербезпеці. Систематичний огляд поєднань штучного інтелекту з відкритою розвідкою показує стрімке зростання кількості таких робіт і водночас брак узгодженої методології відбору ознак [9]. Прикладні фреймворки автоматизують збір пасивних відкритих даних і перетворюють його на оцінки ризику активів градієнтним бустингом та кластеризацією [10]. Український доробок охоплює аналіз програмних засобів реалізації OSINT [11], узагальнення технологій відкритих джерел у сфері безпеки держави та дослідження небезпек, які несуть OSINT-інструменти для самої організації [12]. Окремо вивчають поведінкові та психосоціальні індикатори інсайдера: поєднання психосоціальних даних з аудитом кіберактивності посилює прогнозу здатність моделей [13]. Передові наукові праці застосовували OSINT-аналіз публічної активності для оцінки психологічного стану користувачів [3, 14], а також аналізували організаційний ризик у публічних комунікаціях [4].

Реляційний напрям представлено систематичним оглядом графового виявлення інсайдерів [15]. Огляд обґрунтовує перевагу графового подання інтранету над табличними методами: традиційні класифікатори дають F1-score близько 60 %, тоді як графові моделі краще вловлюють відношення між користувачами, хостами та ресурсами. Запропоновано згорткові моделі на графах журналів [16], дводоменну графову згортку [17] та навчання на графі знань із керуванням оцінкою ризику [18]. Практично всі ці моделі спираються на гомофільне припущення передавання повідомлень, яке для інсайдерів систематично порушується. Для гетерофільних графів розроблено подання з розділенням его-та сусідських ознак [19] і навчання ваги узагальненого PageRank [20]. Проти надзгладжування на глибині застосовують персоналізовану PageRank-пропагацію [21] та початковий резидуальний зв'язок [22]. Показовим є домен графового виявлення шахрайства: аномальні вузли там рідкісні й переважно з'єднані з нормальними, тому графи інгертентно гетерофільні [5], [23], [24].

Змагальний аспект досліджено переважно поза межами інсайдерської проблематики. Показано вразливість графових моделей до цілеспрямованих збурень структури та ознак [25], систематизовано атаки й захисти [26]. Водночас у літературі з прогнозування інсайдерів стійкість до спотворення сигналів самим порушником практично не оцінювали, хоча саме інсайдер контролює власний публічний слід.

Поза увагою наявних робіт залишилося методичне ядро – власне операціоналізація. Дослідження



зосереджуються або на технічних інструментах збору, або на класифікаторах, а готовність ознак приймають як належне. Питання змістової валідності, надійності, часової випереджальності та правомірності OSINT-ознак, як і їх інтеграції з внутрішньою телеметрією, розглянуто фрагментарно. Правові межі та етичні ризики відкритої розвідки, зокрема дотримання вимог захисту персональних даних і GDPR, опрацьовано переважно в контексті протидії інформаційним впливам [27], а не кадрового прогнозування; оглядові праці прямо наголошують на потребі балансу між безпекою та правами особи [28]. Нерозв'язаними залишаються також інтеграція операціоналізованих відкритих ознак зі структурним сигналом графа, збереження якості прогнозу на сегменті привілейованих користувачів та кількісна оцінка змагальної стійкості такої інтеграції.

Мета статті. Мета полягає в розробленні методичного підходу до операціоналізації OSINT-ознак і побудові на його основі гібридної моделі прогнозування інсайдерських інцидентів. Модель має об'єднувати відкриті сигнали, внутрішню телеметрію та структуру корпоративного графа і зберігати працездатність на сегменті привілейованих користувачів за умов змагальної маніпуляції. Поставлено такі завдання:

- формалізувати операціоналізацію як перехід від конструкта до вимірюваної ознаки моделі та визначити критерії її якості;
- побудувати таксономію OSINT-ознак інсайдерського ризику з прив'язкою до джерел і способів вимірювання;
- оцінити на обчислювальному експерименті приріст якості прогнозу від додавання OSINT-ознак до внутрішньої телеметрії;
- систематизувати недоліки наявних графових підходів і сформулювати вимоги до моделі, що їх усуває;
- розробити гібридну модель, яка поєднує операціоналізовані ознаки зі структурним сигналом графа на основі гетерофілійно-стійких операторів пропагації, та оцінити її якість, зокрема на привілейованому сегменті;
- дослідити стійкість моделі до маскуванню відкритого сліду й приховування зв'язків і запропонувати механізм захисту;
- окреслити операційні та етико-правові механізми, що утримують застосування OSINT у межах правомірності й пропорційності.

ТЕОРЕТИЧНІ ОСНОВИ ДОСЛІДЖЕННЯ

Операціоналізація OSINT-ознак

Операціоналізація – це перетворення абстрактного поняття на вимірювану величину через систему індикаторів. Для OSINT подамо її як ланцюг відображень (рис. 1): прихована сутність (латентний конструкт c) розбивається на набір чітких індикаторів $I(c)$, доступних для спостереження; кожному індикатору зіставляють конкретне відкрите джерело s ; для джерела визначають спосіб вимірювання m : $s \rightarrow R$ та правило нормалізації; одержану ознаку $g = z(m)$ подають у модель разом із внутрішньою телеметрією. Наприклад, конструкт «невдоволення» розкладається на індикатори «негативні дописи» й «публічні скарги», джерелом слугує соціальна мережа, виміром – частка негативних дописів за місяць, а нормалізацією – z -оцінка цієї частки.

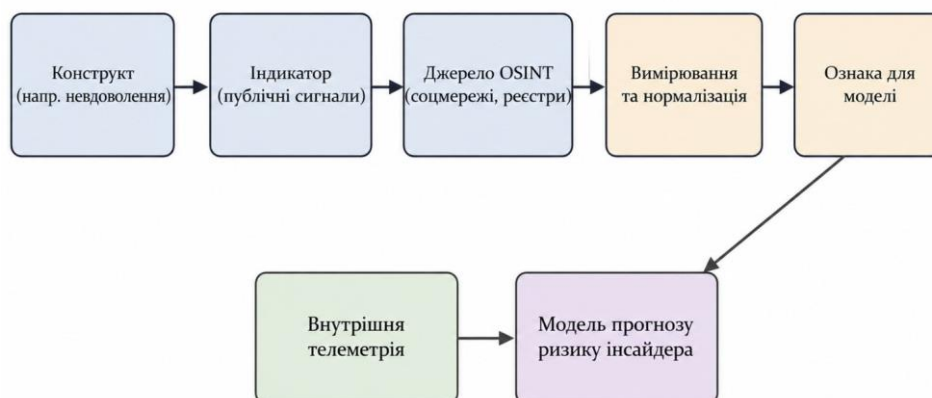


Рис. 1. Послідовність операціоналізації OSINT-ознаки та її інтеграції з внутрішньою телеметрією
Якість операціоналізованої ознаки визначають чотири критерії. Недотримання будь-якого з них



знецінює всю подальшу обробку:

- змістова валідність – ознака вимірює саме той конструкт, для якого призначена, а не випадкову кореляцію з публічною активністю;
- надійність вимірювання – значення ознаки відтворюється в часі та між джерелами й є стійким до шуму й маніпуляцій;
- часова випереджальність – ознака змінюється раніше за внутрішні технічні індикатори (саме це створює прогнозу цінність);
- правомірність збору: обробка відкритих даних відповідає законодавству про захист персональних даних і принципу пропорційності [27].

Систематизацію OSINT-ознак за вісьмома категоріями, які взято в роботі як основні, наведено в табл. 1. Кожній категорії відповідають конструкт, спостережуваний індикатор, типове відкрите джерело та нормалізована ознака моделі.

Числову частину операціоналізації описують два співвідношення. Сирий показник x відкритого джерела приводять до зіставної шкали z -оцінку за формулою:

$$z = \frac{x - \mu}{\sigma}, \quad (1)$$

де μ і σ – середнє та стандартне відхилення показника. Їх оцінюють тільки на навчальному вікні спостережень, що унеможливило витік інформації з тестових даних. У розгортанні μ і σ періодично переоцінюють на кожному історичному вікні, оскільки цифрова поведінка є нестационарною. Підсумкову оцінку ризику p отримують злиттям внутрішніх і OSINT-ознак через логістичне перетворення:

$$p = \sigma(\sum_i w_i f_i + \sum_j w'_j g_j + b), \quad (2)$$

де f_i – нормалізовані внутрішні ознаки, g_j – OSINT-ознаки, w_i та w'_j – їхні ваги, b – зміщення, $\sigma(\cdot)$ – логістична функція, яка відображає оцінку в інтервалі $[0, 1]$ і надає їй ймовірнісне тлумачення.

Таблиця 1

Операціоналізація OSINT-ознак інсайдерського ризику

Категорія	Конструкт та індикатор	Джерело OSINT	Ознака моделі (нормалізація)
Тональність	Невдоволення, негативні висловлювання про роботодавця	Соціальні мережі, форуми	Індекс негативної тональності, z -оцінка
Публічна образа	Відкритий конфлікт, скарги, погрози	Публічні дописи, коментарі	Частота грубих/конфліктних маркерів
Фінансовий стрес	Боргове навантаження, фінансова скрута	Оголошення, публічні реєстри	Бінарна та шкальована ознака тиску
Пошук роботи	Намір змінити місце праці	Професійні мережі, дошки вакансій	Інтенсивність оновлень профілю
Зовнішні зв'язки	Контакти з конкурентами чи ризиковими групами	Відкриті графи зв'язків	Зовнішня центральність вузла
Витік даних	Наявність облікових даних у відкритих витоках	Бази витоків, dark web	Кількість і свіжість збігів
Волатильність сліду	Раптові зміни цифрової поведінки	Сукупність профілів	Міра відхилення від базової лінії
Згадка в dark web	Пропозиція послуг або даних організації	Закриті форуми, моніторинг	Зважений індикатор згадок

Ваги визначають навчанням на розмічених даних із контролем за критеріями якості, описаними вище. Співвідношення (2) відіграє роль інтерпретовної опорної моделі злиття, зручної для пояснення рішень кадровим і безпековим підрозділам. В обчислювальному експерименті лінійну комбінацію замінено нелінійним ансамблем градієнтного бустингу над тим самим об'єднаним вектором ознак; калібрування виходу зберігає ймовірнісне трактування оцінки ризику.

Недоліки наявних графових підходів

В ході аналізу наукових джерел виокремлено п'ять системних недоліків наявних підходів. Недолік Н1: задача виявлення домінує над задачею прогнозування, а відкриті сигнали не інтегровано з телеметрією.



Недолік Н2: передавання повідомлень спирається на гомофільне припущення, тобто вважає, що сусідні вузли мають однакову мітку; для інсайдера це припущення порушується, бо його аномальні зв'язки ведуть саме до «чистих» носіїв ресурсів. Недолік Н3: із зростанням глибини подання вузлів-хабів надзгладжуються й колапсують до спільного середнього, тому якість на привілейованому сегменті падає. Недолік Н4: відсутня оцінка калібрування та контролю хибнопозитивних спрацьовувань за сильного дисбалансу класів. Недолік Н5: не досліджено стійкість моделі до того, що сам інсайдер спотворює власні сигнали. Ці п'ять недоліків і визначають вимоги до моделі, яку побудовано в цій роботі.

МЕТОДИКА ДОСЛІДЖЕННЯ

Гібридна модель із урахуванням структури графа

Організацію подано неорієнтованим графом $G = (V, E)$. Вузли множини V – це облікові записи, а ребра множини E відповідають комунікаціям, спільним ролям і доступу до чутливих ресурсів; $n = |V|$ – кількість вузлів. Матрицю суміжності та похідні від неї матриці степенів задано співвідношеннями:

$$A \in \{0,1\}^{n \times n}, A^T = A, A_{ii} = 0, D_{ii} = \sum_j A_{ij}, \tilde{D} = D + I, \quad (3)$$

де A – матриця суміжності; рівність $A^T = A$ означає симетричність, тобто неорієнтованість графа; умова $A_{ii} = 0$ виключає самопетлі; D – діагональна матриця степенів вузлів; $\tilde{D}$ – матриця степенів графа, доповненого самопетлями; I – одинична матриця порядку n . Вузли нульового степеня виключаються з графової частини моделі, оскільки для них $D^{-1/2}$ не визначена; прогноз для таких вузлів формується лише за їхніми ознаками.

Кожному вузлу i зіставлено вектор внутрішньої телеметрії та вектор операціоналізованих OSINT-ознак. Обидва блоки нормалізовано за формулою (1), тому їх можна зчепити на спільній шкалі згідно з формулою:

$$x_i = [x_i^{\text{tel}} \parallel x_i^{\text{osint}}] \in \mathbb{R}^d, d = d_{\text{tel}} + d_{\text{osint}}, X = [x_1; \dots; x_n] \in \mathbb{R}^{n \times d}, \quad (4)$$

де $\parallel$ – операція зчеплення (конкатенації) векторів; $d_{\text{tel}} = 10$ і $d_{\text{osint}} = 8$ – розмірності телеметричного та OSINT-блоків; $d = 18$ – повна розмірність вектора ознак вузла; X – матриця ознак усіх вузлів, кожен рядок якої відповідає одному вузлу. Наголосимо, що верхні індекси tel та osint є мітками блоків ознак, а не операцією транспонування; транспонування позначаємо прямим символом T , як у формулі (3).

Структурний сигнал уводимо двома нормалізованими операторами суміжності:

$$\hat{S} = \tilde{D}^{-1/2}(A + I)\tilde{D}^{-1/2}, \hat{A} = D^{-1/2}AD^{-1/2}, \quad (5)$$

де $\hat{S}$ – симетрично нормалізований оператор із самопетлями, $\hat{A}$ – той самий оператор без самопетель. Оператор $\hat{S}$ застосовують моделі, яким потрібне згладжування разом із власним сигналом вузла, тобто формули (6), (7) та (11)–(14). Оператор $\hat{A}$ застосовує H2GCN, де самопетлі виключено навмисно: інакше власне подання вузла змішалось б із сусідським і розділення, заради якого будується ця модель, зникло б.

Як базові розглянуто дві згорткові моделі. Перша, GCN-lin, є лінеаризованою двокроковою згорткою без проміжних нелінійностей:

$$H_{\text{GCN-lin}} = \hat{S}^2 X, K = 2. \quad (6)$$

Друга базова модель, GCN-8, є глибокою згорткою з восьми шарів:

$$H^{(l+1)} = \text{ReLU}(\hat{S}H^{(l)}W^{(l)}), H^{(0)} = X, H_{\text{GCN-8}} = H^{(L)}, L = 8, \quad (7)$$

де $W^{(l)}$ – навчена вагова матриця шару l , L – кількість шарів. GCN-8 слугує контрольним прикладом надзгладжування: зі зростанням L подання вузлів-хабів дедалі ближчі до спільного середнього, і роздільна здатність на привілейованому сегменті падає.

Модель H2GCN [19] реалізує три принципи: розділення власного та сусідського подань, використання околів вищих порядків і комбінування проміжних подань. Околи задаються матрицями точних відстаней:



$$(A_r)_{uv} = 1 \Leftrightarrow d_G(u, v) = r, \quad r \in \{1, 2\}, \quad \bar{A}_r = D_r^{-1/2} A_r D_r^{-1/2}, \quad (8)$$

де $d_G(u, v)$ – довжина найкоротшого шляху між вузлами u та v ; A_1 – матриця суміжності першого околу; A_2 – матриця суміжності околу відстані рівно два; D_r – діагональна матриця степенів матриці A_r . Підкреслимо, що $A_2 \neq A^2$, оскільки друга степінь матриці суміжності містить також шляхи, які повертаються у вихідний вузол або ведуть до його найближчих сусідів. Рекурентний крок агрегує обидва околи окремо:

$$H^{(l)} = [\bar{A}_1 H^{(l-1)} \parallel \bar{A}_2 H^{(l-1)}], \quad H^{(0)} = \text{ReLU}(XW_e), \quad (9)$$

де W_e – матриця его-вкладення, а $H^{(0)}$ – власне подання вузла, яке надалі не зміщується із сусіднім. Підсумкове подання комбінує всі проміжні рівні:

$$H_{H2GCN} = [H^{(0)} \parallel H^{(1)} \parallel \dots \parallel H^{(L)}]. \quad (10)$$

Ширина підсумкового подання становить $(2^{(L+1)} - 1)h_e$, де h_e – розмірність его-вкладення, а L – кількість рівнів агрегації.

Модель GPR-GNN [20] замінює фіксоване згладжування навченою комбінацією кроків пропagaції:

$$H_{GPR} = \sum_{k=0}^K \gamma_k \bar{S}^k X, \quad (11)$$

де γ_k – вага k -го кроку пропagaції, K – глибина пропagaції. Ваги γ_k не обмежені за знаком: від’ємне значення γ_k змінює знак внеску відповідного порядку сусідства, що дає змогу моделі «відштовхувати» подання сусіда з протилежною міткою. Саме це робить оператор придатним для гетерофілійних зв’язків інсайдера, коли аномальні ребра ведуть до легітимних вузлів.

Модель APPNP [21] виконує пропagaцію з телепортацією до початкових ознак:

$$H^{(k+1)} = (1 - \alpha) \bar{S} H^{(k)} + \alpha X, \quad H^{(0)} = X, \quad H_{APPNP} = H^{(K)}, \quad (12)$$

де α – коефіцієнт телепортації. Доданок αX повертає на кожному кроці фіксовану частку початкових ознак, тому подання привілейованих хабів не колапсують до спільного середнього навіть за великої кількості кроків K . Навчуваних ваг у самій пропagaції APPNP немає.

Модель GCNII [22] додає до початкового резидуального зв’язку другий механізм – тотожне відображення:

$$H^{(l+1)} = \text{ReLU} \left(\left((1 - \alpha) \bar{S} H^{(l)} + \alpha H^{(0)} \right) \left((1 - \beta_l) I_h + \beta_l W^{(l)} \right) \right). \quad (13)$$

Коефіцієнт β_l спадає з глибиною за логарифмічним законом, а вхідне подання утворюється лінійним відображенням ознак:

$$\beta_l = \ln \left(\frac{\lambda_\beta}{l} + 1 \right), \quad H^{(0)} = \text{ReLU}(XW_0), \quad H_{GCNII} = H^{(L)}, \quad (14)$$

де λ_β – гіперпараметр розкладу коефіцієнтів β_l ; I_h – одинична матриця порядку h ; h – розмірність прихованого подання. За $\beta_l \rightarrow 0$ множник $((1 - \beta_l) I_h + \beta_l W^{(l)})$ вироджується в тотожний, тому глибокі шари не додають ефективних параметрів і не погіршують подання. Саме цим GCNII відрізняється від APPNP: перша модель має навчувані ваги в пропagaції, друга – ні. Позначення λ_β навмисно відмінне від λ у формулі (16): перший параметр керує розкладом коефіцієнтів за глибиною, другий – силою L2-регуляризації.

Згладжене подання подано на двошаровий перцептрон:

$$p_i = \sigma(\text{ReLU}(H_i W_1 + b_1) W_2 + b_2), \quad (15)$$

де H_i – рядок поданої на класифікатор матриці, що відповідає вузлу i ; W_1 , b_1 , W_2 , b_2 – ваги та зміщення шарів, причому $W_1 \in \mathbb{R}^{d' \times h}$, $b_1 \in \mathbb{R}^h$, $W_2 \in \mathbb{R}^{h \times 1}$, $b_2 \in \mathbb{R}$; $\sigma(u) = (1 + e^{-u})^{-1}$ – логістична



функція; p_i – прогнозована ймовірність належності вузла i до класу інсайдерів. Ширина входу d' залежить від обраного оператора пропагації: d для GPR-GNN і APPNP, h для GCNII, $(2^{(L+1)} - 1)h_e$ для H2GCN, $3d$ для GCN-8 за конкатенації шарів. Зауважимо, що σ у формулі (15) позначає логістичну функцію, тоді як σ у формулі (1) – стандартне відхилення; це різні величини, що традиційно поділяють одне позначення.

За сильного дисбалансу класів навчання виконуємо мінімізацію зваженої бінарної крос-ентропії:

$$L = -\frac{1}{n_{\text{train}}} \sum_{i \in V_{\text{train}}} [w_+ y_i \ln p_i + (1 - y_i) \ln(1 - p_i)] + \lambda (\|W_1\|_F^2 + \|W_2\|_F^2), \quad (16)$$

де $y_i \in \{0, 1\}$ – істинна мітка вузла, причому одиниця відповідає інсайдеру; V_{train} – множина навчальних вузлів, $n_{\text{train}} = |V_{\text{train}}|$ – її потужність; λ – коефіцієнт L2-регуляризації; $\|\cdot\|_F$ – норма Фробеніуса. Вагу позитивного класу беремо оберненою до його частоти: $w_+ = \frac{n_-}{n_+}$, де n_+ та n_- – кількості позитивних і негативних прикладів навчальної вибірки. За частки позитивного класу 0,11 це дає $w_+ \approx 8,1$.

Якість калібрування контролюється очікуваною похибкою калібрування:

$$\text{ECE} = \sum_{b=1}^B \frac{|B_b|}{n_{\text{test}}} |\text{acc}(B_b) - \text{conf}(B_b)|, \quad (17)$$

де B – кількість кошиків, на які розбито прогнози за рівнем упевненості; B_b – множина спостережень b -го кошика, $|B_b|$ – її потужність; n_{test} – обсяг тестової вибірки; $\text{acc}(B_b)$ – частка правильних відповідей у кошику b ; $\text{conf}(B_b)$ – середня прогнозована впевненість у тому самому кошику. Поріг рішення визначаємо як $\tau^* = \arg \max_{\tau} F1(\tau)$ на валідаційній вибірці. Додатково контролюємо показник FPR 90 – частку хибнопозитивних спрацювань за фіксованої повноти 0,90.

Інсайдер контролює власний публічний слід і може його маскувати. Тому до навчальної вибірки додаються копії вузлів з обнуленим блоком OSINT-ознак за незмінної структури графа:

$$\tilde{X} = [X^{\text{tel}} \parallel 0_{n \times d_{\text{osint}}}], \quad L_{\text{def}} = L(X) + L(\tilde{X}), \quad (18)$$

де $\tilde{X}$ – аугментована матриця ознак; 0 – нульовий блок розмірності $n \times d_{\text{osint}}$; L_{def} – сумарна втрата на чистих і замаскованих даних. Оскільки ознаки z-нормалізовані, обнулення блоку еквівалентне заміненню кожної OSINT-ознаки її середнім значенням на навчальному вікні. Саме тому аугментація точно відтворює атаку A1 (п. 4.3) і протидіє їй найкраще, тоді як проти комбінованої атаки A3, яка додатково змінє структуру графа, її дія лише часткова.

Відповідність між п'ятьма недоліками з п. 2.2 і засобами їх усунення у запропонованій моделі наведено в табл. 2. Узагальнену схему моделі, що поєднує обидві складові підходу – операціоналізацію відкритих ознак і гетерофільно-стійку графову пропагацію, – наведено на рис. 2.

Організація обчислювального експерименту

Перевірку виконано у два етапи на повністю специфікованих синтетичних даних. Перший етап оцінює додану цінність OSINT-ознак на рівні ознакового простору. Другий етап оцінює внесок структурного сигналу та стійкість моделі.

На першому етапі побудовано набір із двох тисяч профілів користувачів. Кожен профіль описано десятьма внутрішніми показниками телеметрії та вісьмома OSINT-ознаками згідно з табл. 1. Належність до рідкісного класу підвищеного ризику задано незалежними випробуваннями Бернуллі з імовірністю 0,11. Значення кожної ознаки згенеровано як суму трьох складових: стандартного нормального шуму, складової з важкими хвостами (0,25 від розподілу Стюдента з чотирма ступенями вільності) і (для класу ризику) зсуву середнього. Зсув становить 0,22–0,48 стандартного відхилення для внутрішніх ознак і 0,30–0,55 для OSINT-ознак; дві неінформативні внутрішні ознаки залишено без зсуву. Сигнал класу ризику розкладено на спільний латентний фактор і незалежну OSINT-складову з частками 0,6 та 0,4 відповідно, а між ознаками тональності й публічної образи введено корельований шум (0,55).



Недоліки наявних підходів та їх усунення у пропонованій моделі

	Недолік	Джерела	Усунення в моделі
1	Домінування задачі виявлення над прогнозуванням; відкриті сигнали не інтегровано з телеметрією	[3], [4], [15]	Спільна ознакова шкала телеметрії та операціоналізованих OSINT-ознак (формули (1), (4))
2	Гомофільне припущення передавання повідомлень; ігнорування гетерофільії зв'язків інсайдера	[5], [19], [24]	Гетерофільно-стійкі оператори пропагації: H2GCN (8)–(10), GPR-GNN (11)
3	Надзгладжування та деградація якості на привілейованих вузлах-хабах	[21], [22]	Початковий резидуальний зв'язок (12) і тотожне відображення (13), (14); окремий контроль сегмента
4	Відсутність оцінки калібрування та контролю хибнопозитивних за дисбалансу	[24]	Зважена крос-ентропія (16), калібрування порогів, метрики ECE (17) та FPR@90
5	Не досліджено змагальну стійкість до спотворення сигналів інсайдером	[25], [26]	Атаки маскуванню OSINT і приховування зв'язків; захисне донавчання (18)

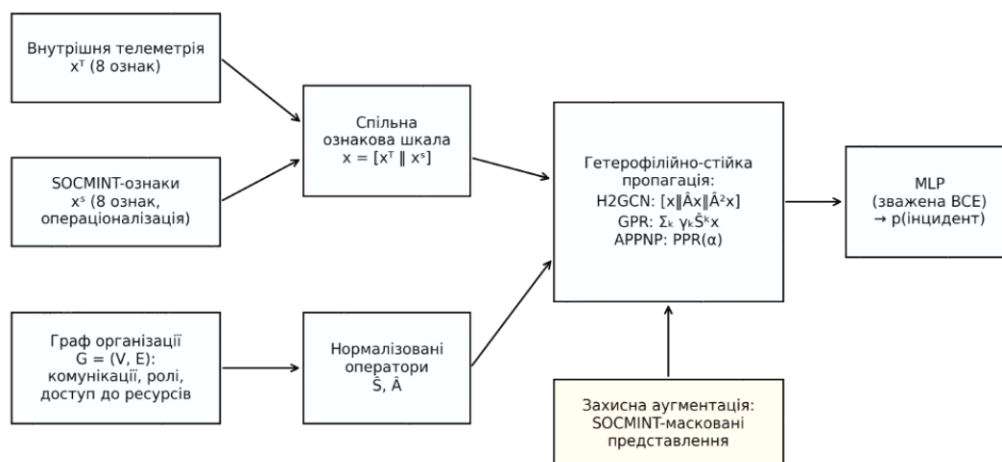


Рис. 2. Структурна схема гібридної моделі прогнозування інсайдерських інцидентів

Вибірку стратифіковано й поділено у співвідношенні 70:30; параметри нормалізації оцінено лише на навчальній частині. Порівнювали дві моделі градієнтного бустингу (300 дерев глибини 3, темп навчання 0,05, субдискретизація 0,8) із сигмоїдним калібруванням за п'ятиблоковою перехресною перевіркою: на самих внутрішніх ознаках і на об'єднаному наборі. Зерно генератора зафіксовано (108). Невизначеність оцінено бутстрепом тестової вибірки (2000 повторень) та двадцятьма незалежними повтореннями всього експерименту.

На другому етапі той самий ознаковий простір доповнено графом організації. Граф породжено стохастичною блоковою моделлю з вісьмома спільнотами ($n=2000$ вузлів, середній степінь $\approx 12,8$, гомофільія міток за ребрами $\approx 0,84$). Аномальний реляційний сигнал інсайдера змодельовано як надлишкові міжспільнотні зв'язки з кастодіанами чутливих ресурсів: це відтворює локальну гетерофільію привілейованих хабів. Привілейованість задано окремою рольовою міткою з помірною ранговою кореляцією зі степенем ($\rho \approx 0,41$), щоб відокремити ефект ролі від тривіального впливу зв'язності. Сили телеметричного та OSINT-сигналів підбрано так, щоб безграфові базові рівні збігалися з першим етапом. Завдяки цьому обидві серії результатів прямо зіставні, а внесок структурного сигналу вимірюється відносно спільної бази. Дані розбито стратифіковано у співвідношенні 60/20/20, результати усереднено за десятьма повтореннями з фіксованими зернами.

Наголосимо на межах такої перевірки. Самостійна дисперсія та часова випередженість OSINT-сигналів, як і реляційний сигнал інсайдера, є припущеннями конструкції генератора. Тому експеримент перевіряє здатність запропонованого конвеєра реалізувати цей потенціал, а не сам факт його існування в

реальних даних.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Додана прогнозна цінність OSINT-ознак

Ранжування ознак за внеском у прогноз, одержане методом перестановки на відкладеній вибірці, показано на рис. 3. Внутрішні та OSINT-ознаки доповнюють одна одну: серед найвпливовіших опиняються і показники доступу й виходу даних, і публічна образа, негативна тональність та витік облікових даних. Отже, OSINT не дублює внутрішні сигнали, а дає самостійну інформацію.

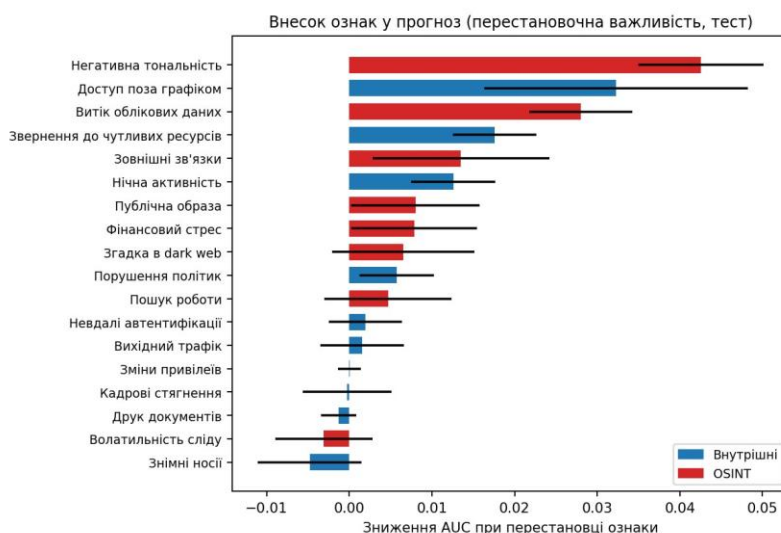


Рис. 3. Внесок ознак у прогноз ризику за методом перестановки

Криві робочих характеристик подано на рис. 4. Додавання OSINT-ознак піднімає площу під ROC-кривою з 0,685 (95 % бутстреп-інтервал 0,606–0,759) до 0,772 (0,710–0,832). За двадцятьма повтореннями експерименту середні значення становлять $0,679 \pm 0,033$ та $0,767 \pm 0,021$, а середній приріст $0,087 \pm 0,022$ стійко перевищує міжзапускову мінливість.

Клас інсайдерів рідкісний, тому інформативнішою є крива «точність – повнота» (рис. 5). Об'єднана модель випереджає базову практично на всьому діапазоні повноти; базовим рівнем точності тут слугує частка позитивного класу 0,11. Це означає суттєве поліпшення виявлення без надмірного зростання хибних триггерів.

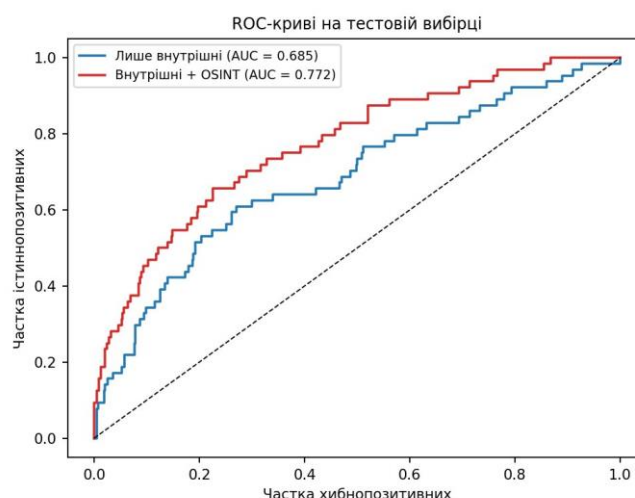


Рис. 4. ROC-криві моделей на внутрішніх ознаках та на об'єднаному наборі

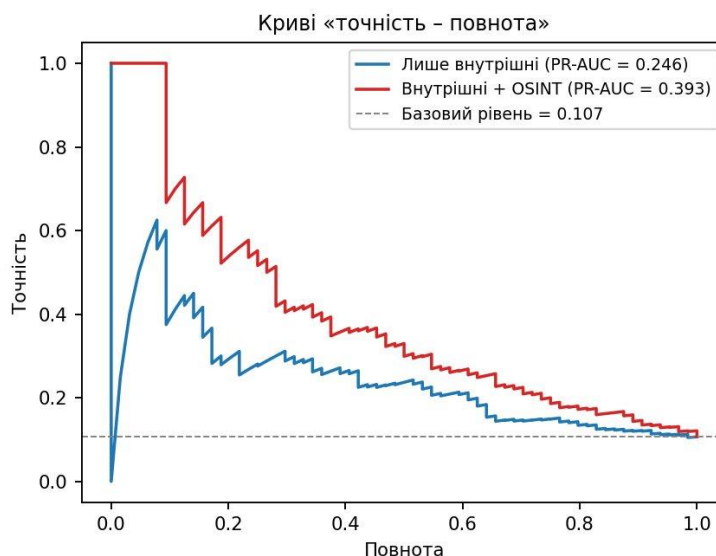


Рис. 5. Криві «точність – повнота» для порівнюваних моделей

Зведене порівняння якості наведено в табл. 3, де у дужках подано 95 % бутстреп-інтервали за 2000 повтореннями на тестовій вибірці. Приріст помітний за всіма показниками. Водночас довірчі інтервали окремих метрик частково перекриваються, а абсолютні значення точності на рідкісному класі залишаються помірними, що застерігає від переоцінювання можливостей прогнозування.

Прогнозна цінність OSINT великою мірою спирається на часову випередженість. Її змодельовано окремою симуляцією щоденних рядів. Для кожного профілю класу ризику початок наростання OSINT-сигналу задано рівномірно за 14–35 днів до інциденту, внутрішнього – за 3–14 днів; поріг тривоги встановлено на рівні 5 % хибних спрацювань на профілях без ризику. Розподіл випередження (рис. 6) показує, що за цих сценарних припущень відкриті сигнали перетинають поріг із медіанним випередженням двадцять один день, а внутрішні – шість (середні значення 22 та 9 днів). Частина внутрішніх сигналів узагалі не встигає перетнути поріг до події. Раннє вікно дає змогу перейти від реагування до запобігання нетехнічними заходами з підтримки персоналу.

Таблиця 3

Якість прогнозування інсайдерського ризику без OSINT-ознак та з ними

Показник якості	Лише внутрішні	Внутрішні + OSINT
AUC (ROC)	0,685 (0,606–0,759)	0,772 (0,710–0,832)
PR-AUC (точність – повнота)	0,246 (0,173–0,371)	0,393 (0,283–0,516)
Повнота за частки хибних 0,05	0,172 (0,088–0,303)	0,297 (0,194–0,442)
Оцінка Браєра (менше – краще)	0,090 (0,071–0,110)	0,081 (0,064–0,098)

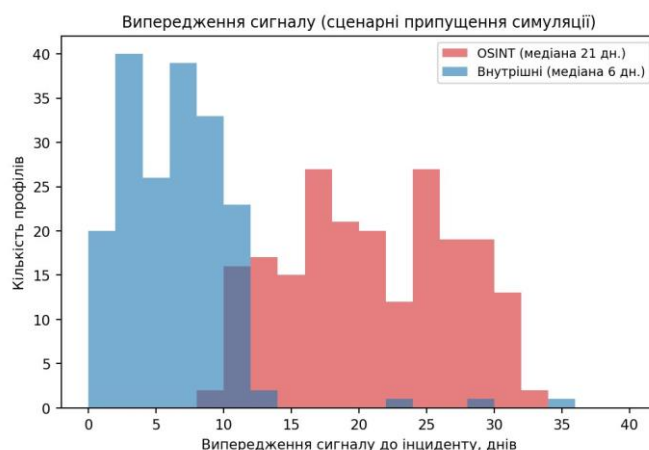


Рис. 6. Випередження сигналу до інциденту для OSINT та внутрішніх джерел

Додана чутливість супроводжується зростанням навантаження через хибні спрацювання. Залежність частки виявлених інсайдерів від частки хибнопозитивних (рис. 7) показує, що навіть у межах п'ятивідсоткового бюджету хибних тривог об'єднана модель виявляє більше інсайдерів. Проте кожне спрацювання стосується реальної особи, тому потребує процедур перевірки та захисту від несправедливих звинувачень.

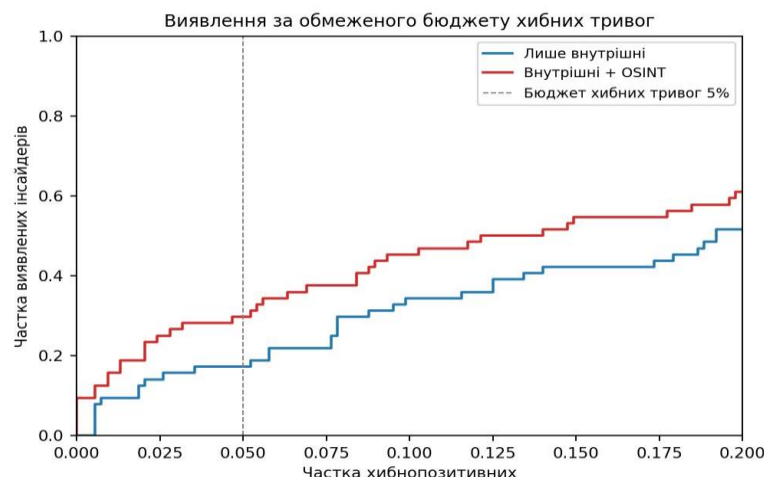


Рис. 7. Залежність частки виявлених інсайдерів від частки хибнопозитивних

Надійність операціоналізації перевіряють і на рівні взаємозв'язків між ознаками. Матриця кореляцій OSINT-ознак (рис. 8) виявляє переважно слабкі залежності, тобто набір ненадлишковий. Підвищені значення між спорідненими індикаторами тональності та публічної образи є сигналом до їх об'єднання або зважування, щоб не враховувати один і той самий конструкт двічі.

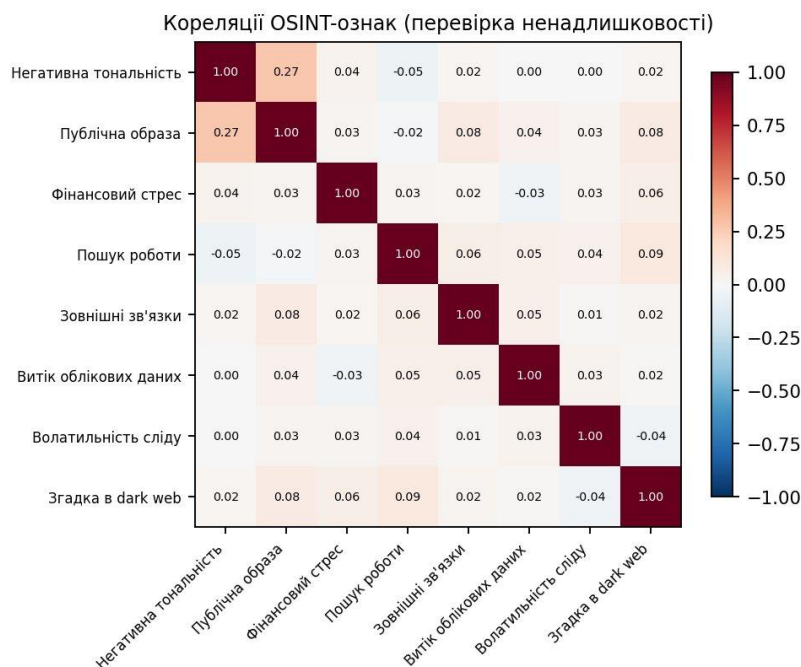


Рис. 8. Матриця кореляцій OSINT-ознак як перевірка ненадлишковості

Окремою умовою застосовності є калібрування. Діаграма надійності (рис. 9) показує прийнятну відповідність прогнозованих ймовірностей спостережуваним частотам інсайдерів. Це дає змогу трактувати вихід моделі як рівень ризику, а не просту бінарну позначку. Каліброваність є передумовою справедливого встановлення порогів втручання.

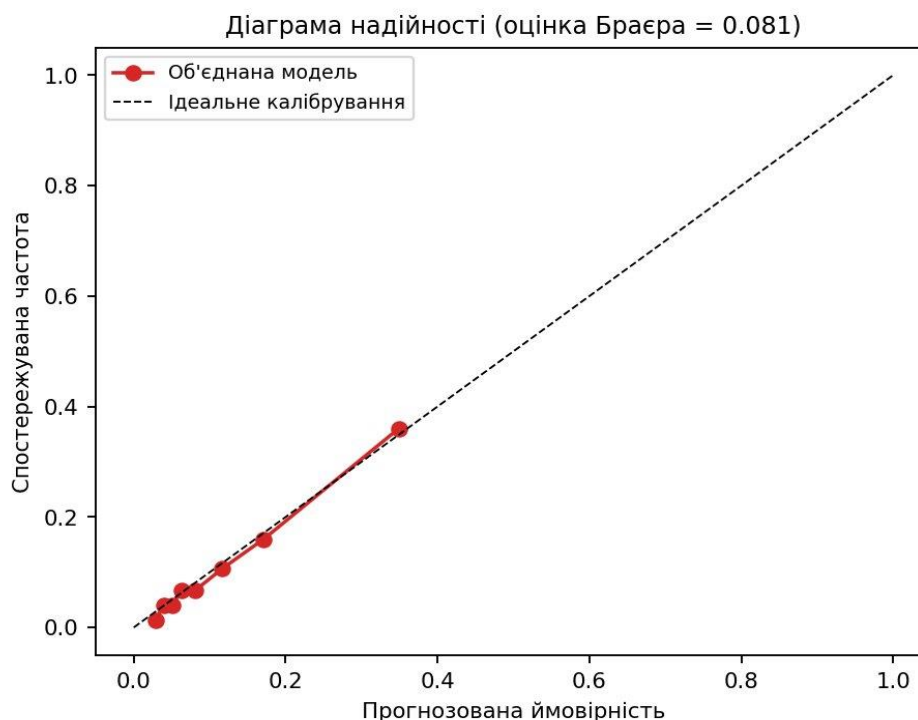


Рис. 9. Діаграма надійності прогнозу об'єднаної моделі

4.Внесок структурного сигналу та поведінка на привілейованому сегменті

Результати другого етапу наведено в табл. 4 та на рис. 10. Додавання OSINT-ознак до телеметрії підвищує AUC-ROC з 0,685 до 0,771, відтворюючи результат першого етапу. Введення структурного сигналу двокроковою згортою підвищує його далі – до 0,831. Отже, приріст від графового контексту (+0,060) співмірний із приростом від самих відкритих сигналів (+0,086) і додається до нього. Це підтверджує доцільність поєднання обох джерел в одній моделі.

Таблиця 4

Порівняння варіантів моделі за чистих умов (10 повторень; MLP-T – лише телеметрія, MLP-TS – телеметрія та OSINT-ознаки)

Модель	AUC	PR-AUC	F1	AUC (прив.)	F1 (прив.)	ECE	FPR@90
MLP-T	0,685±0,035	0,163	0,192	0,705	0,304	0,328	0,660
MLP-TS	0,771±0,037	0,228	0,260	0,799	0,373	0,266	0,526
GCN-lin	0,831±0,033	0,315	0,354	0,794	0,422	0,281	0,415
GCN-8	0,830±0,144	0,348	0,357	0,786	0,396	0,293	0,354
H2GCN	0,787±0,048	0,273	0,262	0,810	0,318	0,193	0,505
GPR/SIGN	0,793±0,046	0,255	0,290	0,815	0,327	0,227	0,497
APNP	0,804±0,026	0,247	0,281	0,821	0,390	0,308	0,457
GCNII	0,779±0,036	0,238	0,264	0,799	0,348	0,268	0,533

На привілейованому сегменті картина інша. Глибока згортка деградує через надзгладжування (GCN-8: 0,786 за різкого зростання дисперсії ±0,144). Гетерофіліїно-стійкі архітектури, навпаки, утримують якість: APNP дає найвищий привілейований AUC-ROC 0,821, GPR/SIGN – 0,815, H2GCN – 0,810 і водночас найнижчу похибку калібрування (ECE 0,193) проти 0,281–0,293 у гомофілійних згорток. Зазначимо, що за показником F1 на привілейованому сегменті найкращою залишається двокрокова згортка (0,422). Тобто перевага гетерофіліїно-стійких архітектур виявляється насамперед у ранжуванні та каліброваності, а не в якості за фіксованого порогу.

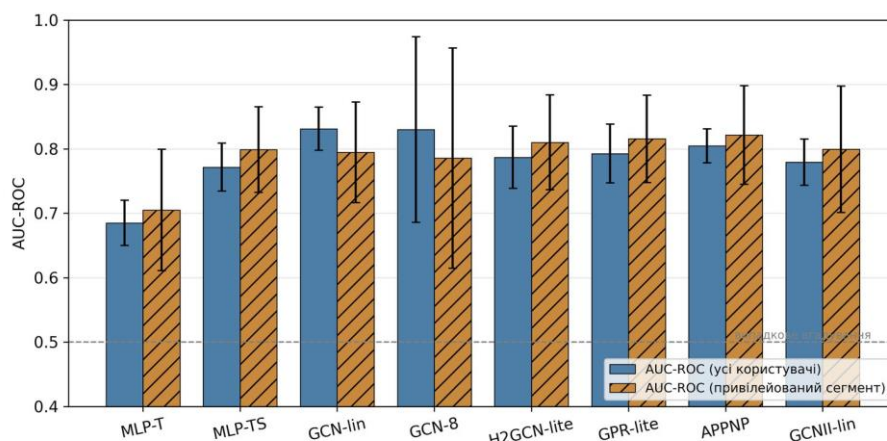


Рис. 10. AUC-ROC загалом та на привілейованому сегменті за чистих умов

Змагальна стійкість і захисне донавчання

Досліджено три сценарії цілеспрямованих атак, що діють лише на тестових інсайдерів: A1 – маскування OSINT-ознак заміщенням медіаною норми, що імітує санітизацію публічного сліду; A2 – приховування 50 % аномальних міжспільнотних зв'язків; A3 – їх комбінація. Результати наведено в табл. 5 та на рис. 11.

Таблиця 5

Змагальна стійкість (AUC-ROC) та ефект захисного донавчання

Модель	Без атаки	A1	A2	A3	A3 + захист
MLP-TS	0,771	0,681	0,771	0,681	0,698
GCN-lin	0,831	0,754	0,757	0,684	0,694
GCN-8	0,830	0,738	0,703	0,620	0,616
H2GCN-lite	0,787	0,717	0,765	0,695	0,735
GPR-lite	0,793	0,707	0,773	0,684	0,743
APPNP	0,804	0,683	0,793	0,673	0,689
GCNII-lin	0,779	0,682	0,778	0,681	0,696

Маскування відкритого сліду є найдієвішим вектором: MLP-TS падає з 0,771 до 0,681. Комбінована атака знижує якість усіх моделей до 0,62–0,70, тобто фактично до рівня самої телеметрії (0,685). Інакше кажучи, успішна маніпуляція нівелює внесок і відкритих сигналів, і структури. Найвразливішою є глибока згортка GCN-8 (0,620), яка опускається навіть нижче за телеметричну базу. Це кількісно підтверджує застереження щодо принципової вразливості OSINT-ознак, які контролює сам працівник.

Захисне донавчання за формулою (18) підвищує залишкову якість під комбінованою атакою саме для гетерофільно-стійких моделей: GPR/SIGN – з 0,684 до 0,743, H2GCN – з 0,695 до 0,735, причому без втрати чистої якості. Для глибокої згортки воно неефективне (0,616). Отже, аугментація зміщує модель до безпечнішого компромісу, але не забезпечує повного захисту самотужки. Це вказує на потребу поєднувати автоматизований прогноз із калібруванням порогів та обов'язковим наглядом людини.

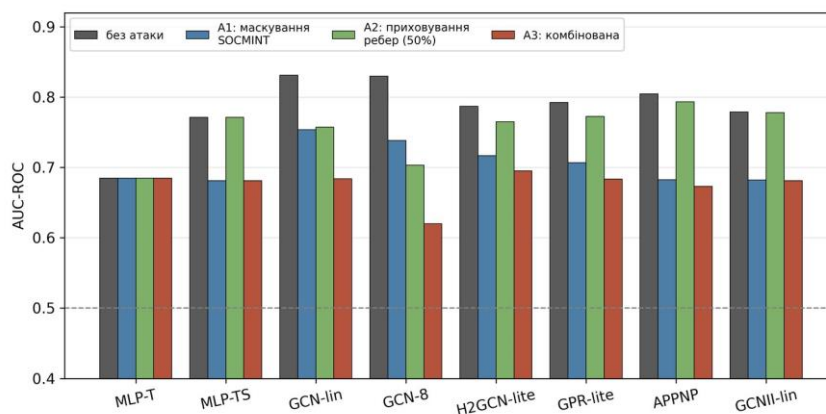


Рис. 11. Деградація AUC-ROC під атаками та ефект захисного донавчання

**Обмеження дослідження та запобіжники відповідального застосування**

Окреслимо обмеження, що визначають межі інтерпретації результатів. По-перше, обидва етапи експерименту виконано на синтетичних даних, у яких цінність OSINT і реляційний сигнал закладені в конструкції генератора; одержані прирости слід розглядати як верхню межу, досяжну за умови справдження цих припущень, а не як емпіричне підтвердження. По-друге, зіставлення публічних профілів із конкретним працівником (розв'язання ідентичності) саме є джерелом похибок і хибної атрибуції, яких експеримент не моделює. По-третє, реальна базова частота інсайдерських інцидентів істотно нижча за прийнятні одиниці відсотків, тому точність на позитивному класі в розгортанні буде нижчою за наведену; нестаціонарність поведінки й дрейф джерел вимагатимуть регулярного перенавчання та переоцінювання порогів. По-четверте, графові моделі реалізовано в декомпонованій схемі «пропаганди – класифікатор», що спрощує відтворення, але не вичерпує можливостей повних нелінійних архітектур.

Приріст якості поєднується з людськими наслідками хибних спрацювань. Це визначає набір обов'язкових запобіжників відповідального застосування OSINT-ознак:

- правомірність і пропорційність збору: перелік джерел обмежено лише відкритими даними з законною підставою обробки, прихований моніторинг приватного спілкування виключено [27];
- мінімізація даних і знеособлення на етапах, де ідентифікація особи не потрібна, разом із чіткими строками зберігання;
- калібрування ймовірностей і контроль порогів, що утримують навантаження хибними тривогами в прийнятних межах і запобігають масовим безпідставним підозрам;
- обов'язковий нагляд людини та інтерпретовність рішень, які унеможливають автоматичні кадрові санкції на підставі самого лише модельного бала [6];
- організаційне розмежування ролей і захист від зловживання самими OSINT-засобами, оскільки відкрита розвідка є інструментом подвійного призначення [12].

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Операціоналізація OSINT-ознак є вузловою ланкою. Від неї залежить, чи стане відкрита розвідка корисним джерелом раннього прогнозу інсайдерського ризику, чи перетвориться на генератор шуму та неправомірного втручання. Подання процедури як строгого ланцюга переходів від конструкта до нормалізованої ознаки, підпорядкованого критеріям валідності, надійності, випереджальності та правомірності, дає відтворюваний спосіб включення відкритих сигналів у моделі. Обчислювальний експеримент продемонстрував, що запропонований конвеєр здатний реалізувати потенціал відкритих сигналів, закладених у даних: поєднання внутрішньої телеметрії з OSINT-ознаками підвищило площу під ROC-кривою з 0,685 до 0,772 та поліпшило точність на рідкісному класі.

Розширення підходу в напрямі гібридного моделювання показало, що ізольований розгляд користувача не вичерпує наявної інформації. Систематизація недоліків наявних графових підходів виявила п'ять системних проблем; головні з них – гомофілійне припущення та недосліджена змагальна стійкість. Побудована з їх урахуванням гібридна модель, що поєднує операціоналізовані OSINT-ознаки, внутрішню телеметрію та структурний сигнал графа організації, підвищила площу під ROC-кривою до 0,831, причому приріст від графового контексту виявився співмірним із приростом від самих відкритих сигналів. Ключовим є висновок щодо привілейованих облікових записів: саме на цьому найризикованішому сегменті глибока гомофілійна згортка деградує через надзгладжування подань, тоді як гетерофілійно-стійкі оператори утримують якість виявлення на рівні 0,80–0,82 за площею під ROC-кривою та забезпечують найнижчу похибку калібрування. Це безпосередньо знижує ціну хибних тривог для кадрових і безпекових підрозділів.

Уперше для цієї задачі кількісно оцінено змагальну стійкість. Установлено, що маскуваність відкритого цифрового сліду є найдієвішим вектором атаки, а його поєднання з приховуванням аномальних міжспільнотних зв'язків знижує якість усіх розглянутих моделей до рівня самої телеметрії; найвразливішою виявляється глибока згортка. Запропоноване захисне донавчання з маскованою аугментацією частково відновлює стійкість без істотної втрати чистої якості, проте не забезпечує повного захисту самотужки. Це підтверджує необхідність поєднувати автоматизований прогноз із калібруванням порогів, мінімізацією даних та обов'язковим наглядом людини.

Практична значущість результату подвійна: він окреслює методичний шлях підвищення чутливості прогнозу й одночасно фіксує ціну у вигляді хибнопозитивних спрацювань і ризиків для приватності. Усі одержані оцінки мають статус верхньої межі, оскільки отримані на повністю специфікованих синтетичних даних. Перспективу подальших досліджень становлять валідація моделі на реальних знеособлених наборах, зокрема CERT Insider Threat і TWOS; розроблення темпорального механізму раннього попередження з калібруванням ймовірностей ризику в часі; дослідження адаптивних стратегій



змагальника, що враховують наявність захисту; а також формалізація метрик правомірності для автоматизованого контролю меж застосування OSINT у кадрових процесах.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Cybersecurity Insiders. (2025). 2025 insider risk report: The shift to predictive whole-person insider risk management. <https://www.cybersecurity-insiders.com/2025-insider-risk-report-the-shift-to-predictive-whole-person-insider-risk-management/>
2. Global Market Insights. (2025). Open-source intelligence (OSINT) market size, share and growth forecast. <https://www.gminsights.com/industry-analysis/open-source-intelligence-osint-market>
3. Park, W., You, Y., & Lee, K. (2018). Detecting potential insider threat: Analyzing insiders' sentiment exposed in social media. *Security and Communication Networks*, 2018, Article 7243296. <https://doi.org/10.1155/2018/7243296>
4. Osteritter, L., & Carley, K. M. (2021). Conversations around organizational risk and insider threat. In *Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 613–621). IEEE. <https://doi.org/10.1145/3487351.3492721>
5. Gao, Y., Wang, X., He, X., Liu, Z., Feng, H., & Zhang, Y. (2023). Addressing heterophily in graph anomaly detection: A perspective of graph spectrum. In *Proceedings of the ACM Web Conference 2023* (pp. 1528–1538). Association for Computing Machinery. <https://doi.org/10.1145/3543507.3583268>
6. Kirichenko, L., Radivilova, T., Pichugina, O., Chala, L., & Khandak, D. (2022). Trend detection in short time series using discrete wavelet transform. In *Proceedings of the IX International Scientific Conference "Information Technology and Implementation" (IT&I-2022) (CEUR Workshop Proceedings, Vol. 3347, pp. 159–168). CEUR-WS. https://ceur-ws.org/Vol-3347/*
7. Іванченко, С., Аверічев, І., & Рижаков, М. (2025). Узагальнена модель прогнозування та виявлення кібербезпекових аномалій на основі штучного інтелекту. *Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка»*, 4(28), 529–546. <https://doi.org/10.28925/2663-4023.2025.28.823>
8. Radivilova, T. i Panteliev, V. (2026) «Баєсівська часова модель для раннього прогнозування інсайдерських загроз», *Сучасний стан наукових досліджень та технологій в промисловості*, (3(37), с. 107–119. doi: 10.30837/2522-9818.2026.3.107.
9. Browne, T. O., Abedin, M., & Chowdhury, M. J. M. (2024). A systematic review on research utilising artificial intelligence for open source intelligence (OSINT) applications. *International Journal of Information Security*, 23, 2911–2938. <https://doi.org/10.1007/s10207-024-00868-2>
10. Babenko, T., Kolesnikova, K., Abramkina, O., & Vitulyova, Y. (2025). Automated OSINT techniques for digital asset discovery and cyber risk assessment. *Computers*, 14(10), Article 430. <https://doi.org/10.3390/computers14100430>
11. Drozd, M., & Nesterenko, S. (2025). Analysis of software for the implementation of OSINT in the field of information security. *Bulletin of V.N. Karazin Kharkiv National University, Series «Mathematical Modeling. Information Technology. Automated Control Systems»*, 66, 45–55. <https://doi.org/10.26565/2304-6201-2025-66-04>
12. Легомінова, С., Щавінський, Ю., Рабчун, Д., Запорожченко, М., & Будзинський, О. (2024). Небезпека інструментів OSINT та способи пом'якшення наслідків їх використання для організації. *Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка»*, 1(25), 294–303. <https://doi.org/10.28925/2663-4023.2024.25.294303>
13. Akinsola, F. A., Ogwueleka, F. N., & Mbanaso, U. M. (2025). A comprehensive survey of insider threat landscape and detection indicators. *Kwaghe International Journal of Engineering and Information Technology*, 2(3). <https://doi.org/10.58578/kijeit.v2i3.7704>
14. Pichugina, O., Kirichenko, L., & Radivilova, T. (2022). Binary classification: Ensemble methods utilizing decision theory tools. In *Proceedings of the 2nd International Workshop of IT-professionals on Artificial Intelligence (ProfIT AI 2022) (CEUR Workshop Proceedings, Vol. 3348, pp. 19–33). CEUR-WS. https://ceur-ws.org/Vol-3348/*
15. Gong, Y., Cui, S., Liu, S., Jiang, B., Dong, C., & Lu, Z. (2024). Graph-based insider threat detection: A survey. *Computer Networks*, 254, Article 110757. <https://doi.org/10.1016/j.comnet.2024.110757>
16. Fei, K., Zhou, J., Su, L., Wang, W., & Chen, Y. (2025). Log2Graph: A graph convolution neural network based method for insider threat detection. *Journal of Computer Security*, 33(2), 37–56. <https://doi.org/10.3233/JCS-230092>
17. Li, X., Li, X., Jia, J., Li, L., Yuan, J., Gao, Y., & Yu, S. (2023). A high accuracy and adaptive anomaly detection model with dual-domain graph convolutional network for insider threat detection. *IEEE*



- Transactions on Information Forensics and Security, 18, 1638–1652. <https://doi.org/10.1109/TIFS.2023.3245413>
18. Thi, V. D., Duc, T. T., Nguyen, T. V., & Luong, H.-M. P. (2026). Insider threat detection using knowledge graphs and RiskScore-guided graph neural networks. *Engineering, Technology & Applied Science Research*, 16(2), 33712–33721. <https://doi.org/10.48084/etasr.17187>
 19. Zhu, J., Yan, Y., Zhao, L., Heimann, M., Akoglu, L., & Koutra, D. (2020). Beyond homophily in graph neural networks: Current limitations and effective designs. In *Advances in Neural Information Processing Systems* (Vol. 33). Curran Associates. <https://doi.org/10.48550/arXiv.2006.11468>
 20. Chien, E., Peng, J., Li, P., & Milenkovic, O. (2021). Adaptive universal generalized PageRank graph neural network. In *Proceedings of the 9th International Conference on Learning Representations (ICLR 2021)*. <https://doi.org/10.48550/arXiv.2006.07988>
 21. Gasteiger, J., Bojchevski, A., & Günnemann, S. (2019). Predict then propagate: Graph neural networks meet Personalized PageRank. *arXiv*. <https://doi.org/10.48550/arXiv.1810.05997>
 22. Chen, M., Wei, Z., Huang, Z., Ding, B., & Li, Y. (2020). Simple and deep graph convolutional networks. In *Proceedings of the 37th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 119, pp. 1725–1735)*. PMLR. <https://doi.org/10.48550/arXiv.2007.02133>
 23. Радівілова, Т., Добринін, І., Пантелєєв, В., Фісенко, Д., Мазепа, А., & Білодід, В. (2025). Аналіз методів прогнозування внутрішніх загроз на основі аналізу даних соціальної мережі Twitter. *Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка»*, 4(28), 478–489. <https://doi.org/10.28925/2663-4023.2025.28.818>
 24. Xu, F., Wang, N., Wu, H., Wen, X., Zhao, X., & Wan, H. (2024). Revisiting graph-based fraud detection in sight of heterophily and spectrum. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(8), 9214–9222. <https://doi.org/10.1609/aaai.v38i8.28773>
 25. Zügner, D., Akbarnejad, A., & Günnemann, S. (2018). Adversarial attacks on neural networks for graph data. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2847–2856). Association for Computing Machinery. <https://doi.org/10.1145/3219819.3220078>
 26. Радівілова, Т., & Пантелєєв, В. (2026). Метод темпорального раннього прогнозування внутрішніх інцидентів з калібруванням ймовірностей ризику. *Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка»*, 2(34), 687–701. <https://doi.org/10.28925/2663-4023.2026.34.1362>
 27. Думчиков, М. О. (2025). Відкрита розвідка (OSINT) у протидії інформаційним війнам: аналітичні інструменти та правові межі. *Аналітично-порівняльне правознавство*, 2(6), 273–277. <https://doi.org/10.24144/2788-6018.2025.06.2.43>
 28. Nonum, E. O., Avwokuruaye, O., & Ezemonye, T. M. (2025). Role of open source intelligence (OSINT) in cybersecurity and threat analysis. *International Journal of Latest Technology in Engineering, Management & Applied Science*, 14(3), 189–200. <https://doi.org/10.51583/IJLTEMAS.2025.140300023>

**Vadym O. Pantieliev**

PhD student, V. V. Popovskyy Department of Infocommunication Engineering
Kharkiv National University of Radio Electronics, Kharkiv, Ukraine
ORCID: 0009-0008-7824-8782
vadym.pantieliev@nure.ua

Volodymyr Bilodid

Researcher
Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine
ORCID: 0009-0002-8976-2310
vibos111@gmail.com

OPERATIONALISATION OF OSINT FEATURES AND A HYBRID MODEL FOR PREDICTING INSIDER INCIDENTS

Abstract. The article proposes a methodological approach to predicting insider incidents. The approach combines the operationalisation of open-source intelligence (OSINT) features with a hybrid model that accounts for the structure of the corporate graph. The behavioural precursors of internal incidents – disgruntlement, financial pressure, external affiliations and job-search activity – often surface in the public digital space earlier than in an organisation’s internal telemetry. Yet their conversion into valid, reliable and lawfully admissible model features remains an unsolved task. Operationalisation is presented as a chain of mappings: a latent construct, an indicator, an OSINT source, a measurement with normalisation, and a ready feature that is merged with internal access and activity metrics. A taxonomy of OSINT features across eight categories is proposed together with four quality criteria: content validity, measurement reliability, temporal lead and lawfulness of collection. A computational experiment on a synthetic set of two thousand user profiles showed that, under the stated scenario assumptions, adding OSINT features raises the area under the ROC curve from 0.685 to 0.772, improves precision on the rare class and provides a noticeable lead of the signal before an incident. The second part of the study addresses the limitations of existing graph-based approaches. Five shortcomings are systematised; the principal ones are the homophily assumption of message passing, oversmoothing of representations at privileged hub nodes, and unexamined adversarial robustness. A hybrid model is built with these in mind. It integrates OSINT features, internal telemetry and the structural signal of the communication and resource-access graph, and performs propagation with heterophily-robust operators. In an experiment harmonised with the first stage in terms of baseline levels, the structural signal raises the area under the ROC curve to 0.831. On the privileged account segment, deep homophilous convolution degrades, whereas heterophily-robust architectures maintain quality at 0.80–0.82 and achieve the lowest calibration error. For the first time in this task, adversarial robustness is quantified. Masking the open digital footprint together with concealing anomalous ties reduces predictive quality to the level of telemetry alone. The proposed defensive fine-tuning with masked augmentation partially restores robustness without a substantial loss of clean-condition quality. The ethical and legal limits to applying OSINT in personnel processes are also outlined in accordance with personal data protection principles.

Keywords: OSINT; insider threats; operationalization of indicators; graph neural networks; security events; security incidents; forecasting; personal data protection.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Cybersecurity Insiders. (2025). 2025 insider risk report: The shift to predictive whole-person insider risk management. <https://www.cybersecurity-insiders.com/2025-insider-risk-report-the-shift-to-predictive-whole-person-insider-risk-management/>
2. Global Market Insights. (2025). Open-source intelligence (OSINT) market size, share and growth forecast. <https://www.gminsights.com/industry-analysis/open-source-intelligence-osint-market>
3. Park, W., You, Y., & Lee, K. (2018). Detecting potential insider threat: Analyzing insiders’ sentiment exposed in social media. *Security and Communication Networks*, 2018, Article 7243296. <https://doi.org/10.1155/2018/7243296>
4. Ostertter, L., & Carley, K. M. (2021). Conversations around organizational risk and insider threat. In *Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis*



- and Mining (pp. 613–621). IEEE. <https://doi.org/10.1145/3487351.3492721>
5. Gao, Y., Wang, X., He, X., Liu, Z., Feng, H., & Zhang, Y. (2023). Addressing heterophily in graph anomaly detection: A perspective of graph spectrum. In *Proceedings of the ACM Web Conference 2023* (pp. 1528–1538). Association for Computing Machinery. <https://doi.org/10.1145/3543507.3583268>
 6. Kirichenko, L., Radivilova, T., Pichugina, O., Chala, L., & Khandak, D. (2022). Trend detection in short time series using discrete wavelet transform. In *Proceedings of the IX International Scientific Conference “Information Technology and Implementation” (IT&I-2022)* (CEUR Workshop Proceedings, Vol. 3347, pp. 159–168). CEUR-WS. <https://ceur-ws.org/Vol-3347/>
 7. Ivanchenko, E., Averichev, I., & Ryzhakov, M. (2025). A Generalized Model for Predicting and Detecting Cybersecurity Anomalies Based on Artificial Intelligence. *Electronic Professional Scientific Journal “Cybersecurity: Education, Science, Technology,”* 4(28), 529–546. <https://doi.org/10.28925/2663-4023.2025.28.823>
 8. Radivilova, T., & Panteliev, V. (2026). “A Bayesian Time Series Model for Early Prediction of Insider Threats,” *Current State of Scientific Research and Technologies in Industry*, 3(37), pp. 107–119. doi: 10.30837/2522-9818.2026.3.107.
 9. Browne, T. O., Abedin, M., & Chowdhury, M. J. M. (2024). A systematic review on research utilising artificial intelligence for open source intelligence (OSINT) applications. *International Journal of Information Security*, 23, 2911–2938. <https://doi.org/10.1007/s10207-024-00868-2>
 10. Babenko, T., Kolesnikova, K., Abramkina, O., & Vitulyova, Y. (2025). Automated OSINT techniques for digital asset discovery and cyber risk assessment. *Computers*, 14(10), Article 430. <https://doi.org/10.3390/computers14100430>
 11. Drozd, M., & Nesterenko, S. (2025). Analysis of software for the implementation of OSINT in the field of information security. *Bulletin of V.N. Karazin Kharkiv National University, Series «Mathematical Modeling. Information Technology. Automated Control Systems»*, 66, 45–55. <https://doi.org/10.26565/2304-6201-2025-66-04>
 12. Legominova, S., Shchavinsky, Y., Rabchun, D., Zaporozhchenko, M., & Budzinsky, O. (2024). The Risks of OSINT Tools and Ways to Mitigate the Consequences of Their Use for Organizations. *Electronic Professional Scientific Journal “Cybersecurity: Education, Science, Technology,”* 1(25), 294–303. <https://doi.org/10.28925/2663-4023.2024.25.294303>
 13. Akinsola, F. A., Ogwueleka, F. N., & Mbanaso, U. M. (2025). A comprehensive survey of insider threat landscape and detection indicators. *Kwaghe International Journal of Engineering and Information Technology*, 2(3). <https://doi.org/10.58578/kijeit.v2i3.7704>
 14. Pichugina, O., Kirichenko, L., & Radivilova, T. (2022). Binary classification: Ensemble methods utilizing decision theory tools. In *Proceedings of the 2nd International Workshop of IT-professionals on Artificial Intelligence (ProFIT AI 2022)* (CEUR Workshop Proceedings, Vol. 3348, pp. 19–33). CEUR-WS. <https://ceur-ws.org/Vol-3348/>
 15. Gong, Y., Cui, S., Liu, S., Jiang, B., Dong, C., & Lu, Z. (2024). Graph-based insider threat detection: A survey. *Computer Networks*, 254, Article 110757. <https://doi.org/10.1016/j.comnet.2024.110757>
 16. Fei, K., Zhou, J., Su, L., Wang, W., & Chen, Y. (2025). Log2Graph: A graph convolution neural network based method for insider threat detection. *Journal of Computer Security*, 33(2), 37–56. <https://doi.org/10.3233/JCS-230092>
 17. Li, X., Li, X., Jia, J., Li, L., Yuan, J., Gao, Y., & Yu, S. (2023). A high accuracy and adaptive anomaly detection model with dual-domain graph convolutional network for insider threat detection. *IEEE Transactions on Information Forensics and Security*, 18, 1638–1652. <https://doi.org/10.1109/TIFS.2023.3245413>
 18. Thi, V. D., Duc, T. T., Nguyen, T. V., & Luong, H.-M. P. (2026). Insider threat detection using knowledge graphs and RiskScore-guided graph neural networks. *Engineering, Technology & Applied Science Research*, 16(2), 33712–33721. <https://doi.org/10.48084/etasr.17187>
 19. Zhu, J., Yan, Y., Zhao, L., Heimann, M., Akoglu, L., & Koutra, D. (2020). Beyond homophily in graph neural networks: Current limitations and effective designs. In *Advances in Neural Information Processing Systems* (Vol. 33). Curran Associates. <https://doi.org/10.48550/arXiv.2006.11468>
 20. Chien, E., Peng, J., Li, P., & Milenkovic, O. (2021). Adaptive universal generalized PageRank graph neural network. In *Proceedings of the 9th International Conference on Learning Representations (ICLR 2021)*. <https://doi.org/10.48550/arXiv.2006.07988>
 21. Gasteiger, J., Bojchevski, A., & Günnemann, S. (2019). Predict then propagate: Graph neural networks meet Personalized PageRank. *arXiv*. <https://doi.org/10.48550/arXiv.1810.05997>
 22. Chen, M., Wei, Z., Huang, Z., Ding, B., & Li, Y. (2020). Simple and deep graph convolutional networks. In *Proceedings of the 37th International Conference on Machine Learning (Proceedings of Machine*



- Learning Research, Vol. 119, pp. 1725–1735). PMLR. <https://doi.org/10.48550/arXiv.2007.02133>
23. Radivilova, T., Dobrynin, I., Panteleev, V., Fisenko, D., Mazepa, A., & Bilodid, V. (2025). Analysis of methods for forecasting internal threats based on an analysis of data from the social network Twitter. *Electronic Professional Scientific Journal “Cybersecurity: Education, Science, Technology,”* 4(28), 478–489. <https://doi.org/10.28925/2663-4023.2025.28.818>
 24. Xu, F., Wang, N., Wu, H., Wen, X., Zhao, X., & Wan, H. (2024). Revisiting graph-based fraud detection in sight of heterophily and spectrum. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(8), 9214–9222. <https://doi.org/10.1609/aaai.v38i8.28773>
 25. Zügner, D., Akbarnejad, A., & Günnemann, S. (2018). Adversarial attacks on neural networks for graph data. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2847–2856). Association for Computing Machinery. <https://doi.org/10.1145/3219819.3220078>
 26. Radivilova, T., & Panteleev, V. (2026). A Method for Temporal Early Prediction of Internal Incidents with Risk Probability Calibration. *Electronic Professional Scientific Journal “Cybersecurity: Education, Science, Technology,”* 2(34), 687–701. <https://doi.org/10.28925/2663-4023.2026.34.1362>
 27. Dumchikov, M. O. (2025). Open-Source Intelligence (OSINT) in Countering Information Wars: Analytical Tools and Legal Limits. *Analytical and Comparative Jurisprudence*, 2(6), 273–277. <https://doi.org/10.24144/2788-6018.2025.06.2.43>
 28. Nonum, E. O., Avwokuruaye, O., & Ezemonye, T. M. (2025). Role of open source intelligence (OSINT) in cybersecurity and threat analysis. *International Journal of Latest Technology in Engineering, Management & Applied Science*, 14(3), 189–200. <https://doi.org/10.51583/IJLTEMAS.2025.140300023>

Отримано редакцією журналу / Received: 03.09.26

Прорецензовано / Revised: 16.09.26

Схвалено до друку / Accepted: 26.10.26



This work is licensed under Creative Commons Attribution-noncommercial-sharealike 4.0 International License.