УДК 004.91

Є. О. Фещенко; Т. М. Заболотня, канд. техн. наук, доц.

МЕТОД АВТОМАТИЗОВАНОГО ВИЯВЛЕННЯ ФІШИНГУ В ЕЛЕКТРОННИХ ЛИСТАХ НА ОСНОВІ ГІБРИДНОЇ НЕЙРОМЕРЕЖЕВОЇ АРХІТЕКТУРИ

У статті запропоновано новий метод для автоматизованого виявлення фішингового вмісту в текстовій частині електронних листів, який базується на застосуванні гібридної нейромережевої архітектури. Основу запропонованого методу складає інтеграція згорткових нейронних мереж, двонаправлених рекурентних мереж з довгою короткостроковою пам'яттю та механізму уваги. Таке поєднання дає можливість враховувати локальні та глобальні семантичні ознаки тексту одночасно, підвищуючи ефективність класифікації фішингових повідомлень порівняно з класичними методами машинного навчання та традиційними нейромережевими підходами. В роботі здійснено поглиблене дослідження впливу розмірності векторних представлень слів GloVe на результати класифікації. Проведені експерименти охоплювали кілька наборів даних різного розміру та характеру текстів, включаючи як загальні набори, так і спеціалізовані для виявлення фішингу. Аналіз показав, що оптимальний вибір розміру ембедингів суттєво впливає на здатність моделі розпізнавати контекстуальні особливості шахрайських повідомлень, що, у свою чергу, позитивно впливає на загальну точність та повноту ідентифікації. Додаткову увагу приділено використанню механізму уваги, який забезпечує автоматичне визначення важливості слів і виразів, що є ключовими для розпізнавання фішингового контенту. Цей механізм дозволяє моделі точніше концентруватися на потенційно небезпечних ознаках, зменшуючи кількість хибних спрацьовувань. Порівняльні експериментальні дослідження продемонстрували перевагу запропонованого методу над класичними моделями, такими як логістична регресія та машини опорних векторів, а також схожу або вищу ефективність порівняно з іншими сучасними нейромережевими методами, включаючи трансформерні архітектури. У статті також запропоновано рекомендації щодо практичного впровадження розробленого методу та визначено перспективні напрямки подальших досліджень, зокрема щодо адаптації методу до умов реального часу.

Ключові слова: *фішинг, електронні листи, гібридна архітектура, CNN, BiLSTM, шар уваги, семантичні ембедінги, GloVe.*

Вступ

Проблема автоматизованого виявлення фішингових електронних листів залишається однією з найактуальніших у галузі інформаційної безпеки та захисту персональних даних [1]. Фішинг – це шахрайська діяльність, що передбачає розсилання оманливих електронних повідомлень з метою отримання конфіденційної інформації, такої як паролі, номери банківських карток чи доступ до онлайн-сервісів. Незважаючи на численні технологічні досягнення, кількість і складність фішингових атак продовжують зростати, створюючи серйозні загрози як для окремих користувачів, так і для великих компаній і державних установ [2].

Автоматичне виявлення фішингових листів важливе в різних контекстах, таких як:

– безпека персональних даних користувачів: захист особистої інформації, банківських рахунків та конфіденційних відомостей є пріоритетом, оскільки наслідки успішних фішингових атак можуть бути катастрофічними для окремих осіб;

– корпоративна інформаційна безпека: великі компанії зазнають значних фінансових втрат через фішингові атаки, які націлені на дані співробітників з метою викрадення корпоративних паролів, документів або доступу до внутрішніх мереж;

– захист державних установ та критичної інфраструктури: фішингові атаки часто є першим етапом більш серйозних загроз, таких як кібершпигунство або проникнення в

системи критичної інфраструктури.

Ефективність виявлення фішингових листів ускладнюється значною різноманітністю текстових формулювань, маніпуляціями з лексикою та стилем, що унеможливлює застосування простих евристичних правил або базових статистичних методів для гарантованого виявлення. Крім того, короткі повідомлення з обмеженою кількістю слів створюють додаткові труднощі для точного аналізу семантики та контексту. Традиційні алгоритми класифікації, що базуються на статистичних характеристиках слів (наприклад, Naive Bayes, SVM, Random Forest), хоча і працюють швидко, але часто мають недостатню точність і схильні до помилкових спрацьовувань.

З іншого боку, сучасні нейромережеві моделі, такі як CNN, RNN та трансформерні архітектури (наприклад, BERT), демонструють високу ефективність у задачах оброблення природної мови, враховуючи контекстуальні особливості слів та фраз [3]. Однак, ці моделі також мають певні недоліки, такі як висока обчислювальна складність, вимогливість до обчислювальних ресурсів та велика кількість навчальних параметрів.

Актуальність роботи зумовлена необхідністю створення ефективного рішення для автоматичного виявлення фішингового змісту в електронних листах, яке не вимагає значних витрат обчислювальних ресурсів. Поєднання різних типів нейронних мереж із механізмом уваги дає змогу досягати високих показників точності.

Метою статті є підвищення точності автоматизованої ідентифікації фішингового вмісту в текстовій частині електронних листів шляхом застосування гібридної нейромережевої архітектури, яка об'єднує згорткові та двонаправлені рекурентні нейронні мережі із шаром уваги.

Огляд наявних методів та підходів

Автоматизована ідентифікація фішингового вмісту в електронних листах є предметом активних наукових досліджень в останні роки. За цей час було розроблено багато різноманітних рішень – від класичних алгоритмів на основі простих статистичних моделей до сучасних складних нейромережевих архітектур глибокого навчання. Розглянемо основні категорії цих методів та їх особливості.

Класичні статистичні методи. Найбільш поширеними у задачі виявлення фішингу є такі класичні алгоритми машинного навчання, як логістична регресія, метод опорних векторів (SVM) та Random Forest [3]. Ці методи базуються на вилученні із тексту таких статистичних ознак, як частота слів, присутність певних ключових термінів, кількість спеціальних символів та посилань. Логістична регресія має перевагу у простоті та швидкості, але її точність суттєво падає на складних і різноманітних текстах. Метод SVM дозволяє враховувати нелінійні залежності між ознаками за допомогою спеціальних ядер (kernel), проте цей підхід також вразливий до варіативності фішингових повідомлень. Метод Random Forest забезпечує кращу стійкість до незбалансованих даних завдяки ансамблевому підходу, однак його ефективність також залежить від якості ручного підбору ознак [4].

Методи на основі N-грам і сигнатур. Цей клас методів використовує текстові фрагменти (N-грами) та спеціальні сигнатури тексту для виявлення типових ознак фішингу. Метод N-грам передбачає розгляд послідовностей з N слів чи символів, що дозволяє виявляти частково змінені або шаблонні повідомлення. Коефіцієнт Жаккара, який визначає подібність через частку спільних N-грам, є одним із найпопулярніших у цьому підході. Подібні методи демонструють високу стійкість до дрібних змін у формулюваннях, але можуть бути неефективними проти повністю нових типів фішингу, що не містять характерних шаблонних ознак [5].

Векторні представлення тексту. Для задачі автоматизованого виявлення фішингових листів традиційно застосовуються статистичні методи, такі як TF-IDF (Term Frequency–Inverse Document Frequency), що дозволяють перетворити текст повідомлення у векторне

представлення. Цей метод ґрунтується на визначенні важливості слів за їх частотою у конкретному листі, порівняно з усією вибіркою повідомлень. Косинусна подібність між такими векторами надає змогу ефективно порівнювати листи й виявляти характерні ознаки фішингових повідомлень. Однак суттєвим недоліком TF-IDF у контексті фішингу є нездатність враховувати семантичні та контекстуальні особливості слів і фраз, через що фішингові листи, які використовують синоніми, змінюють стиль написання або застосовують різноманітні лексичні маніпуляції, можуть класифікуватися неправильно [6].

Семантичні моделі та ембедінги. Останнім часом активно використовуються семантичні моделі на основі ембедінгів слів, зокрема Word2Vec, GloVe та трансформерні моделі (наприклад, BERT). Ці методи дозволяють отримати семантично значущі векторні представлення слів і текстів, враховуючи їх контекст. Трансформерні моделі, такі як BERT, особливо ефективні завдяки генерації контекстуальних ембедінгів, що враховують різні значення одного і того ж слова у різних контекстах. Це дозволяє суттєво підвищити точність виявлення фішингових листів, але за рахунок значних витрат обчислювальних ресурсів [7].

Гібридні нейромережеві підходи. Для вирішення недоліків окремих моделей останні дослідження все частіше звертаються до гібридних підходів, які комбінують переваги CNN (згорткових нейронних мереж), RNN (рекурентних нейронних мереж) та шарів уваги [8]. CNN ефективно ідентифікують локальні патерни в тексті, такі як сталі фрази або часті шаблонні конструкції. Двонаправлені LSTM здатні враховувати семантичні залежності слів у двох напрямках, що є критичним для правильного розуміння контексту повідомлення. Шари уваги дозволяють нейромережам акцентувати увагу на найважливіших словах або фразах, що суттєво покращує точність класифікації. Цей комбінований підхід демонструє значні переваги перед окремими моделями, забезпечуючи баланс між точністю та обчислювальною складністю.

Таким чином, аналіз наявних методів показує, що класичні статистичні методи – швидкі та прості, але не завжди достатньо точні для складних і різноманітних фішингових повідомлень. Семантичні та нейромережеві моделі, хоч і забезпечують високу точність, часто є надто ресурсоємними. Комбінування різних підходів, зокрема гібридних архітектур нейромереж, дозволяє досягти балансу між точністю та швидкістю обробки, що актуально для практичного застосування у системах автоматизованого виявлення фішингового вмісту в електронних листах.

Запропонований комбінований метод

Для вирішення задачі автоматизованого виявлення фішингового вмісту в текстовій частині електронних листів запропоновано новий комбінований метод, що поєднує згорткові нейронні мережі (CNN), двонаправлені рекурентні нейронні мережі з довгою короткостроковою пам'яттю (BiLSTM) та механізм уваги. Така комбінація дозволяє одночасно аналізувати як локальні, так і глобальні семантичні особливості тексту, забезпечуючи високий рівень точності класифікації фішингових повідомлень.

Архітектура гібридної моделі. Архітектура нейронних мереж у запропонованому методі включає такі компоненти:

- CNN-шари призначені для вилучення локальних патернів у текстах, таких як характерні фрази та шаблонні конструкції, які є типовими для фішингових листів.
- BiLSTM-шари забезпечують аналіз послідовностей слів у двох напрямках, що дозволяє ефективно моделювати семантичні залежності між словами та враховувати контекст.
- шар уваги забезпечує динамічне виділення найбільш значущих слів та словосполучень, підвищуючи точність моделі за рахунок зосередження на семантично важливих елементах.

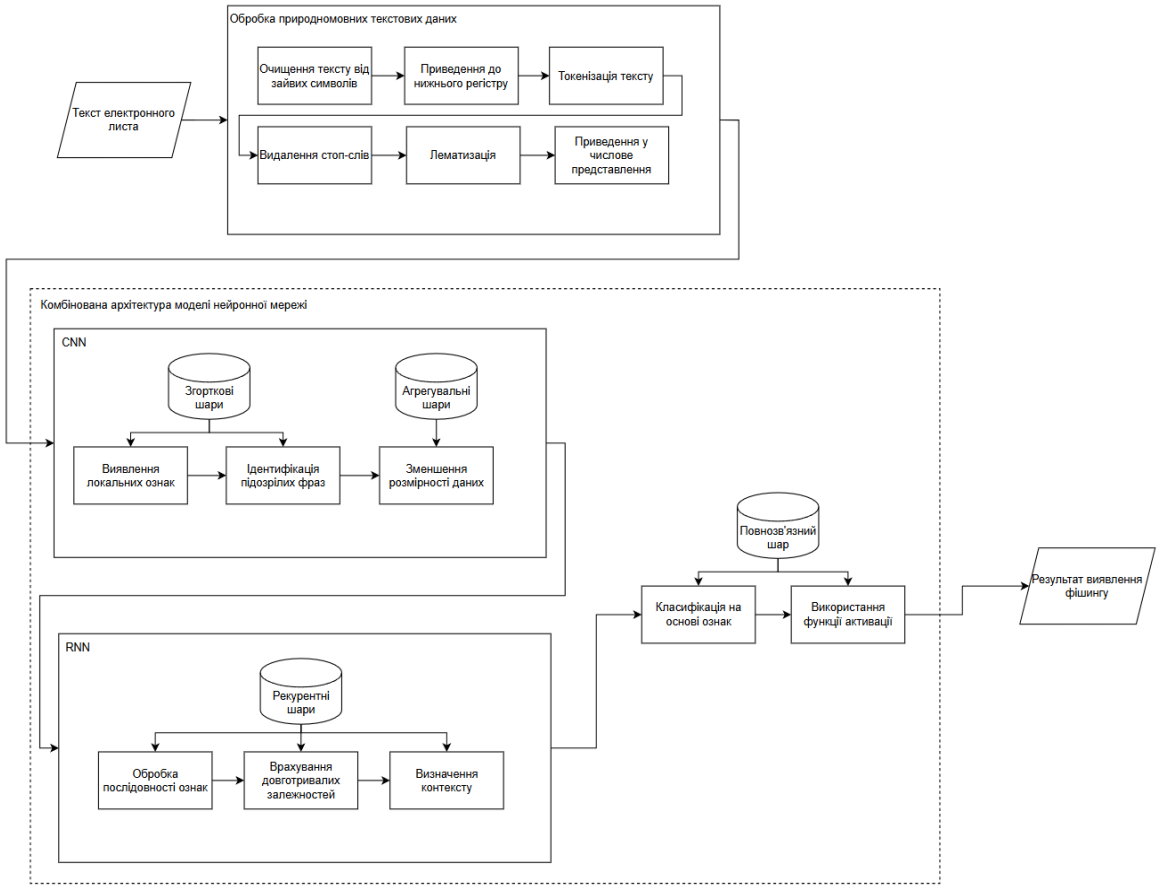

Рис. 1. Схематичне представлення роботи методу

Нижче більш детально опишемо зміст кожного з етапів запропонованого методу.

Попередня обробка тексту є ж важливим першим етапом у запропонованому методі, оскільки вона забезпечує коректну підготовку даних для ефективної роботи гібридної нейромережі. На першому етапі виконуються процедури нормалізації, що включають:

– токенізацію – процес поділу тексту на окремі слова або токени.

– видалення стоп-слів – службові слова (наприклад, «і», «в», «на», «що»), які часто зустрічаються у тексті, але мають низьку семантичну значущість, видаляються для зменшення «шуму» у векторних представленнях.

– лематизацію – приведення словоформ до їх базових форм (лем). Це дозволяє уникнути дублювання слів у різних формах (наприклад, «банківський», «банківська», «банківське» → «банківський»), що знижує розмірність і підвищує якість моделі.

– нормалізацію тексту – перетворення всього тексту у нижній регістр, вилучення спеціальних символів і цифр (за винятком випадків, коли вони є важливими ознаками фішингу), а також усунення зайвих пробілів.

Ці кроки дозволяють значно скоротити розмірність векторних представлень та зберегти чітку семантичну інформацію, що підвищує точність та швидкість класифікації текстів за допомогою гібридної моделі.

Векторизація тексту. Підготовлені текстові дані переводяться у векторне представлення за допомогою семантичних ембедінгів GloVe. Це дозволяє враховувати семантичні взаємозв'язки між словами, що підвищує подальшу класифікацію фішингового вмісту.

Виявлення фішингових ознак з використанням нейромереж. Отримані вектори текстових даних подаються до комбінованої нейромережевої моделі, що включає CNN-шари для виявлення локальних ознак та BiLSTM-шари для аналізу семантичних залежностей і контексту.

Застосування механізму уваги. Застосування шару уваги є критично важливим елементом запропонованого комбінованого методу. Цей компонент працює таким чином, що модель навчається автоматично ідентифікувати слова і словосполучення у тексті, які є найважливішими для прийняття рішення щодо класифікації листа як фішингового чи легітимного, та надавати їм більшу вагу.

Шар уваги формує вагові коефіцієнти, які автоматично підкреслюють значущість певних частин тексту, зокрема тих, які містять характерні для фішингу слова або вирази (наприклад, «терміново», «перевірка», «оновіть пароль», «безпека вашого рахунку»). В результаті, цей шар дозволяє моделі краще зосередитись на семантично важливих елементах тексту, зменшуючи вплив менш значущих слів, що суттєво знижує кількість хибних спрацьовувань.

Отже, запропонований метод поєднує нейромережеві методи CNN та BiLSTM з застосуванням шару уваги, що має забезпечувати високу точність класифікації фішингових електронних листів.

Особливості програмної реалізації методу

Мова програмування та бібліотеки. Програмна реалізація запропонованого методу виконана мовою програмування Python з використанням бібліотек TensorFlow та Keras, які є стандартом для побудови моделей глибокого навчання. TensorFlow надає зручні інструменти для ефективного створення, навчання та оцінювання нейронних мереж, а Keras забезпечує інтуїтивно зрозумілий інтерфейс для швидкої розробки нейромережевих архітектур.

Попередня обробка текстових даних. Для виконання попередньої обробки природномовних текстових даних застосовуються бібліотеки SpaCy та NLTK, які забезпечують ефективну токенізацію, лематизацію та видалення стоп-слів. Ці кроки попередньої обробки необхідні для отримання якісних та компактних векторних представлень, що є критично важливими для досягнення високих показників точності класифікації.

Семантичні ембедінги GloVe. Для забезпечення ефективності векторного представлення слів, які використовуються у запропонованому підході, застосовуються попередньо навчені семантичні ембедінги GloVe (Global Vectors for Word Representation) [9]. Вибір GloVe зумовлений здатністю враховувати не лише локальні, а й глобальні статистичні особливості використання слів у великих текстових корпусах. На відміну від класичних методів векторного представлення слів, таких як TF-IDF, які базуються виключно на частоті появи слів, ембедінги GloVe відображають глибокі семантичні взаємозв'язки між словами. Завдяки навчанню на великих обсягах текстових даних, GloVe забезпечує високий рівень узагальнення та враховує смислові особливості слів, що критично важливо для ефективного розпізнавання фішингового контенту, який часто маскується за допомогою перефразувань, синонімів та інших лексичних маніпуляцій.

Реалізація CNN. Згорткова нейронна мережа реалізована з використанням шару *Conv1D* бібліотеки Keras, який дозволяє виявляти локальні текстові патерни. Для кожного текстового фрагмента використовувалися фільтри з різними розмірами ядер (3, 4, 5), що дозволяє враховувати особливості різної довжини слів та фраз. Після згорткового шару застосовується шар *GlobalMaxPooling1D*, який зменшує розмірність вихідних даних, виділяючи найсуттєвіші ознаки.

Реалізація RNN. Для реалізації двонаправлених рекурентних мереж з довгою короткостроковою пам'яттю використовувався шар *Bidirectional LSTM* бібліотеки Keras. Такий підхід дозволяє моделі враховувати контекст слів у двох напрямках (від початку до кінця та навпаки), що забезпечує глибше розуміння семантичних зв'язків у тексті. Використання двонаправленого аналізу є доцільним для точного моделювання контексту в задачах класифікації текстів.

Реалізація механізму уваги. Для реалізації шару уваги використано власноруч

налаштований шар *Attention* на основі бібліотеки TensorFlow та Keras. Цей механізм дозволяє моделі автоматично надавати додаткову вагу найбільш значущим словам і словосполученням. Під час навчання нейромережа формує вагові коефіцієнти, що підкреслюють значущість окремих частин тексту для прийняття остаточного рішення про класифікацію. Такий підхід значно покращує точність ідентифікації фішингового контенту за рахунок ефективнішого фокусування на критичних елементах тексту.

Експериментальна оцінка

Тестові дані. Для перевірки ефективності запропонованого комбінованого методу використано три незалежні набори даних, які широко застосовуються для задач ідентифікації фішингових електронних листів написаних англійською мовою:

– Phishing Email Dataset – корпус, що містить понад 160 000 фішингових і нефішингових електронних листів. Листи мають середню довжину до 250 слів і надають достатньо текстових ознак для аналізу моделі [10].

– Phishing Email Detection by Subhajournal – великий набір даних, що містить близько 18 000 листів різної тематики та походження. Цей корпус дозволяє перевірити ефективність моделі у реалістичних умовах з великою кількістю різноманітних шаблонів та формулювань [11].

– Fraud Email Dataset – менший датасет, що містить приблизно 11 000 листів, типових для шахрайських схем. Особливістю цього набору є висока однорідність та часта повторюваність певних зловмісних шаблонів [12].

Метрики оцінювання. Для отримання об'єктивної оцінки якості реалізованого методу використовуються стандартні метрики класифікації: Accuracy (точність класифікації всіх повідомлень), Precision (частка правильно класифікованих фішингових листів серед усіх визначених моделлю як фішингові), Recall (частка виявлених фішингових листів серед усіх справжніх фішингових листів) та F1-score (гармонійне середнє між Precision і Recall). Вибір цих метрик зумовлений їх широким застосуванням та здатністю об'єктивно оцінювати ефективність і надійність моделі.

Вплив розміру ембедингів на ефективність моделі. Експериментально досліджено вплив розмірності ембедингів GloVe (50, 100, 200) на ефективність гібридної моделі. Використання Twitter-GloVe ембедингів є виправданим через те, що ці ембединги попередньо навчені на великому корпусі коротких повідомлень із соціальних мереж, що семантично і стилістично близькі до текстів фішингових листів. Виявлено, що збільшення розмірності ембедингів позитивно впливає на точність, але лише до певної межі. На рис. 2 наведено значення основних метрик:

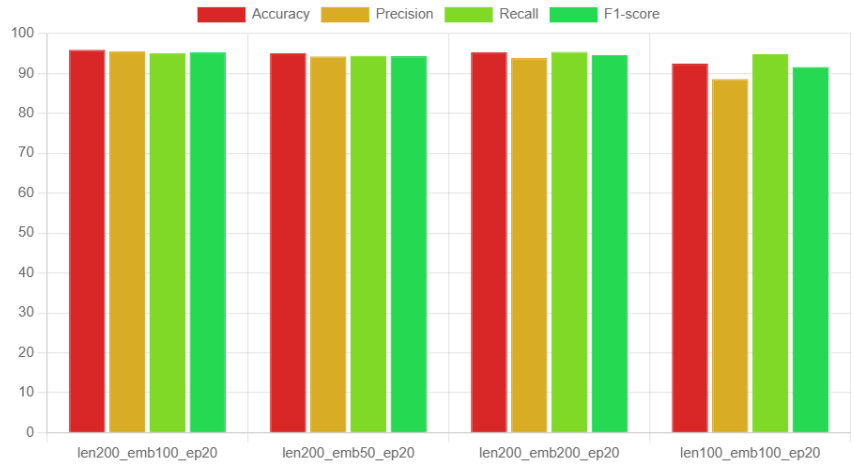

Рис. 2. Порівняння метрик ефективності моделей з різними гіперпараметрами на датасеті «Fraud Email Dataset»

Модель із 50-вимірним GloVe демонструє достатньо високу ефективність (точність

близько 95 %), але поступається у здатності чітко диференціювати близькі за значенням слова і фрази. При збільшенні розміру до 100 та 200 значень точність покращується до 97 – 99 %, що пов'язано зі зростанням кількості семантичної інформації, яка дозволяє краще враховувати контекст та тонкі семантичні нюанси. Однак, подальше збільшення розміру ембедингів (понад 200) значно збільшує витрати на обчислення та обсяг необхідної пам'яті, не забезпечуючи суттєвого приросту якості. На основі отриманих результатів доцільним є використання 200-вимірних GloVe ембедингів, оскільки вони забезпечують високий рівень точності.

Результати експериментів. Експериментально було проведено порівняння запропонованої CNN-BiLSTM моделі з шаром уваги з альтернативними підходами: окремими CNN та BiLSTM, трансформерною моделлю BERT та класичними алгоритмами машинного навчання (логістична регресія, SVM, Random Forest). На датасеті Phishing Email Dataset результати порівняння представлені в Таблиці 1.

Таблиця 1

Порівняння ефективності методів на Phishing Email Dataset

Метод	Accuracy, %	Precision, %	Recall, %	F1-score, %
Логістична регресія	94.5	91.0	86.0	88.4
SVM	95.2	92.5	87.5	89.9
Random Forest	94.0	90.1	86.8	88.4
CNN	92.8	95.0	87.0	90.8
BiLSTM	94.1	94.5	91.5	92.9
BERT	98.1	98.0	97.5	97.7
Запропонована гібридна модель	99.6	99.6	93.9	96.7

Як видно з таблиці, запропонована CNN-BiLSTM модель з шаром уваги забезпечує вищу ефективність порівняно з окремими CNN та BiLSTM моделями, а також суттєво перевершує класичні алгоритми за всіма метриками. При цьому, незважаючи на те, що трансформерна модель BERT демонструє дещо вищі абсолютні показники, запропонована гібридна модель виграє в швидкості навчання та обчислювальних ресурсах, що робить її більш практичною для реальних умов.

На великому датасеті від Subhajournal CNN-BiLSTM модель також продемонструвала високі результати: точність (Accuracy) досягає 95 %, Precision і Recall – понад 94 %, а F1-score стабільно перевищує 94 %. Це свідчить про хорошу масштабованість і узагальнюваність запропонованого підходу навіть за умов значного обсягу та різноманіття даних.

На меншому Fraud Email Dataset запропонована модель також показує високу стабільність, забезпечуючи Accuracy на рівні 94 %, Precision понад 95 %, а Recall близько 93 %, що є значним покращенням порівняно з класичними алгоритмами (наприклад, Random Forest має F1 близько 88 %, а SVM – до 90 %). Це підтверджує ефективність застосування CNN-BiLSTM архітектури з механізмом уваги навіть для специфічних і невеликих за обсягом наборів даних.

Таким чином, запропонований метод, що складається з CNN-BiLSTM моделі з шаром уваги підтверджує свою ефективність на різноманітних наборах даних, забезпечуючи стабільно високі показники точності, повноти і збалансованості результатів.

Обговорення результатів

Переваги запропонованого методу. Запропонований комбінований метод на основі поєднання CNN, BiLSTM та механізму уваги демонструє суттєві переваги порівняно з іншими підходами до автоматичної класифікації фішингових електронних листів. Однією з ключових переваг є досягнення високої точності та низьким рівнем хибнопозитивних спрацювань (Accuracy до 99 %, F1-score до 0.98), що забезпечується експериментально підібраною розмірністю GloVe-ембедингів (200-вимірних). Порівняно з трансформерними моделями на основі BERT, які потребують значних ресурсів для навчання і виконання прогнозів, гібридний підхід забезпечує конкурентну точність та узагальнювальну здатність, що підтверджено експериментами класифікації фішингу на різних незалежних корпусах [13].

Також, інтеграція шару уваги дозволяє не лише покращити ефективність класифікації, але й отримати інтерпретовані результати за рахунок візуалізації важливих слів та фраз, що є критично важливим для аналізу та подальшого вдосконалення моделі. Завдяки цій особливості метод має значний потенціал для практичного застосування у задачах моніторингу безпеки електронної пошти, забезпечуючи прозорість та зрозумілість прийняття рішень системою.

Аналіз обмежень запропонованого методу. Попри загальну високу ефективність, запропонований метод має певні обмеження. Одним із суттєвих обмежень є залежність якості класифікації від обсягу і якості тренувального датасету. У випадках, коли фішингові листи мають надто короткий текст або містять специфічну лексику, яка не представлена у навчальному наборі або в попередньо навчених ембедингах GloVe, модель може демонструвати знижену ефективність через брак семантичної інформації або недостатньо чіткі контекстуальні ознаки.

Крім того, хоча гібридна модель ефективно обробляє семантично складні повідомлення, вона може мати труднощі з класифікацією листів, у яких фішингові наміри виражені за допомогою тонких маніпуляцій. Такі нюанси потребують додаткових семантичних та контекстних даних, які можуть бути недостатньо представлені у попередньо навчених ембедингах або стандартних навчальних наборах.

Перспективи застосування. Запропонований метод має широкі перспективи для практичного застосування в системах автоматичного моніторингу безпеки електронної пошти в корпоративному та державному секторах. Завдяки високій точності та швидкості оброблення, гібридна архітектура може бути легко інтегрована у системи реального часу для моніторингу і фільтрації електронних повідомлень.

Однією з перспективних сфер є інтеграція програмного забезпечення, що реалізує метод, у хмарні сервіси та платформи електронної пошти, що забезпечить додатковий рівень захисту від фішингових атак. Також, гібридна модель може бути використана в системах аналітики загроз, де здатність чітко ідентифікувати фішингові повідомлення є важливою для швидкого реагування на загрози та попередження поширення атак у мережах.

У подальшому перспективним напрямом є адаптація моделі під конкретні галузі та спеціалізовані тематичні домени, з використанням специфічних словників і додаткових тренувальних наборів. Також, варто звернути увагу на можливості подальшого розвитку моделі через інтеграцію додаткових шарів семантичного аналізу та контекстуальних трансформерних моделей для зменшення виявлених обмежень.

Висновки

У статті запропоновано метод для автоматизованого виявлення фішингового вмісту в текстовій частині електронних листів на основі гібридної нейромережевої архітектури, заснований на поєднанні згорткових нейронних мереж, двонаправлених рекурентних мереж з довгою короткостроковою пам'яттю та механізму уваги. Запропонований метод дозволяє поєднувати переваги виявлення локальних і глобальних семантичних ознак тексту, а також

динамічно акцентувати увагу на найбільш значущих словах і виразах, що суттєво підвищує точність ідентифікації фішингових електронних листів.

Проведена експериментальна оцінка на трьох незалежних наборах даних продемонструвала високу ефективність запропонованого методу. Зокрема, на тренувальному наборі «Phishing Email Dataset» отримано точність (Accuracy) 99,6 %, Precision 99,6 % та F1-score 96,7 %. На іншому великому корпусі «Phishing Email Detection by Subhajournal» модель продемонструвала Accuracy 95 % і стабільно високий F1-score понад 94 %. На меншому наборі даних, а саме на «Fraud Email Dataset», також зафіксовано високу ефективність: Accuracy близько 94 %, Precision понад 95%, а Recall – близько 93 %, що значно перевищує показники класичних методів, таких як SVM (F1 близько 90 %) та Random Forest (F1 близько 88 %).

Проведений аналіз впливу розмірності ембедингів показав, що оптимальним для запропонованого методу є використання 200-вимірних GloVe-векторів. При використанні саме такого розміру ембедингів точність моделі зростає до 97 – 99 %, що підтверджує доцільність цього вибору з позицій ефективності та точності.

Таким чином, розроблена гібридна нейромережева архітектура є високоефективним рішенням, що досягає точності до 99,6 % у задачі автоматизованої ідентифікації фішингових повідомлень в електронних листах. Отримані результати підтверджують доцільність практичного впровадження методу в системах інформаційної безпеки для корпоративного й державного секторів. Подальші дослідження рекомендовано спрямувати на адаптацію моделі до роботи в реальному часі, інтеграцію з трансформерними архітектурами та розширення навчальних даних для покращення обробки коротких і стилістично складних текстів.

СПИСОК ЛІТЕРАТУРИ

1. Chiew K. L., Yong K. S. C., Tan C. L. A Survey of Phishing Attacks. *Expert Systems with Applications*. 2018. URL: https://www.researchgate.net/publication/324036997_A_Survey_of_Phishing_Attacks_Their_Types_Vectors_and_Technical_Approaches (дата звернення: 30.03.2025).

2. Verizon. Data Breach Investigations Report 2024. 2024. URL: https://www.verizon.com/business/resources/reports/2024-dbir-data-breach-investigations-report.pdf (дата звернення: 30.03.2025).

3. Abdelhamid N., Thabtah F., Abdel-Jaber H. Phishing Detection – Intelligent ML Comparison. *IEEE International Symposium on Information Security and Privacy*. 2017. P. 1–7. URL: https://ieeexplore.ieee.org/document/8004877 (дата звернення: 10.05.2025).

4. Khonji M., Iraqi Y., Jones A. Phishing Detection: A Literature Survey. *IEEE Communications Surveys & Tutorials*. 2013. Vol. 15, №4. P. 2091–2121. URL: https://www.researchgate.net/publication/256841808_Phishing_Detection_A_Literature_Survey (дата звернення: 10.05.2025).

5. Euna N., Hossain S., Anwar M., Sarker I. Content-based Spam Email Detection Using N-gram Machine Learning Approach. *Preprints*. 2021. DOI: 10.20944/preprints202109.0236.v1. URL: https://www.researchgate.net/publication/354603645_Content-based_Spam_Email_Detection_Using_N-gram_Machine_Learning_Approach (дата звернення: 10.05.2025).

6. Widianto A., Pebriyanto E., Fitriyanti F., Marna M. Document Similarity Using Term Frequency-Inverse Document Frequency Representation and Cosine Similarity. *Journal of Dinda: Data Science, Information Technology, and Data Analytics*. 2024. Vol. 4, №2. P. 149–153. URL: https://www.researchgate.net/publication/383453541_Document_Similarity_Using_Term_Frequency-Inverse_Document_Frequency_Representation_and_Cosine_Similarity (дата звернення: 10.05.2025).

7. Vinayakumar R., Alazab M., Soman K. P., Poornachandran P. Al-Nemrat A. DeepAnti-PhishNet: Applying Deep Neural Networks for Phishing Email Detection CEN-AISecurity@IWSPA-2018. *IEEE International Workshop on Security and Privacy Analytics (IWSPA)*. 2018. P. 77–84. URL: https://www.researchgate.net/publication/326211143_DeepAnti-PhishNet_Applying_Deep_Neural_Networks_for_Phishing_Email_Detection_CEN-AISecurityIWSPA-2018 (дата звернення: 10.05.2025).

8. Fang Y., Zhang C., Huang C., Liu L., Yang Y. Phishing Email Detection Using Improved RCNN Model With Multilevel Vectors and Attention Mechanism. *IEEE Access*. 2019. Vol. 7. P. 56329–56340. URL: https://ieeexplore.ieee.org/document/8701426 (дата звернення: 10.05.2025).

9. Pennington J., Socher R., Manning C. D. GloVe: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* (EMNLP). 2014. P 1532–1543. URL: https://www.researchgate.net/publication/284576917_Glove_Global_Vectors_for_Word_Representation (дата звернення: 16.03.2025).

10. Phishing Email Dataset. Kaggle: веб-сайт. URL: https://www.kaggle.com/datasets/naserabdullahalam/phishing-email-dataset?select=phishing_email.csv (дата звернення: 25.04.2025).

11. Phishing Email Detection Dataset. Kaggle: веб-сайт. URL: https://www.kaggle.com/datasets/subhajournal/phishingemails (дата звернення: 25.04.2025).

12. Fraud Email Dataset. Kaggle: веб-сайт. URL: https://www.kaggle.com/datasets/llabhishekll/fraud-email-dataset/data (дата звернення: 25.04.2025).

13. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference on Human Language Technologies* (NAACL-HLT). 2019. P. 4171–4186. URL: https://aclanthology.org/N19-1423.pdf (дата звернення: 16.03.2025).

Стаття надійшла до редакції 13.06.2025.
Стаття пройшла рецензування 16.06.2025.

Фещенко Єгор Олександрович – магістр кафедри програмного забезпечення комп'ютерних систем, e-mail: yehor.feschenko@gmail.com.

Заболотня Тетяна Миколаївна – канд. техн. наук, доцент кафедри програмного забезпечення комп'ютерних систем.

Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського".