КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 3 (31), 2025
ISSN 2663 - 4023
DOI 10.28925/2663-4023.2025.31.1032
УДК 004.42:004.056.5

Чернящук Наталія Леонідівна
д.пед.н., професор
Волинський національний університет імені Лесі Українки, Луцьк, Україна
ORCID: 0000-0002-3178-8377
cherniashchuk.nataliia@vnu.edu.ua

ЗАСТОСУВАННЯ НЕЙРОННИХ МЕРЕЖ ДЛЯ АВТОМАТИЗОВАНОГО РОЗПІЗНАВАННЯ ТЕКСТУ ТА СИМВОЛІВ У ЗАДАЧАХ КІБЕРБЕЗПЕКИ

Анотація. У цій роботі представлено розробку та дослідження системи оптичного розпізнавання тексту (OCR) для низькоякісних зображень із застосуванням методів машинного навчання. Для виконання завдання було створено два набори зображень у відтінках сірого: перший містив окремі символи англійського алфавіту та цифри (4 960 зображень розміром 250×50 пікселів), а другий – фрагменти осмисленого тексту з книги «Голодні ігри» (4 010 зображень розміром 680×50 пікселів). Щоб підвищити стійкість моделі, зображення було спеціально спотворено за допомогою розмиття та цифрового шуму. OCR-модель була побудована на основі поєднання згорткових нейронних мереж (CNN) і рекурентних мереж (RNN) із шаром Connectionist Temporal Classification (CTC) для корекції послідовностей. Навчання проводилося протягом 70–80 епох із розподілом даних 9:1 для навчання та валідації. Було проведено порівняльний аналіз між розробленою системою та Tesseract OCR. Експериментальні результати показали, що запропонована модель забезпечує кращу точність розпізнавання на низькоякісних зображеннях, особливо тих, що містять цифровий шум, тоді як Tesseract OCR значно втрачає точність у таких умовах. Отримані результати підтверджують ефективність гібридних архітектур нейронних мереж для розпізнавання спотвореного тексту. У подальшій роботі планується зосередитися на розпізнаванні багаторядкового осмисленого тексту та підвищенні стійкості моделі до різних типів візуальних спотворень.

Ключові слова: розпізнавання тексту, OCR, CNN, RNN, CTC, машинне навчання, спотворені зображення, Tesseract OCR.

ВСТУП

Технологія оптичного розпізнавання символів (OCR) сьогодні є важливим інструментом для автоматизації роботи з текстами та переведення інформації в цифрову форму. Вона дає змогу перетворювати друкований чи рукописний текст у формат, який може оброблятися комп’ютером – зокрема його можна зберігати, шукати або аналізувати. Проте точність роботи OCR суттєво погіршується, якщо зображення мають низьку якість або містять спотворення: розмитість, цифровий шум чи нерівномірне освітлення. Такі ситуації характерні для реальних умов, коли документи фотографують або сканують без спеціального обладнання [1].

Сучасні системи розпізнавання дедалі частіше використовують методи глибинного навчання – передусім згорткові та рекурентні нейронні мережі (CNN і RNN). На практиці вони забезпечують кращі результати, ніж класичні алгоритми OCR. Комбінація CNN і RNN дозволяє моделі водночас виділяти важливі візуальні ознаки та враховувати послідовність символів у тексті. Крім того, функція втрат CTC (Connectionist Temporal

Classification) спрощує вирівнювання передбачених символів з еталонним текстом, оскільки не потребує попередньої сегментації [2].

У цьому дослідженні поставлено завдання створити та протестувати нейронну мережу, здатну розпізнавати як осмислений текст, так і набори випадкових символів на зображеннях низької якості у градаціях сірого. Для цього було сформовано два набори даних: один містив окремі літери та цифри, а інший – фрагменти тексту роману. В обох наборах зображення штучно погіршували за допомогою розмиття та шуму, щоб відтворити типові реальні умови.

Щоб оцінити ефективність розробленої моделі, її результати порівнювали з роботою популярної OCR-системи Tesseract. Основна мета дослідження – показати переваги підходів, що базуються на машинному навчанні, у розпізнаванні спотвореного тексту, а також створити підґрунтя для подальших досліджень, спрямованих на роботу з більш складним і багаторядковим текстом за умов суттєвих візуальних перешкод.

Постановка проблеми. Технології оптичного розпізнавання тексту (OCR) за останні десятиліття пройшли шлях від класичних систем, що працювали за жорстко заданими правилами та шаблонами, до повністю навчальних нейромережевих моделей. Розглянуто найбільш актуальні підходи до розпізнавання друкованого тексту, який може бути спотворений або мати низьку якість.

Традиційні OCR-системи, що послідовно виконують сегментацію символів, їх класифікацію та подальшу мовну обробку, добре працюють на чистих, якісно відсканованих документах. Проте за наявності шуму, розмиття, різноманітних шрифтів або нерівного розташування символів точність таких методів різко падає. Через це дослідження останніх років переорієнтувалися на нейронні моделі, які розпізнають текст «від початку до кінця», без ручної сегментації, та краще адаптуються до реальних умов.

Одним із найпоширеніших рішень є архітектура, що поєднує згорткові нейронні мережі (CNN) для виділення візуальних ознак та рекурентні мережі (RNN – наприклад, LSTM або GRU) для моделювання послідовностей символів. Навчання здійснюється з використанням функції втрат CTC, що дозволяє моделі обходитися без явного поділу тексту на окремі символи. Така CRNN-архітектура довела свою ефективність у задачах розпізнавання тексту і на сценах, і на друкованих зображеннях.

Функція втрат CTC залишається стандартом у випадках, коли навчальні дані не містять точної прив’язки між фрагментами зображення та символами. Вона дає змогу моделі передбачати послідовності змінної довжини та працювати з неповною або ослабленою розміткою – що важливо для наборів даних із шумами та іншими спотвореннями [3].

Останнім часом у галузі спостерігається стрімкий перехід до трансформерних архітектур. Одним із прикладів є TrOCR, де використано візуальний трансформер для обробки зображення та текстовий трансформер для декодування. Такі моделі базуються на попередньо навчених мережах і демонструють високі результати як на друкованих, так і на рукописних текстах. Завдяки ширшому охопленню контексту вони краще справляються зі складними спотвореннями, нерівномірним освітленням та великою різноманітністю шрифтів, ніж системи, побудовані лише на CNN і RNN.

Залучення масштабних попередньо навчених моделей суттєво підвищує точність у випадках, коли навчальні дані обмежені або містять шум.

Використання синтетично згенерованих прикладів, розширених технік спотворення (шум, розмиття, геометричні зміни) та спеціалізованих модулів попередньої обробки підвищує стійкість моделей.

Оптимізовані варіанти CRNN чи трансформерів із багатомасштабними ознаками та меншою кількістю параметрів відкривають можливість застосування на мобільних та вбудованих пристроях.

Створення універсальних моделей для різних мов і систем письма, включно з нелатинськими та рукописними текстами, залишається одним із ключових напрямів розвитку.

Хоча такі витривалі OCR-рушії, як Tesseract, залишаються ефективними для стандартних документів хорошої якості, їхні можливості суттєво обмежуються, коли зображення містять значні спотворення. У таких умовах сучасні нейронні моделі – особливо трансформерні та посилені варіанти CRNN – забезпечують набагато вищу точність. Тому саме підходи, побудовані на глибинному навчанні, нині вважаються найкращими для розпізнавання тексту з шумних або низькоякісних зображень.

Для задачі розпізнавання розмитих або зашумлених зображень тексту найбільш ефективними виявилися моделі, що поєднують потужні механізми виділення ознак (CNN або візуальні трансформери), обробку послідовностей (RNN або трансформер-декодер) та методи навчання без вирівнювання (CTC чи loss функції форматів sequence-to-sequence). Сучасні тенденції – перехід до трансформерів, використання великих попередньо навчених моделей і складніша аугментація – напряму підсилюють стійкість таких систем і становлять природну основу для подальшого розвитку OCR-моделі, описаної в цій роботі.

Аналіз останніх досліджень і публікацій. Робота Chen (2020) [7] показує, що OCR-системи можуть бути ціллю атак. Це підкреслює необхідність розробки захищених моделей, здатних виявляти або ігнорувати адверсаріальні спотворення, що має пряму цінність для кібербезпеки.

Моделі як у Kumar 2025 [1] або PP-OCR [9] демонструють, що сучасні нейронні мережі можуть бути не лише точними, але й оптимізованими за обчислювальними ресурсами – це важливо для вбудованих безпекових пристроїв або систем, що працюють на кінцевих пристроях.

Пост-корекція (Drobac, 2020) [4] – ключовий механізм підвищення надійності OCR, особливо там, де навіть невеликі помилки можуть призвести до серйозних наслідків (наприклад, під час обробки конфіденційних документів).

Підходи Bernasconi (2025) [2], Mask TextSpotter (Liao et al. 2019) [10] та CRNN-сцени (Liu, 2023) [3] показують, як розпізнавати символи або текст на нетрадиційних носіях – це відкриває можливості для безпекових аналізів документів, зображень, фото з камер, PDFсканів із застарілими або підробленими шрифтами.

Дослідження [1] представляє OCR-систему на базі поєднання згорткової та рекурентної нейронних мереж (CRNN), доповнену декодером Word Beam Search (WBS). Вони показують, що за допомогою WBS досягається більш точне розпізнавання слів, особливо в складних або спотворених зображеннях – це важливо в контексті кібербезпеки, оскільки при аналізі документів або зображень, що можуть бути частиною загроз (наприклад, фішингових повідомлень, підроблених документів), саме точність розпізнавання слів, а не окремих символів, може мати вирішальне значення.

У статті [2] запропоновано підхід із використанням **CNN** для класифікації символів у старовинних або рукописних документах (включно з понад 6000 символів). Натренована модель демонструє високу стійкість за рахунок аугментації даних і сегментації символів. З точки зору кібербезпеки це корисно, оскільки символи з нестандартних або рідкісних джерел (наприклад, шифрованих, рукописних або

спеціальних документів) можуть бути коректно розпізнані й інтерпретовані, знижуючи ризик помилкових або пропущених символів у автоматизованому аналізі.

У статті [3] йдеться про модель на базі CRNN, яка інтегрує попередню обробку зображення (алгоритм Retinex), детекцію тексту (DBNet) та класифікатор орієнтації тексту. Це дозволяє моделі успішно працювати з текстом у природних сценах, навіть якщо він багатоорієнтований або перекручений. У контексті кібербезпеки така модель важлива, оскільки багато загроз можуть надходити як зображення наприклад, скриншоти, фотографії з документами або відео з текстом – і розпізнавання цього тексту є критичним для подальшого аналізу.

У роботі [4] автор розглядає, як нейронні мережі можуть бути використані в поєднанні з техніками посткорекції (post-correction) для OCR після первинного розпізнавання система аналізує помилки й виправляє неточності. У кібербезпеці це може бути вкрай корисним, адже неправильне розпізнавання символів або слів може призвести до неправильної інтерпретації важливої інформації (наприклад, у юридичних або фінансових документах) – саме постобробка підвищує надійність системи.

Мета статті: показати переваги підходів, що базуються на машинному навчанні, у розпізнаванні спотвореного тексту, а також створити підґрунтя для подальших досліджень, спрямованих на роботу з більш складним і багаторядковим текстом за умов суттєвих візуальних перешкод.

РЕЗУЛЬТАТИ ДОСЛІДЖЕНЬ

Для виконання поставленого завдання було сформовано два набори зображень низької якості у відтінках сірого. Текст на цих зображеннях був представлений у десяти різних шрифтах:

```
# get list font
font_lst = ['arial', 'arialbd', 'times', 'timesbd', 'timesi','ariblk',
'arialbd', 'arialbi', 'ariali', 'timesbi']
```

Перший набір містить 4 960 зображень розміром 250×50 пікселів. На них випадковим чином розміщені великі та малі букви англійського алфавіту, а також цифри, виконані шрифтом розміром 24 px. Другий набір включає 4 010 зображень з роздільністю 680×50 пікселів, на яких подано осмислені фрагменти тексту. Розмір шрифту також становить 24 px. Спочатку для навчання та тестування OCR-системи використовували перший набір, до якого додавали розмиття та цифровий шум. Після цього аналогічну обробку застосували і до другого набору [4].

Рис. 1 – Приклади зображень із першого набору даних після внесення спотворень

ISSN 2663 - 4023

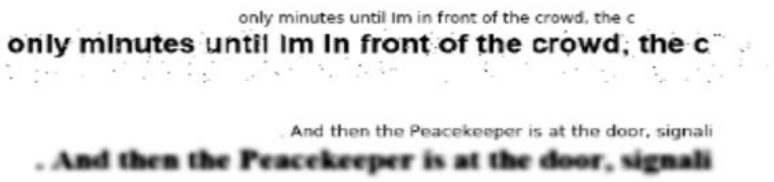

Рис. 2 – Приклади зображень із другого набору даних після внесених спотворень

Модель для розпізнавання тексту була реалізована на основі матеріалів статті «OCR Model for Reading Captchas» [5]. У ній наведено приклад простої OCR-моделі, побудованої за допомогою API-функцій, яка поєднує CNN і RNN. Також показано, як створити власний шар і використати його як «кінцевий шар» для реалізації функції втрат CTC. API-функції, що застосовувалися під час побудови моделі, подані на рисунку 3 у стовпці «Layer (type)».

Model: "ocr_model_v1"

Layer (type)	Output Shape	Param #	Connected to
image (InputLayer)	[(None, 200, 50, 1)]	0	[]
Conv1 (Conv2D)	(None, 200, 50, 32)	320	['image[0][0]']
pool1 (MaxPooling2D)	(None, 100, 25, 32)	0	['Conv1[0][0]']
Conv2 (Conv2D)	(None, 100, 25, 64)	18496	['pool1[0][0]']
pool2 (MaxPooling2D)	(None, 50, 12, 64)	0	['Conv2[0][0]']
reshape (Reshape)	(None, 50, 768)	0	['pool2[0][0]']
dense1 (Dense)	(None, 50, 64)	49216	['reshape[0][0]']
dropout_2 (Dropout)	(None, 50, 64)	0	['dense1[0][0]']
bidirectional_4 (Bidirectional)	(None, 50, 256)	197632	['dropout_2[0][0]']
bidirectional_5 (Bidirectional)	(None, 50, 128)	164352	['bidirectional_4[0][0]']
label (InputLayer)	[(None, None)]	0	[]
dense2 (Dense)	(None, 50, 64)	8256	['bidirectional_5[0][0]']
ctc_loss (CTCLayer)	(None, 50, 64)	0	['label[0][0]', 'dense2[0][0]']

Total params: 438,272
Trainable params: 438,272
Non-trainable params: 0

Рис. 3 – Модель, побудована для першого набору зображень

Model: "ocr_model_v1"

Layer (type)	Output Shape	Param #	Connected to
image (InputLayer)	[(None, 680, 50, 1)]	0	[]
Conv1 (Conv2D)	(None, 680, 50, 32)	320	['image[0][0]']
pool1 (MaxPooling2D)	(None, 340, 25, 32)	0	['Conv1[0][0]']
Conv2 (Conv2D)	(None, 340, 25, 64)	18496	['pool1[0][0]']
pool2 (MaxPooling2D)	(None, 170, 12, 64)	0	['Conv2[0][0]']
reshape (Reshape)	(None, 170, 768)	0	['pool2[0][0]']
dense1 (Dense)	(None, 170, 64)	49216	['reshape[0][0]']
dropout (Dropout)	(None, 170, 64)	0	['dense1[0][0]']
bidirectional (Bidirectional)	(None, 170, 256)	197632	['dropout[0][0]']
bidirectional_1 (Bidirectional)	(None, 170, 128)	164352	['bidirectional[0][0]']
label (InputLayer)	[(None, None)]	0	[]
dense2 (Dense)	(None, 170, 73)	9417	['bidirectional_1[0][0]']
ctc_loss (CTCLayer)	(None, 170, 73)	0	['label[0][0]', 'dense2[0][0]']

Total params: 439,433
Trainable params: 439,433
Non-trainable params: 0

Рис. 4 – Модель, створена для другого набору зображень

Набори зображень, підготовлені для навчання та валідації моделі, складаються з низькоякісних прикладів і були поділені у співвідношенні 9:1 відповідно [6; 7]. Навчання

моделі на першому наборі тривало 70 епох, при цьому одна епоха займала приблизно 124 секунди. Значення функції втрат під час навчання та валідації наведено на рисунку 5.

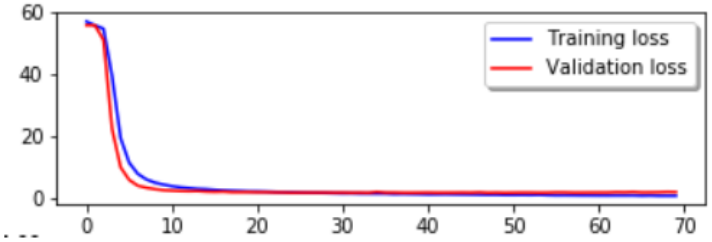

Рис. 5 – Значення функції втрат

Навчання моделі на другому наборі зображень проводилося протягом 80 епох, при цьому одна епоха тривала в середньому 246 секунд. Значення функції втрат під час навчання та валідації наведено на рисунку 6.

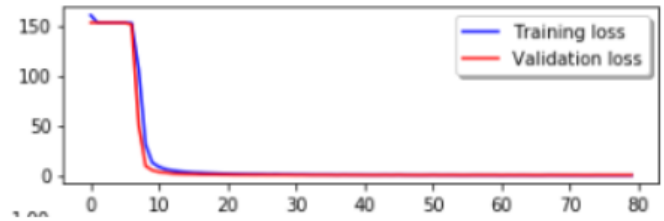

Рис. 6 – Значення функції втрат

Результати роботи моделі на першому наборі зображень наведено на рисунках 7, 8 та 9 (результати розпізнавання).

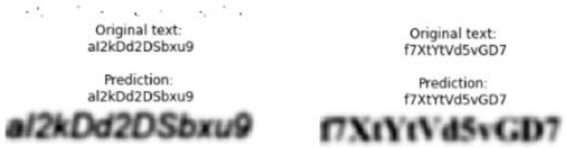

Рис. 7 – Результати розпізнавання

Результати роботи моделі на другому наборі зображень наведено на рисунках 8–11.

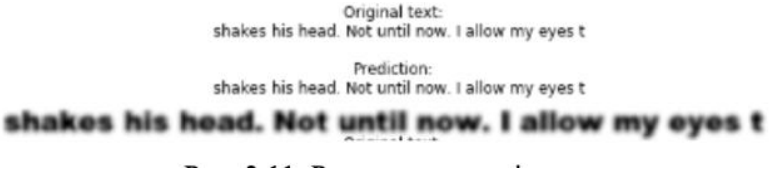

Рис. 8 – Результати розпізнавання

ISSN 2663 - 4023

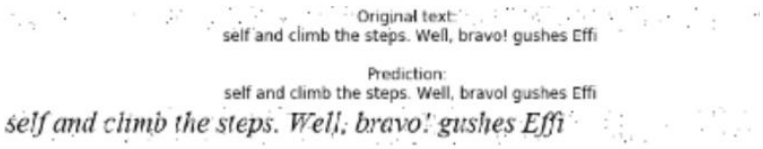

Рис. 9 – Результати розпізнавання

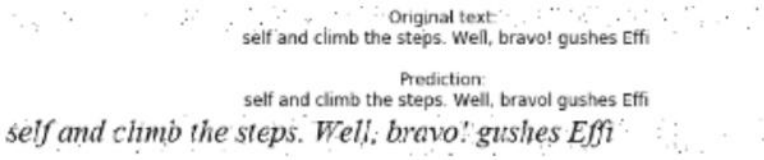

Рис. 10 – Результати розпізнавання

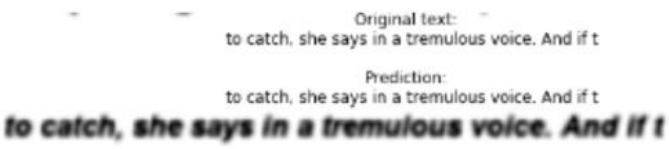

Рис. 11 – Результати розпізнавання

Для порівняння було використано Tesseract OCR. Результати розпізнавання тексту наведено на рисунках 12–15.

Рис. 12 – Результати розпізнавання за допомогою Tesseract OCR та розробленої моделі

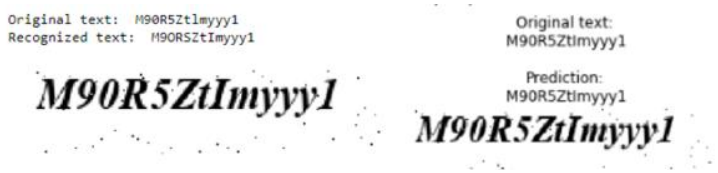

Рис. 13 – Результати розпізнавання за допомогою Tesseract OCR та розробленої моделі

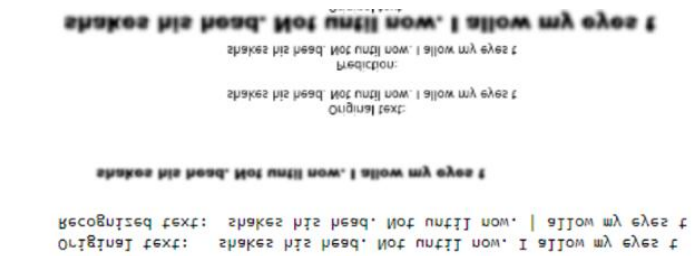

Рис. 14 – Результати розпізнавання за допомогою Tesseract OCR та розробленої моделі

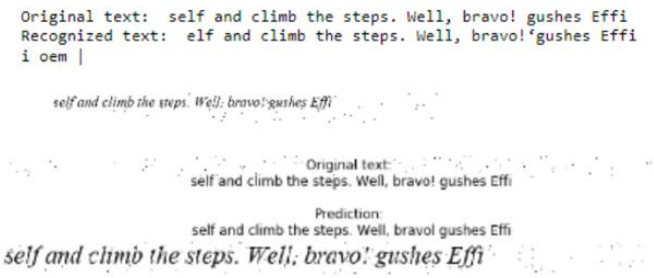

Рис. 15 – Результати розпізнавання за допомогою Tesseract OCR та розробленої моделі

З огляду на наведені результати можна зробити висновок, що система Tesseract OCR показує низьку точність при роботі з низькоякісними зображеннями. З власних спостережень у розпізнаванні тексту на спотворених зображеннях, моделі краще справляються з текстом, пошкодженим цифровим шумом, ніж із розмитими зображеннями.

Проведені експерименти показують, що запропонована OCR-модель, яка поєднує згорткові та рекурентні нейронні мережі з функцією втрат CTC, ефективно розпізнає текст на низькоякісних зображеннях у відтінках сірого. Використання синтетичних наборів даних зі спотвореннями, такими як розмиття та цифровий шум, дозволило всебічно оцінити стійкість моделі у реалістичних умовах [8; 9].

Результати свідчать, що модель демонструє значно кращу точність порівняно з класичною системою Tesseract OCR при обробці спотворених зображень. Хоча Tesseract зберігає прийнятну точність на чистих або добре відсканованих текстах, він не справляється зі спотвореннями. На відміну від нього, запропонована модель підтримує стабільну точність розпізнавання навіть при розмитих зображеннях або наявності різних видів шуму [10].

Важливою спостереженою особливістю є те, що модель точніше розпізнає текст на зображеннях із цифровим шумом порівняно з розмитими. Це пояснюється тим, що шум не порушує просторову структуру символів так сильно, як розмиття. Розмиття призводить до злиття країв символів, що зменшує чіткі візуальні патерни, на яких базуються шари CNN для виділення ознак. Це свідчить про те, що додаткові методи попередньої обробки – наприклад, фільтри для видалення розмиття, методи суперроздільної здатності або адаптивне підвищення контрасту – можуть ще більше покращити точність розпізнавання розмитого тексту.

Процес навчання показав, що запропонована архітектура нейронної мережі здатна ефективно сходитися протягом розумної кількості епох, досягаючи низьких значень функції втрат як на тренувальному, так і на валідаційному наборах. Стабільність кривих сходження свідчить про те, що модель добре узагальнює дані і не проявляє перенавчання,

незважаючи на відносно невеликі набори даних. Це підкреслює відповідність обраної архітектури і ефективність обраної стратегії навчання, включно з аугментацією даних та оптимізацією на основі CTC.

Крім того, дослідження підтверджує практичний потенціал створення легких OCR-систем «від початку до кінця» на базі глибинного навчання. Навчаючи моделі на завданнях із наборами даних, що відображають реальні спотворення, можна створювати системи розпізнавання, адаптовані до конкретних умов – наприклад, низькоякісні скани, відеоспостереження або документи, зняті смартфоном [11; 12].

Загалом, обговорення результатів експериментів підтверджує, що запропонована архітектура успішно заповнює розрив між традиційними алгоритмами OCR та сучасними рішеннями на основі глибинного навчання. Вона демонструє, що системи на базі CNN–RNN–CTC здатні забезпечувати надійні результати навіть у складних умовах, що відкриває можливості для подальших досліджень, спрямованих на підвищення точності розпізнавання складного, багаторядкового та багатомовного тексту.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

У цій роботі була розроблена та оцінена OCR-система на основі нейронних мереж для розпізнавання як беззмістовного, так і осмисленого тексту на низькоякісних зображеннях у відтінках сірого. Для моделювання реальних умов було створено два спеціальні набори даних – один містив окремі буквено-цифрові символи, а інший – суцільні фрагменти тексту з книги. Кожен набір навмисно погіршили за допомогою розмиття та цифрового шуму, щоб перевірити стійкість моделі.

Запропонована модель, побудована на поєднанні згорткових (CNN) та рекурентних нейронних мереж (RNN) із функцією втрат Connectionist Temporal Classification (CTC), показала високу ефективність у розпізнаванні спотвореного тексту. Експериментальні результати підтвердили, що модель перевершує класичну систему Tesseract OCR, особливо на зображеннях із різними видами спотворень. Аналіз також показав, що текст на зображеннях із цифровим шумом зазвичай розпізнається легше, ніж на розмитих, оскільки розмиття приховує важливі межі символів та їхні структурні особливості.

Отримані результати підтверджують ефективність методів глибинного навчання для задач OCR, де якість зображень знижена. Гібридний підхід CNN–RNN–CTC забезпечує гнучкість і стійкість до шуму, що робить його придатним для застосувань, таких як цифровізація документів, автоматичне вилучення даних та розпізнавання тексту в режимі реального часу на низькоякісних джерелах, наприклад, із камер спостереження чи мобільних пристроїв.

Перспективи подальшої роботи включають кілька ключових напрямів:

- Розширення на багаторядковий і багатомовний текст – покращення моделі для обробки довгих текстових блоків, абзаців та різних алфавітів.
- Інтеграція трансформерних архітектур – дослідження TrOCR або гібридних структур CNN–Transformer для кращого врахування глобального контексту та підвищення точності розпізнавання.
- Впровадження методів покращення зображень – включно з фільтрацією розмиття, зменшенням шуму та модулями суперроздільної здатності для підвищення чіткості спотворених зображень.
- Оптимізація для роботи в реальному часі – адаптація моделі для вбудованих та крайових пристроїв із обмеженими обчислювальними ресурсами.

• Створення більших і різноманітніших наборів даних – розширення тренувальних даних для включення рукописного тексту, різних умов освітлення та шрифтів для забезпечення кращої узагальнюваності моделі.

Загалом, проведене дослідження демонструє, що архітектури глибинного навчання є потужною і гнучкою основою для OCR-систем, здатних ефективно працювати за неідеальних умов. Подальший розвиток дизайну моделей, аугментації даних та методів попередньої обробки очікувано сприятиме підвищенню точності розпізнавання та розширенню можливостей застосування OCR у сучасних цифрових системах.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Kumar, M., Singh, S., Dureja, A., Narula, R., & Shyla, S. (2025). OCR-CRNN (WBS): An optical character recognition system based on convolutional recurrent neural network embedded with word beam search decoder for extraction of text. *International Journal of Information Technology, 17*(7), 849–860. https://doi.org/10.1007/s41870-025-02540-x
2. Bernasconi, E. (2025). Enhancing symbol recognition in library science via convolutional neural networks. *Journal of Computer Science, 16*(2), 119–130. https://doi.org/10.3390/jcs16020119
3. Liu, Y. (2023). A convolutional recurrent neural-network-based model for scene text recognition. *Symmetry, 15*(4), 849–860. https://doi.org/10.3390/sym15040849
4. Drobac, S. (2020). Optical character recognition with neural networks and post-correction techniques. *Pattern Recognition, 100*, 107–118. https://doi.org/10.1007/s10032-020-00359-9
5. Sinthuja, M. (2024). Extraction of text from images using deep learning. *Procedia Computer Science, 187*, 751–758. https://doi.org/10.1016/j.procs.2024.04.099
6. Yousef, M., Hussain, K. F., & Mohammed, U. S. (2018). Accurate, data-efficient, unconstrained text recognition with convolutional neural networks. *arXiv preprint* arXiv:1812.11894. https://arxiv.org/abs/1812.11894
7. Chen, L. (2020). Attacking optical character recognition (OCR) systems with adversarial examples. *arXiv preprint* arXiv:2002.03095. https://arxiv.org/abs/2002.03095
8. Li, B., Tang, X., Qi, X., Chen, Y., & Xiao, R. (2020). Hamming OCR: A locality sensitive hashing neural network for scene text recognition. *arXiv preprint* arXiv:2009.10874. https://arxiv.org/abs/2009.10874
9. Du, Y., et al. (2020). PP-OCR: A practical ultra lightweight OCR system. *arXiv preprint* arXiv:2009.09941. https://arxiv.org/abs/2009.09941
10. Liao, M., et al. (2019). Mask TextSpotter: An end-to-end trainable neural network for spotting text with arbitrary shapes. *arXiv preprint* arXiv:1908.08207. https://arxiv.org/abs/1908.08207
11. Shi, B., Bai, X., & Yao, C. (2017). An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(11), 2298–2304. https://doi.org/10.1109/TPAMI.2016.2645393
12. Liu, Y. (2024, October 14). Convolutional recurrent neural network for text recognition. *XenonStack Insights*. https://www.xenonstack.com/insights/crnn-for-text-recognition

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 3 (31), 2025
ISSN 2663 - 4023

Nataliia Cherniiashchuk
Doctor of Pedagogical Sciences, Professor
Lesya Ukrainka Volyn National University, Lutsk, Ukraine
ORCID: 0000-0002-3178-8377
cherniashchuk.nataliia@vnu.edu.ua

APPLICATION OF NEURAL NETWORKS FOR AUTOMATED TEXT AND SYMBOL RECOGNITION IN CYBERSECURITY TASKS

Abstract. This work presents the development and investigation of an Optical Character Recognition (OCR) system for low-quality text using machine learning methods. To address the task, two grayscale image datasets were created: the first consisting of isolated English alphabet characters and digits (4,960 images of 250×50 pixels), and the second containing fragments of meaningful text from the book *The Hunger Games* (4,010 images of 680×50 pixels). To enhance model robustness, the images were distorted using blur and digital noise functions. The OCR model was built based on a combination of Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) with a Connectionist Temporal Classification (CTC) layer for sequence correction. Training was conducted over 70–80 epochs with a 9:1 split for training and validation datasets. A comparative analysis was carried out between the developed system and Tesseract OCR. Experimental results demonstrated that the proposed model achieves superior recognition performance on low-quality images, particularly those affected by digital noise, whereas Tesseract OCR significantly loses accuracy under such conditions. The results confirm the effectiveness of hybrid neural network architectures for recognizing distorted text. Future work will focus on recognizing multi-line meaningful text and improving the model's resilience to various types of visual distortions.

Keywords – text recognition, OCR, CNN, RNN, CTC, machine learning, distorted images, Tesseract OCR.

REFERENCES

1. Kumar, M., Singh, S., Dureja, A., Narula, R., & Shyla, S. (2025). OCR-CRNN (WBS): An optical character recognition system based on convolutional recurrent neural network embedded with word beam search decoder for extraction of text. *International Journal of Information Technology*, *17*(7), 849–860. https://doi.org/10.1007/s41870-025-02540-x
2. Bernasconi, E. (2025). Enhancing symbol recognition in library science via convolutional neural networks. *Journal of Computer Science*, *16*(2), 119–130. https://doi.org/10.3390/jcs16020119
3. Liu, Y. (2023). A convolutional recurrent neural-network-based model for scene text recognition. *Symmetry*, *15*(4), 849–860. https://doi.org/10.3390/sym15040849
4. Drobac, S. (2020). Optical character recognition with neural networks and post-correction techniques. *Pattern Recognition*, *100*, 107–118. https://doi.org/10.1007/s10032-020-00359-9
5. Sinthuja, M. (2024). Extraction of text from images using deep learning. *Procedia Computer Science*, *187*, 751–758. https://doi.org/10.1016/j.procs.2024.04.099
6. Yousef, M., Hussain, K. F., & Mohammed, U. S. (2018). Accurate, data-efficient, unconstrained text recognition with convolutional neural networks. *arXiv preprint* arXiv:1812.11894. https://arxiv.org/abs/1812.11894
7. Chen, L. (2020). Attacking optical character recognition (OCR) systems with adversarial examples. *arXiv preprint* arXiv:2002.03095. https://arxiv.org/abs/2002.03095
8. Li, B., Tang, X., Qi, X., Chen, Y., & Xiao, R. (2020). Hamming OCR: A locality sensitive hashing neural network for scene text recognition. *arXiv preprint* arXiv:2009.10874. https://arxiv.org/abs/2009.10874
9. Du, Y., et al. (2020). PP-OCR: A practical ultra lightweight OCR system. *arXiv preprint* arXiv:2009.09941. https://arxiv.org/abs/2009.09941
10. Liao, M., et al. (2019). Mask TextSpotter: An end-to-end trainable neural network for spotting text with arbitrary shapes. *arXiv preprint* arXiv:1908.08207. https://arxiv.org/abs/1908.08207

11. Shi, B., Bai, X., & Yao, C. (2017). An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(11), 2298–2304. https://doi.org/10.1109/TPAMI.2016.2645393
12. Liu, Y. (2024, October 14). Convolutional recurrent neural network for text recognition. *XenonStack Insights*. https://www.xenonstack.com/insights/crnn-for-text-recognition