КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 4 (28), 2025
ISSN 2663 - 4023
DOI 10.28925/2663-4023.2025.28.805
УДК 004. 056 : 519

Лавров Вадим Валерійович
аспірант кафедри захисту інформації
Вінницький національний технічний університет, Вінниця, Україна
ORCID ID: 0009-0005-7031-3891
vadim.lavrovv@gmail.com

Дудатьєв Андрій Веніамінович
к.т.н., доцент, доцент кафедри захисту інформації
Вінницький національний технічний університет, Вінниця, Україна
ORCID ID: 0000-0002-7944-2404
dydatyev.av@gmail.com

Гарнага Володимир Анатолійович
доцент кафедри захисту інформації
Вінницький національний технічний університет, Вінниця, Україна
ORCID ID: 0009-0000-9180-6445
garnaga.volodymyr@vntu.edu.ua

НЕЙРО-НЕЧІТКА СИСТЕМА ANFIS ДЛЯ ОЦІНЮВАННЯ РИЗИКУ ДЕЗІНФОРМАЦІЇ В УМОВАХ ІНФОРМАЦІЙНОЇ ВІЙНИ

Анотація. Стаття присвячена аналізу та розробці методики нейро-нечіткої системи ANFIS для оцінювання ризику дезінформації в умовах інформаційної війни. Запропоновано інтеграцію контентних, мережевих та психологічних факторів для розрахунку єдиного показника ризику (RiskScore) поширення дезінформації. В рамках моделі як вхідні змінні розглядаються: характеристики контенту (наприклад, правдивість чи емоційна забарвленість повідомлення), мережеві показники (наприклад, швидкість і масштаб розповсюдження, базове число репродукції R_0) та психологічні чинники аудиторії (рівень довіри, сприйнятливість до маніпуляцій тощо). Ці змінні подаються до адаптивної нечіткої системи, де на основі набору правил типу IF-THEN проводиться нечітка логічна оцінка ризику. Архітектура ANFIS дозволяє навчатися на обмежених або синтетичних даних, налаштовуючи параметри правил та функцій приналежності для підвищення точності. Запропонований підхід демонструє здатність об'єднати різнорідні фактори в єдину метричну оцінку ризику, що дає змогу виявляти потенційно небезпечні дезінформаційні кампанії на ранніх стадіях. Показано, що поєднання епідеміологічних показників (таких як базове число репродукції R_0) з контентно-психологічними характеристиками через нечітку логіку підвищує інформативність оцінки ризику, особливо в умовах інформаційних війн. Застосування ANFIS-моделі для оцінки ризику дезінформації може допомогти державним та оборонним структурам пріоритезувати зусилля з протидії інформаційним атакам, перевершуючи за гнучкістю традиційні лінійні моделі, епідеміологічні SIR-моделі та класичні нечіткі системи Мамдані. Результати можуть бути використані для подальшої розробки моделі.

Ключові слова: ризик дезінформації; нейро-нечітка система; ANFIS; оцінка ризику; інформаційна війна; моделювання дезінформації; інформаційний простір; кібербезпека.

ВСТУП

Постановка проблеми. В сучасних умовах інформаційні війни та цілеспрямовані кампанії дезінформації становлять значну загрозу для суспільства і національної безпеки [1], [2]. Соціальні мережі стали потужним засобом поширення фейкових новин, пропаганди та маніпулятивних повідомлень, які можуть впливати на громадську думку,

© В.В. Лавров, А.В. Дудатьєв, В.А. Гарнага, 2025

підривати довіру до офіційних джерел і навіть спричиняти паніку або суспільні заворушення. Зокрема, платформи соціальних мереж використовуються як інструмент «дипломатії» або неоголошеної війни, де ворожі актори запускають дезінформацію, аби вплинути на настрої вразливої аудиторії [2], [3].

Аналіз останніх досліджень і публікацій. Традиційні методи виявлення та протидії дезінформації — такі як фактчекінг або фільтрація контенту — часто не встигають за швидкістю вірусного поширення неправдивих відомостей [2]. Вже на початкових етапах поширення окремого фейку важливо оцінити, наскільки високим є ризик перетворення його на масову дезінформаційну «епідемію». У класичних епідеміологічних моделях поширення інформації, зокрема на основі SIR (Susceptible-Infected-Recovered), ризик визначається параметрами типу коефіцієнта передачі β та базового числа репродукції R_0 [4]. Якщо для певної інформації, це означає, що один «інфікований» (той, хто повірив або поділився дезінформацією) в середньому заражає більше ніж одну нову особу; як наслідок, поширення набуває характеру вибухового зростання [5]. Це свідчить про те, що хвиля дезінформації з часом затухне сама собою [5]. Однак, одного лише епідеміологічного порогу недостатньо: навіть при помірному дезінформація може бути дуже небезпечною, якщо її зміст спричиняє значний психологічний ефект (наприклад, паніку) або якщо вона цілеспрямовано поширюється через впливових осіб чи ботоферми (мережевий ефект). Більшість існуючих підходів фокусуються або на змісті повідомлень (наприклад, класифікація фейкових новин за допомогою машинного навчання), або на динаміці поширення (епідеміологічні моделі, аналіз соціальних мереж), або на індивідуальних особливостях аудиторії (дослідження психологічної сприйнятливості). Окремо ці підходи не дають повної картини ризику. Наприклад, контент-орієнтовані моделі визначають, чи є новина неправдивою, але не враховують, скільки людей її побачать і повірять. Мережеві моделі (SIR, SI, тощо) добре описують динаміку охоплення, але в них всі «інфіковані» повідомленнями трактуються однаково, без урахування відмінностей у змісті та реакції аудиторії [6]. Психологічні дослідження показують, що певні групи людей більш уразливі до дезінформації (наприклад, через вікові особливості, когнітивні упередження або політичні вподобання) [6], але ці фактори важко прямо підключити до моделей розповсюдження. Таким чином, виникає задача об’єднання контентних, мережевих та психологічних чинників у єдину модель, яка надаватиме інтегральну оцінку ризику дезінформації.

Мета статті. Метою статті є розробка методики застосування моделі, що дозволить оцінити ризик кожного окремого інформаційного повідомлення або кампанії з урахуванням: наскільки небезпечний зміст (дезінформативний, провокаційний), який потенціал поширення в мережі (структура соціального графа, наявність ботів, початкове), психологічну вразливість аудиторії (довіра до джерела, емоційний стан суспільства, когнітивні упередження тощо). Рішенням є побудова нейро-нечіткої моделі оцінки ризику, яка поєднає ці різнорідні фактори. У межах цієї мети передбачається:

- розробити методику для оцінки ризиків, що об’єднує контентні, мережеві та психологічні фактори, для формування єдиного показника ризику (RiskScore);
- виконати аналіз математичної моделі, яка поєднує нечітку логіку та штучні нейронні мережі, що дозволяє адаптувати модель до змінних умов і навчатися на обмежених даних;
- провести порівняння з іншими існуючими методами, підкреслюючи переваги адаптивності та здатності моделювати складні взаємодії між різними факторами.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Для розв'язання поставленої задачі використано Adaptive Neuro-Fuzzy Inference System (ANFIS) — адаптивну нейромережеву нечітку систему. ANFIS поєднує в собі переваги нечіткої логіки (робота з нечіткими множинами, інтерпретованість правил) та нейронних мереж (адаптивне навчання параметрів) [9]. В основі ANFIS лежить нечітка інференційна система типу Такаґі–Суґено (TSK), в якій правила мають вигляд «Якщо ... то ...» з лінгвістичними передумовами та функціональним (зазвичай лінійним) наслідком. Така архітектура здатна апроксимувати довільні нелінійні залежності між вхідними та вихідними даними, що підтверджує універсальність ANFIS як оцінювача функцій [10]. У нашому випадку це дозволяє побудувати модель, не задаючи явної аналітичної формули, а навчаючи її на даних.

На вході ANFIS-модель отримує набір з кількох параметрів, котрі описують ключові аспекти дезінформації:

- Контекнті фактори;
- Мережеві фактори;
- Психологічні фактори.

Контентні фактори — це характеристики самого повідомлення або новини. Серед них: правдивість / достовірність контенту (за оцінками фактчекерів або ймовірність того, що новина хибна, — наприклад, «низька» достовірність означає сильну дезінформативність); емоційна забарвленість та тональність (чи викликає страх, гнів, шок — сильні емоції сприяють віральності повідомлення); сенсаційність / провокаційність заголовків; тип контенту (текст, зображення, мем, відео — наприклад, вірaльні меми можуть швидше поширюватися). Ці фактори можуть бути агреговані у інтегральний показник, наприклад Content Risk Factor, що лінгвістично оцінюється як «низький», «середній» або «високий». Так, завідомо брехливий або маніпулятивний текст з емоційно забарвленими образами матиме високий Content Risk.

Мережеві фактори відображають, як інформація розповсюджується мережею. Сюди належать: потенційне охоплення — розмір аудиторії початкового джерела (наприклад, кількість підписників акаунту, що поширив повідомлення); швидкість розповсюдження — скільки репостів або переглядів набрано за одиницю часу після публікації; топологія мережі — наявність «центрів впливу» (hub nodes) або ехо-камер (груп вузлів, де інформація циркулює без зовнішньої перевірки) [11] координованість поширення — ознаки того, що новину розганяють боти чи скоординовані облікові записи (синхронність постів, спільні хештеги тощо). Ключовим інтегральним показником тут виступає епідеміологічний параметр R_0 — базове число репродукції дезінформації. Його можна оцінити на ранній стадії, вимірявши, скільки вторинних поширень генерує одне поширення: наприклад, якщо один користувач, що побачив фейк, ділиться ним з двома іншими, то попередньо $R_0 \approx 0$. Крім R_0 можна використовувати коефіцієнт інфікування β (ймовірність передачі при одному «контакті», аналізуючи частку переглядів, що перетворилися на репости). Мережеві фактори у моделі можуть бути згорнуті до показника Network Spread Factor з лінгвістичними значеннями «повільне/локальне», «помірне», «швидке/масове» поширення.

Група психологічних факторів враховує сприйнятливість аудиторії та контекст сприйняття інформації. Сюди входить рівень довіри населення до офіційних джерел (низька довіра підвищує ризик, що люди повірять альтернативним, навіть сумнівним повідомленням), поширеність когнітивних упереджень та стереотипів (наприклад, confirmation bias — схильність вірити інформації, яка підтверджує існуючі переконання, чи illusory truth effect — схильність вважати повторювані твердження правдивими;

емоційний стан аудиторії (страх, гнів під час кризи збільшує уразливість до неправдивих «заспокійливих» чи навпаки панічних вістей) та рівень цифрової грамотності і критичного мислення. Також важливими є емоційний та соціальний інтелект самих користувачів. Дослідження показують, що люди з нижчим рівнем емоційного та соціального інтелекту частіше бездумно поширюють фейки [12]. Для моделі ці аспекти можна об'єднати у показник Audience Susceptibility (вразливість аудиторії), значення якого оцінюються як «низька», «середня», «висока» вразливість до дезінформації.

Іноді до уваги можна взяти важкість потенційних наслідків (наприклад, наскільки серйозну шкоду може заподіяти віра у цю дезінформацію — здоров'ю, суспільному порядку, обороноздатності). Цей фактор впливає на трактування RiskScore: навіть якщо ймовірність поширення невелика, але можливі наслідки катастрофічні, можна трактувати такий фейк як високоризиковий. Цей аспект за потреби враховується через окремий модуль експертної оцінки або через модифікацію вихідної функції належності ризику.

Обрана нейро-нечітка система має багатошарову архітектуру із п'яти шарів, типову для ANFIS [10]. На рис. 1 показано блок-схему цієї архітектури, адаптовану під поставлену задачу.

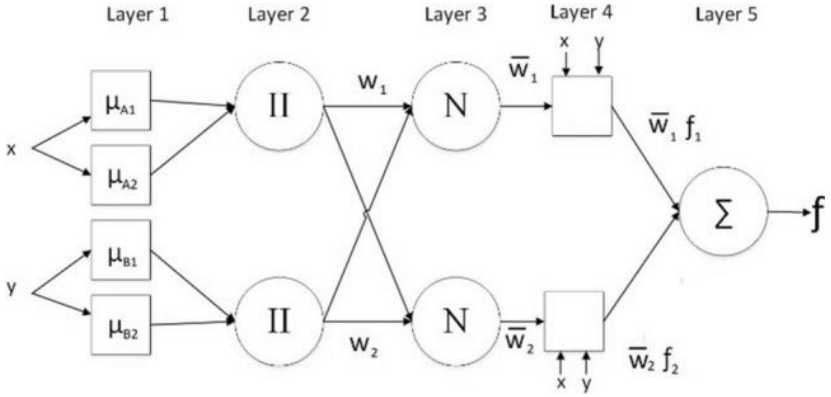

Рис. 1. Архітектура ANFIS

Коротко розглянемо функції шарів. Шар 1 це фазифікація, кожен вузол цього шару відповідає за обчислення ступеня приналежності значення вхідної змінної до відповідного нечіткого терму. Для кожного вхідного фактору визначено кілька нечітких множин. Наприклад, для Content Risk Factor визначені терми «Low», «Medium», «High» (низький, середній, високий ризик контенту); для Network Spread — «Localized», «Moderate», «Viral» (локальний, помірний, віральний); для Audience Susceptibility — «Low», «Medium», «High» (низька, середня, висока уразливість). Вузли шару 1 обчислюють значення функцій приналежності $\mu(x)$ для кожного терму. Як функції приналежності можуть бути використані гаусові, трикутні, трапецоїдальні тощо. Наприклад, для параметра R_0 можна ввести терми: $R_0 < 1$ (низький, «Contained»), $R_0 \approx 1$ («Marginal»), $R_0 > 1$ (високий, «Outbreak») — з відповідними плавними функціями приналежності.

У другому шарі здійснюється агрегація умов правил. Кожен вузол відповідає окремому нечіткому правилу, і виконує операцію кон'юнкції (AND) над усіма вхідними передумовами правила. Результат на виході вузла — це сила спрацювання (firing strength) правила, що чисельно дорівнює, наприклад, добутку (для мультиплікативної моделі) або мінімуму (для моделі Мін-макс) значень приналежності всіх умов.

Наприклад, розглянемо одне з можливих правил: «IF ContentRisk є High AND NetworkSpread є Viral AND AudienceSusceptibility є High THEN RiskScore є High». Це правило спрацює сильно (близько до 1) лише тоді, коли всі три умови мають високі ступені істинності. Для кількісного виходу правила ANFIS зазвичай використовує модель Суґено: наслідок правила — це не просто лінгвістична категорія, а функція від входів. У найпростішому випадку — нульового порядку — висновок правила задається константою (наприклад, при виконанні вищезгаданого правила RiskScore = 0.9 у шкалі 0–1). В більш складному випадку першого порядку — наслідок лінійно залежить від входів $f_k(x_1,\ldots,x_n) = a_k + b_{k1}x_1 + \ldots + b_{kn}x_n$. У нашій моделі доцільно використовувати нульовий або перший порядок залежно від бажаного рівня інтерпретованості: нульовий дає зрозумілий фіксований внесок правила, перший — дозволяє тонше налаштувати вихід в рамках правила.

Третій шар — це нормалізація. Кількість правил у системі дорівнює декартовому добутку кількостей нечітких термів усіх входів (якщо не застосовано оптимізацію). Щоб уникнути «вибуху» кількості правил при багатьох входах, застосовуються методи скорочення, як-от кластеризація або експертне виключення малозначущих комбінацій. Незалежно від кількості, виходи всіх правил у шарі 2 підлягають нормалізації в шарі 3. Тут кожен вузол обчислює нормалізовану силу правила:

$$\bar{w}k = \frac{w_k}{\sum jw_j}, \quad (1)$$

де w_k сила спрацювання k-го правила. Цей крок забезпечує, що ваги правил далі інтерпретуються як відносні (сума нормалізованих ваг = 1).

Вузли четвертого шару обчислюють зважені внески правил у вихід. Кожен вузол k приймає на вхід нормалізовану силу $\bar{w}k$ з шару 3 і обчислює вихід

$$y_k = \bar{w}k \cdot f_k(X), \quad (2)$$

де $f_k(x)$ — функція наслідку правила k (вказана в шарі 2). Якщо ми використовуємо нульовий порядок, то $f_k(x) = c_k$ (константа, що відображає «RiskScore, внесений правилом k»). Якщо першого порядку

$$f_k(X) = a_k + \sum b\,kix_i \quad (3)$$

то c, a_k, bki є настроюваними параметрами моделі — саме їх значення буде оптимізуватися під час навчання ANFIS, щоб найкраще відповідати навчальним даним. Шар 5 — останній шар, в якому один вузол сумує всі входи від шару 4: $RiskScore = \sum y_k = \sum \bar{w}_k\, f_k(X)$. Таким чином отримується єдиний числовий вихід — оцінка ризику дезінформації. Власне, для Sugeno-типу систем ця сума вже є дефазифікованим результатом. RiskScore інтерпретується, наприклад, у межах 0–1, де 1 (або 100%) відповідає максимально можливому ризику (надзвичайно вірална та небезпечна дезінформація), а 0 — мінімальному (нешкідлива або непоширювана інформація).

Унікальною властивістю ANFIS є її здатність навчатися на основі даних, автоматично підлаштовуючи як параметри функцій приналежності (передумови правил), так і параметри висновків правил. Навчання відбувається з використанням гібридного алгоритму: поєднання зворотного поширення помилки (backpropagation) для налаштування параметрів функцій приналежності та методу найменших квадратів для оптимізації лінійних параметрів висновків. Так, якщо у нас є набір тренувальних прикладів:

$$X^{(p)}, y_{true}^{(p)} \quad (4)$$

де $X^{(p)}$ вектор вхідних факторів для p-тої дезінформації, а y_{true} — відома оцінка ризику або інший таргет (можливо, апроксимація фактичного охоплення чи бальної

оцінки експертів), то ANFIS мінімізує суму квадратів помилок $\sum\left(y_{true}^{(p)} - y_{pred}^{(p)}\right)^2$ шляхом поступового коригування своїх параметрів. На практиці, через обмеженість реальних розмічених даних про дезінформаційні атаки, можливе залучення синтетичних даних для навчання. Ми генеруємо синтетичні сценарії (використовуючи, наприклад, агент-орієнтовані моделі поширення в соціальних мережах), варіюючи контентні, мережеві та психологічні параметри, й оцінюємо для них «істинний» RiskScore за допомогою моделювання. Такий підхід дозволяє отримати достатню кількість навчальних прикладів навіть при дефіциті реальних подій, зберігаючи при цьому правдоподібність ситуацій. Літературні джерела підтверджують ефективність застосування синтетичних даних для навчання ANFIS у випадках браку емпіричних вибірок [13], [14].

Варто зазначити, що система може починати з експертно визначених правил (наприклад, очевидних правил на кшталт «якщо R_0 високий і довіра низька, ризик високий») — це полегшує початкову інтерпретацію. Далі ANFIS шляхом навчання може уточнювати ці правила або ваги. Такий підхід дає баланс між експертною інтуїцією та даними. Крім того, використання механізмів скорочення правил (через попереднє групування або видалення незначущих правил) допомагає побороти комбінаторний вибух числа правил при великій кількості факторів.

АНАЛІЗ МАТЕМАТИЧНОЇ МОДЕЛІ ANFIS

Для формалізації викладемо математичну модель ANFIS у термінах вхідних змінних, нечітких множин, правил та алгоритмів навчання. Нехай $x_1, x_2, \ldots, x_m$ — вхідні змінні моделі (у нашому випадку m може дорівнювати 3 або 4: ContentFactor, NetworkFactor, AudienceFactor, і опційно епідеміологічний R_0 як окрема змінна). Кожна змінна x_i визначена на універсальній множині U_i і розбита на кілька термів (лінгвістичних значень) з відповідними нечіткими підмножинами

$$A_{i1}, A_{i2}, \ldots, A_{iLi} \tag{5}$$

де L_i кількість термів для змінної x_i. Кожна нечітка множина характеризується функцією приналежності $\mu A_{ij}(xi) \in [0,1]\mu$, що визначає ступінь відповідності значення x_i цьому терму. Наприклад, для змінної $X_{Content}$ із термами A_{cont}^1 “Low”, A_{cont}^2 = “Medium”, A_{cont}^3 = «High» маємо функції $\mu_{LowCont}(x_{Content}), \mu_{MedCont}(x_{Content}), \mu_{HighCont}(x_{Content})$, що набувають значень 1 при відповідно дуже низькому, середньому, дуже високому значенні контентного фактору, і плавно змінюються між 0 та 1 для проміжних значень. Наприклад, можна задати гаусову функцію приналежності:

$$\mu_{LowCont}(x) = expr\left(\frac{(x-c)^2}{2\sigma^2}\right) \tag{6}$$

де c — центр, σ — ширина функції. Параметри функцій приналежності σ_{ij}, cij є настроюваними і налаштовуються у процесі навчання.

Кожна можлива комбінація термів вхідних змінних визначає правило типу Такані–Суґено. Якщо припустити, що використовуються всі комбінації (повна база правил), то загальна кількість правил $N_{rules} = \prod_{i=1}^{m} L_i$. Однак, як згадано, на практиці база правил може бути оптимізована/скорочена. Запишемо загальний вигляд правила k (зі скороченим позначенням термів):

R_k: $IFx_1isA_1(k)$AND$x_2isA_2(k)$AND … AND$x_misA_m(k)THENyk = f_k(x_1, x_2, \ldots, x_m)$.

Тут A_i^k — вибраний терм для x_i в правилі k. Висновок fk(·)fk(·) є функцією від входів. У Sugeno-моделі першого порядку береться лінійна функція:

$$f_k * (x_1, \dots, x_m) = a_k + \sum_{i=1}^{m} b_{ki}\, x_i \quad (7)$$

де параметри a_k, b_{ki}, — постійні (коефіцієнти), специфічні для правила k. У нульового порядку спрощено:$f_k = const = c_k$ (тобто $a_k = c_k$, а всі $b_{ki} = 0$). Наш вихід y системи — це RiskScore.

В системі ANFIS зі структурою, описаною вище, вихід обчислюється за формулою:

1. Визначення сил спрацювання правил: для кожного правила k обчислюється вага w_k як добуток (або інша t-норма) всіх відповідних значень приналежності:

$$w_k = \mu_{(A1(k))}(x_1) \cdot \mu_{(A2(k))}(x_2) \dots \mu_{(Am(k))}(x_m) \quad (8)$$

Наприклад, для правила «R_k:IFx_1 є High ANDx_2 є Viral THEN …» маємо

$$w_k = \mu High(x1) \cdot \mu Viral(x_2) \cdot \dots \quad (9)$$

і т.д. $w_k \in [0,1]$ показує, наскільки вхід $X = (x_1, \dots, x_m)$ відповідає умовам правила R_k.

2. Нормалізація: обчислюються нормалізовані ваги

$$\bar{w}k = \frac{w_k}{\sum \frac{N_{rules}}{j=1} w_j} \quad (10)$$

Це означає, що $\sum \bar{w_k} \in [0,1]$. Якщо всі $wj = 0$ (жодне правило не спрацювало) — ситуація малоймовірна, бо зазвичай вхід належить хоча б одній нечіткій комбінації; але тоді RiskScoreRiskScore може братись як 0 або наперед задане значення.

3. Обчислення вкладень правил у вихід: для кожного правила k рахується зважений вклад $w_k \cdot f_k(x_1, \dots, x_m)$. З урахуванням нормалізації можна писати $\bar{w_k} f_k(\cdot)$ — це вихід четвертого шару.

4. Отримання остаточного виходу: це сума вкладень всіх правил:

$$y_{out} = \sum_{k=1}^{Nrules} \bar{w_k}\, f_k(x_1, \dots, x_m) \quad (11)$$

Підставляючи f_k, маємо (для моделі першого порядку):

$$\mathrm{y_{out}} = \sum_{\mathrm{k=1}}^{\mathrm{Nrules}} \mathrm{\bar{w}\,k} \left((\mathrm{a_k}) + \sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{b_{ki}\, x_i} \right) \quad (12)$$

Це і є формула нечіткого висновку Суґено, яка дає безпосередньо чітке значення RiskScoreRiskScore на виході. Відзначимо, що при нульовому порядку $f_k = c_k$, формула спрощується до

$$y_{out} = \sum_{K} \bar{w_k}\, c_k \quad (13)$$

тобто середньозважене значення вихідних констант правил з вагами $\bar{w_k}$ — типова дефазифікація методом центру ваги для дискретного випадку (аналог Мамдані з центроїдом, але підрахованим ефективно через Sugeno-константи).

ПОРІВНЯННЯ МОДЕЛІ З КЛАСИЧНИМИ МЕТОДАМИ

Запропонований підхід слід співставити з альтернативними (більш традиційними) методами оцінки та прогнозування ризиків поширення дезінформації: лінійними моделями, епідеміологічними SIR-подібними моделями та класичними нечіткими системами (типу Мамдані).

Лінійна модель могла б, приміром, обчислювати RiskScore як зважену суму факторів. Такий підхід є простим та інтерпретованим, але він припускає адитивність і незалежність впливів. У реальності вплив факторів не є лінійно незалежним: наприклад, високий R_0 особливо небезпечний саме тоді, коли контент неправдивий — тут виникає ефект взаємного підсилення (нелінійна взаємодія). Лінійна модель не зможе врахувати порогові явища (як стрибок ризику при $R_0 > 1$) чи логічні умови (як «ризик низький, якщо хоча б один із ключових факторів низький, незалежно від інших»). Нечіткі правила природним чином моделюють такі нелінійні взаємодії та умовні залежності. Наприклад, правило з кон'юнкцією фактично реалізує логічне «і» між ознаками — те, чого лінійна комбінація не зробить. Тому ANFIS переважає лінійні методи за здатністю описати складні сценарії. Щодо точності, то для складних процесів (як поширення фейків) нейро-нечіткі моделі зазвичай точніші після навчання, ніж лінійні регресії. Лінійні моделі можуть недооцінювати ризик у критичних випадках (бо усереднюють вплив) або переоцінювати у неважливих комбінаціях, тоді як ANFIS диференціює ці ситуації.

Епідемічні моделі розглядалися у ряді досліджень для опису поширення чуток і дезінформації. Вони вводять змінні типу S (вразливі), I (інфіковані дезінформацією), R («вилікувані» — хто більше не ведеться або хто поінформований) та диференціальні рівняння для їх взаємодії. Перевага таких моделей — простота та теоретична обґрунтованість: можна вивести явні формули для порогів (R_0) та динаміки. Проте вони мають ряд допущень: однорідне змішування популяції (всі контактують з усіма з рівною ймовірністю), однаковість «штаму» інформації (не враховується контент — будь-яка дезінформація моделюється однаково), обмежена кількість станів аудиторії (нехтуються проміжні стани типу сумнівного ставлення, емоційних факторів, різних рівнів сприйнятливості). У складніших варіантах дослідники додавали, наприклад, стан «Doubter» (сумнівається) чи розбивали інфікованих за рівнем емоційного забарвлення [15], але це ускладнює модель і потребує більше параметрів, які важко оцінити на практиці. Наша ж модель ANFIS фактично навчається на даних без потреби явно задавати всі параметри диференційних рівнянь. Вона може апроксимувати поведінку SIR-моделі, якщо дані це відображають (наприклад, вивчить значення R_0 як сильний фактор), але також врахує інші виміри, недоступні SIR. Таким чином, ANFIS-гнучкість вища: ми можемо включити ефект ехо-камер чи різних когнітивних груп просто як додаткові вхідні змінні, а не перебудовуючи всю модель. Слід зауважити, що епідеміологічні моделі добре дають прогноз динаміки в часі (наприклад, коли пік «інфодемії»), тоді як ANFIS наразі ми використовуємо для статичної оцінки ризику конкретного вкиду. Однак, RiskScore можна інтерпретувати як прогнозований рівень охоплення чи впливу, тож опосередковано він узгоджується з поняттями SIR (наприклад, високий RiskScore прогноз великого I_{max} у моделі поширення). Отже, ANFIS-підхід перевершує SIR за широтою врахування факторів, хоча SIR-модель залишається цінною для теоретичного розуміння і верифікації (наприклад, можна симулювати SIR для оцінки R_0, який потім подати в ANFIS як вхід).

Класична нечітка логіка (типу Мамдані) теж використовується для оцінки ризиків у різних доменах завдяки здатності працювати з лінгвістичними змінними. Наприклад, можна вручну задати правила: «Якщо контент дуже брехливий І поширення швидке, ТО ризик високий» тощо. Така система дасть вихід після дефазифікації (скажімо, методом центру ваги). Основна різниця: Мамдані-система не навчається автоматично. Правила і функції приналежності визначаються експертами. У простих випадках (кілька факторів) це можливо; але у складних (як наш, з 3–4 факторами та багатьма градаціями) експерт може не врахувати всі комбінації або неточно налаштувати функції. ANFIS же

автоматично підлаштовує як форму функцій (фазифікацію), так і наслідки правил під конкретні дані, що підвищує точність і адаптивність. Наприклад, статична нечітка система може містити правило «якщо аудиторія середньо-вразлива, то ризик середній» незалежно від інших умов, що може спростити логіку надміру. Натомість ANFIS на основі даних може з'ясувати, що навіть середньо вразлива аудиторія дасть високий ризик, якщо R_0 і контент надзвичайні — і відкоригує відповідні наслідки правил. Ще один момент — масштабованість. При зростанні числа факторів або їх градацій, ручне налаштування Мамдані-системи стає дуже громіздким, і продуктивність може страждати. ANFIS надає способи скорочення правил і оптимізації, краще впоруючись зі збільшенням розмірності (хоча теж не безмежно). Дослідники відзначають, що ANFIS виявляється більш гнучким і масштабованим рішенням у задачах оцінки ризиків порівняно з традиційною нечіткою логікою. З іншого боку, Мамдані-моделі іноді вважаються більш інтерпретованими, оскільки їх правила — чисто лінгвістичні, а не з числовими висновками. Проте і в ANFIS можна інтерпретувати правила, особливо якщо використано нульовий порядок (константні висновки, які легко співвідносяться з лінгвістичними оцінками типу «високий ризик»).

Нейро-нечітка модель поєднує позитивні сторони класичних підходів, мінімізуючи їх недоліки. Вона, як і SIR-моделі, враховує ключові порогові параметри типу R_0, але не потребує спрощених припущень про однорідність населення чи однаковість інформації, бо може брати додаткові змінні. Вона, як і нечіткі системи Мамдані, здатна оперувати з невизначеністю та експертними знаннями, але при цьому додає механізм навчання і автоналаштування, що суттєво підвищує адаптивність до нових даних та атак. Розглянемо таблицю порівняння ключових переваг та недоліків досліджених моделей.

Таблиця 1

Ключові переваги та недоліки моделей

Модель	Переваги	Недоліки
ANFIS	Гнучко поєднує різнорідні фактори та нелінійні ефекти у моделі; здатна навчатися на даних (автоматично налаштовує правила під конкретні випадки, підвищуючи точність).	Потребує достатнього набору даних для тренування моделей; архітектура складніша за прості аналогіки (менш прозора, інтерпретація правил ускладнена при великій їх кількості); можлива перенавченість, якщо дані неповні.
Лінійна	Простота реалізації та використання; висока інтерпретованість (коефіцієнти прямо показують вплив кожного параметра); обчислювальна ефективність і мінімальні вимоги до даних.	Занадто спрощує реальність — припускає лише лінійну суму впливів без врахування порогів і взаємодій; може давати значну похибку на нелінійних процесах; не здатна моделювати складні умовні логічні залежності між факторами.
SIR-модель	Теоретично обґрунтована модель із галузі епідеміології; дозволяє описати динаміку поширення (визначати порогове значення R_0, прогнозувати пікові значення «інфодемії»); має невелику кількість параметрів і зрозумілу структуру.	Базується на ряді спрощувальних припущень (однорідність контактів у популяції, однаковий «штам» дезінформації для всіх); не враховує особливостей контенту чи психології аудиторії без суттєвого ускладнення моделі; обмежена кількість станів (ігнорує проміжні стани типу «сумнівається» без додаткових розширень моделі).
Мамдані FIS	Правила у лінгвістичній формі, зрозумілі експертам і користувачам; прямо враховує невизначеність і нечіткість даних через функції приналежності; може моделювати складні нелінійні залежності на основі експертних знань («IF-THEN» логіка).	Відсутнє автоматичне навчання — побудова і налаштування правил вимагають участі експерта; погано масштабується при збільшенні числа змінних або термів (правил стає дуже багато, система ускладнюється і може втрачати продуктивність); точність і актуальність моделі залежать від повноти експертних знань, а не від даних.

У порівнянні з лінійними моделями, ANFIS краще відображає нелінійну природу інформаційних процесів. Тому для задачі оцінки ризику дезінформації, особливо в умовах інформаційної війни, де ставки високі і ситуації унікальні, запропонований нейро-нечіткий підхід виглядає більш ефективним та надійним.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

У результаті проведеного дослідження була досліджена та розроблена методика нейро-нечіткої моделі ANFIS для оцінювання ризику поширення дезінформації в умовах інформаційної війни. Модель успішно інтегрує три основні групи чинників: контентні, мережеві та психологічні аспекти, що дозволяє створити єдиний показник RiskScore для вимірювання рівня ризику дезінформаційної кампанії. Основною особливістю запропонованої моделі є її здатність адаптуватися до нових умов завдяки використанню нечіткої логіки Суґено та нейронних мереж для автоматичного навчання параметрів. Це дає змогу ефективно оцінювати ризик навіть за умов обмежених або синтетичних даних. Розроблена методика показала перевагу перед традиційними методами, такими як лінійні моделі або епідеміологічні SIR-моделі, оскільки враховує неелінійні взаємодії між факторами ризику і дозволяє працювати з комплексними даними, що охоплюють психологічні аспекти, мережеві ефекти і контентні характеристики.

Перспективи подальших досліджень включають кілька напрямів, наприклад, розширення моделі для врахування динаміки зміни RiskScore в часі. Це дозволить моделювати не тільки оцінку поточного ризику, а й прогнозувати зміну ситуації, відстежуючи етапи еволюції дезінформаційної кампанії. Також це дозволить продовжити роботу над реалізацією навчання моделі, тестуванні на синтетичних даних. Загалом, досліджена та розроблена методика має значний потенціал для застосування у реальних умовах інформаційних війн та для запобігання розповсюдженню дезінформації на ранніх етапах її поширення.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Guzmán Rincón, A., Barragán Moreno, S., Rodríguez-Canovas, B., & Mondragón-Cardona, A. (2023). *Social networks, disinformation and diplomacy: A dynamic model for a current problem.* Humanities and Social Sciences Communications, *10(1),* 1–14. https://doi.org/10.1057/s41599-023-01998-z
2. van der Linden, S. (2022). Misinformation: Susceptibility, spread, and interventions to immunize the public. *Nature Medicine, 28,* 460–467. https://doi.org/10.1038/s41591-022-01713-6
3. Maleki, M., Mead, E., Arani, M., & Agarwal, N. (2021). *Using an epidemiological model to study the spread of misinformation during the Black Lives Matter movement* [Preprint]. arXiv. https://arxiv.org/abs/2103.12191
4. Govindankutty, S., & Gopalan, S. P. (2024). Epidemic modeling for misinformation spread in digital networks through a social intelligence approach. *Scientific Reports, 14, 19100*. https://doi.org/10.1038/s41598-024-69657-0
5. Ravichandran, B. D., & Keikhosrokiani, P. (2023). Classification of COVID-19 misinformation on social media based on neuro-fuzzy and neural network: A systematic review. *Neural Computing and Applications, 35(1),* 699–717. https://doi.org/10.1007/s00521-022-07797-y
6. Kozhukhivskyi, A. D., & Kozhukhivska, O. A. (2022). Developing a fuzzy risk assessment model for ERP-systems. *Radio Electronics, Computer Science, Control, (1), 12*. https://doi.org/10.15588/1607-3274-2022-1-12
7. *Adaptive neuro-fuzzy inference system.* (n.d.). Wikipedia. https://en.wikipedia.org/wiki/Adaptive_neuro_fuzzy_inference_system

8. Pérez-Pérez, E.-J., López-Estrada, F.-R., Puig, V., & Ocampo-Martínez, C. (2022). Fault diagnosis in wind turbines based on ANFIS and Takagi–Sugeno interval observers. *Expert Systems with Applications, 206, 117698.* https://doi.org/10.1016/j.eswa.2022.117698
9. Toliupa, S. V., & Kulko, A. A. (2025). Neuro-fuzzy intrusion-detection system for a critical-infrastructure information network. *Cyberbezpeka: Osvita, Nauka, Tekhnika, 3(27),* 233–247. https://doi.org/10.28925/2663-4023.2025.27.750
10. Lakhno, V. K., Blozva, A. V., Chasnovskyi, Ye. O., & Biriukov, O. A. (2021). Information security audit based on a neuro-fuzzy system. *Technical Sciences and Technologies, (3)(25),* 125–137. https://doi.org/10.25140/2411-5363-2021-3(25)-125-137
11. Malyar, M. M., Polishchuk, A. V., Polishchuk, V. V., & Sharkadi, M. M. (2019). Neuro-fuzzy model for multicriteria evaluation. *Radioelektronika, Informatyka, Upravlinnia, (4),* 83–91. https://doi.org/10.15588/1607-3274-2019-4-8
12. Rakytianska, H. B. (2016). Hierarchical neuro-fuzzy inverse-inference model for tuning classification-rule structures. *Informatsiini Tekhnolohii ta Kompiuterna Inzheneriia, 34(3),* 94–99. https://itce.vntu.edu.ua/article/view/214
13. Dmytriienko, K. O., & Korshun, N. I. (2025). Application of improved Kuramoto models for disinformation identification in social networks. *Cyberbezpeka: Osvita, Nauka, Tekhnika, 3(27).* https://doi.org/10.28925/2663-4023.2025.27.754
14. Lavrov, V. V., & Dudatiev, A. V. (2025). Review of existing methods for assessing disinformation risks in hybrid warfare. *Cyberbezpeka: Osvita, Nauka, Tekhnika, 3(27),* 248–256. https://doi.org/10.28925/2663-4023.2025.27.720
15. Obodiak, V. K., & Shelekhov, I. V. (Eds.). (2021). *Modern information technologies in cybersecurity* [Monograph]. Sumy State University. https://essuir.sumdu.edu.ua/bitstream/123456789/82619/3/Obodiak_kiberbezpeka.pdf

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

Vadym Lavrov
Postgraduate Student of the Department of Information Security
Vinnytsia National Technical University, Vinnytsia, Ukraine
ORCID ID: 0009-0005-7031-3891
vadym@lavrovv@gmail.com

Andrii Dudatyev
Ph.D., Associate Professor, Associate Professor of the Department of Information Security
Vinnytsia National Technical University, Vinnytsia, Ukraine
ORCID ID: 0000-0002-7944-2404
dydatyev.av@gmail.com

Volodymyr Harnaha
Associate Professor, Department of Information Protection
Vinnytsia National Technical University, Vinnytsia, Ukraine
ORCID ID: 0009-0000-9180-6445
garnaga.volodymyr@vntu.edu.ua

NEURO-FUZZY ANFIS SYSTEM FOR ASSESSING THE RISK OF DISINFORMATION UNDER INFORMATION-WARFARE CONDITIONS

Abstract. The article analyses and develops a methodology for using a neuro-fuzzy ANFIS system to assess the risk of disinformation during information warfare. We propose integrating content-based, network and psychological factors to calculate a single risk indicator (RiskScore) for disinformation spread. The model's input variables include content characteristics (e.g., truthfulness or emotional tone), network metrics (e.g., propagation speed, reproduction number R_0), and audience psychology (trust level, susceptibility to manipulation, etc.). These variables feed an adaptive fuzzy system that applies a rule-based IF-THEN framework to produce a fuzzy logic risk evaluation. The ANFIS architecture can be trained on limited or synthetic data by tuning rule parameters and membership functions to enhance accuracy. The approach demonstrates an ability to combine heterogeneous factors into a unified metric, enabling early detection of potentially dangerous disinformation campaigns. Combining epidemiological indicators (such as R_0) with content-psychological features through fuzzy logic improves the informativeness of risk estimates, especially in information-warfare settings. Deploying an ANFIS-based risk-assessment model can help government and defence bodies prioritise counter-disinformation efforts, outperforming traditional linear models, epidemiological SIR models and classical Mamdani fuzzy systems in flexibility. The results provide a foundation for further model development.

Keywords: disinformation risk; neuro-fuzzy system; ANFIS; risk assessment; information warfare; disinformation modelling; information space; cybersecurity.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Guzmán Rincón, A., Barragán Moreno, S., Rodríguez-Canovas, B., & Mondragón-Cardona, A. (2023). *Social networks, disinformation and diplomacy: A dynamic model for a current problem.* Humanities and Social Sciences Communications, *10(1),* 1–14. https://doi.org/10.1057/s41599-023-01998-z
2. van der Linden, S. (2022). Misinformation: Susceptibility, spread, and interventions to immunize the public. *Nature Medicine, 28*, 460–467. https://doi.org/10.1038/s41591-022-01713-6
3. Maleki, M., Mead, E., Arani, M., & Agarwal, N. (2021). *Using an epidemiological model to study the spread of misinformation during the Black Lives Matter movement* [Preprint]. arXiv. https://arxiv.org/abs/2103.12191
4. Govindankutty, S., & Gopalan, S. P. (2024). Epidemic modeling for misinformation spread in digital networks through a social intelligence approach. *Scientific Reports, 14, 19100*. https://doi.org/10.1038/s41598-024-69657-0
5. Ravichandran, B. D., & Keikhosrokiani, P. (2023). Classification of COVID-19 misinformation on social media based on neuro-fuzzy and neural network: A systematic review. *Neural Computing and Applications, 35(1),* 699–717. https://doi.org/10.1007/s00521-022-07797-y

6. Kozhukhivskyi, A. D., & Kozhukhivska, O. A. (2022). Developing a fuzzy risk assessment model for ERP-systems. *Radio Electronics, Computer Science, Control, (1), 12*. https://doi.org/10.15588/1607-3274-2022-1-12
7. *Adaptive neuro-fuzzy inference system.* (n.d.). Wikipedia. https://en.wikipedia.org/wiki/Adaptive_neuro_fuzzy_inference_system
8. Pérez-Pérez, E.-J., López-Estrada, F.-R., Puig, V., & Ocampo-Martínez, C. (2022). Fault diagnosis in wind turbines based on ANFIS and Takagi–Sugeno interval observers. *Expert Systems with Applications, 206, 117698.* https://doi.org/10.1016/j.eswa.2022.117698
9. Toliupa, S. V., & Kulko, A. A. (2025). Neuro-fuzzy intrusion-detection system for a critical-infrastructure information network. *Cyberbezpeka: Osvita, Nauka, Tekhnika, 3(27),* 233–247. https://doi.org/10.28925/2663-4023.2025.27.750
10. Lakhno, V. K., Blozva, A. V., Chasnovskyi, Ye. O., & Biriukov, O. A. (2021). Information security audit based on a neuro-fuzzy system. *Technical Sciences and Technologies, (3)(25),* 125–137. https://doi.org/10.25140/2411-5363-2021-3(25)-125-137
11. Malyar, M. M., Polishchuk, A. V., Polishchuk, V. V., & Sharkadi, M. M. (2019). Neuro-fuzzy model for multicriteria evaluation. *Radioelektronika, Informatyka, Upravlinnia, (4),* 83–91. https://doi.org/10.15588/1607-3274-2019-4-8
12. Rakytianska, H. B. (2016). Hierarchical neuro-fuzzy inverse-inference model for tuning classification-rule structures. *Informatsiini Tekhnolohii ta Kompiuterna Inzheneriia, 34(3),* 94–99. https://itce.vntu.edu.ua/article/view/214
13. Dmytriienko, K. O., & Korshun, N. I. (2025). Application of improved Kuramoto models for disinformation identification in social networks. *Cyberbezpeka: Osvita, Nauka, Tekhnika, 3(27).* https://doi.org/10.28925/2663-4023.2025.27.754
14. Lavrov, V. V., & Dudatiev, A. V. (2025). Review of existing methods for assessing disinformation risks in hybrid warfare. *Cyberbezpeka: Osvita, Nauka, Tekhnika, 3(27),* 248–256. https://doi.org/10.28925/2663-4023.2025.27.720
15. Obodiak, V. K., & Shelekhov, I. V. (Eds.). (2021). *Modern information technologies in cybersecurity* [Monograph]. Sumy State University. https://essuir.sumdu.edu.ua/bitstream/123456789/82619/3/Obodiak_kiberbezpeka.pdf

(cc) BY-NC-SA
This work is licensed under Creative Commons Attribution-noncommercial-sharealike 4.0 International License.