

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Сумський державний університет

Центр заочної, дистанційної та вечірньої форм навчання

Кафедра комп'ютерних наук

«До захисту допущено»

В.о. завідувача кафедри

_____ Оксана ШОВКОПЛЯС

27 травня 2024 р.

КВАЛІФІКАЦІЙНА РОБОТА

на здобуття освітнього ступеня бакалавр

зі спеціальності 122 – Комп'ютерні науки,

освітньо-професійної програми «Інформатика»

на тему: «Інформаційна система розпізнавання тексту з використанням штучних нейронних мереж»

здобувачки групи ІНпн-01с Кондратенко Катерини Віталіївни

Кваліфікаційна робота містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело.

_____ Катерина КОНДРАТЕНКО

Керівник

старший викладач, к. ф.-м.н.

кафедри комп'ютерних наук Анна БАДАЛЯН_____

Суми – 2024

Сумський державний університет
Факультет електроніки та інформаційних технологій
Кафедра комп'ютерних наук

«Затверджую»

В.о. завідувача кафедри

Оксана ШОВКОПЛЯС

(підпис)

ІНДИВІДУАЛЬНЕ ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ
на здобуття освітнього ступеня бакалавра

зі спеціальності 122 - Комп'ютерних наук, освітньо-професійної програми «Інформатика»
здобувачки групи ІНпн-01с Кондратенко Катерини Віталіївни

1. Тема роботи: «Інформаційна система розпізнавання тексту з використанням штучних нейронних мереж» затверджую наказом по СумДУ від «26» квітня 2024 р. № 0414-VI
2. Термін здачі здобувачем кваліфікаційної роботи до _____
3. Вхідні дані до кваліфікаційної роботи _____
4. Зміст розрахунково-пояснювальної записки (перелік питань, що їх належить розробити)
1) Аналіз проблеми предметної області, постановка й формування завдань дослідження.
2) Огляд технологій, що використовуються для розпізнавання тексту. 3) Аналіз інтелектуальної системи з розпізнавання тексту. 4) Аналіз результатів.
5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень) _____
6. Консультанти до проекту (роботи), із значенням розділів проекту, що стосується їх

Розділ	Консультант	Підпис, дата	
		Завдання видав	Завдання прийняв

7. Дата видачі завдання «06» травня 2024 р.

Завдання прийняв до виконання _____
(підпис)

Керівник _____
(підпис)

КАЛЕНДАРНИЙ ПЛАН

№ п/п	Назва етапів кваліфікаційної роботи	Термін виконання	Примітка
1	<i>Аналіз предметної області та аналогів, постановка й формування завдань.</i>	06.05.24-08.05.24	
2	<i>Вибір технологій для виконання поставлених завдань</i>	09.05.24-11.05.24	
3	<i>Аналіз існуючих систем розпізнавання тексту</i>	12.05.24-17.05.24	
4	<i>Виклад результатів дослідження та запис результатів дослідження</i>	18.05.24- 31.05.24	

Здобувач вищої освіти _____
(підпис)

Керівник _____
(підпис)

АНОТАЦІЯ

Записка: 30 стр., 10 рис., 14 використаних джерел.

Обґрунтування актуальності теми роботи – Тема кваліфікаційної роботи є актуальною, оскільки розпізнавання тексту є важливою задачею в багатьох галузях, таких як обробка природної мови, машинне перекладання, комп'ютерний зір і робототехніка. Із розвитком штучних нейронних мереж (ШНМ) з'явилися нові можливості для значного покращення точності та ефективності розпізнавання тексту.

Об'єкт дослідження — інформаційні системи розпізнавання тексту з використанням штучних нейронних мереж.

Мета роботи — дослідження теоретичних основ розпізнавання тексту з використанням штучних нейронних мереж та розробка інформаційної системи.

Методи дослідження — аналіз та огляд літератури, вивчення наявних досліджень, статей та книг з теми розпізнавання тексту з використанням штучних нейронних мереж для виявлення поточних тенденцій, проблем та досягнень.

Результати — проведений аналіз та порівняні системи розпізнавання тексту (*OCR*) *Google Cloud Vision*, *Microsoft Azure OCR* і *Tesseract*. Комерційні системи, такі як *Google Cloud Vision* та *Microsoft Azure OCR*, забезпечують високу точність, швидкість та інтеграцію, але вимагають оплати за використання. Відкрита система *Tesseract* пропонує високу гнучкість та економію коштів, але може вимагати більше зусиль для досягнення бажаної точності та швидкості.

ІНФОРМАЦІЙНІ СИСТЕМИ РОЗПІЗНАВАННЯ ТЕКСТУ; ШТУЧНІ НЕЙРОННІ МЕРЕЖІ; ЗГОРТКОВІ НЕЙРОННІ МЕРЕЖІ (CNN); ОПТИЧНЕ РОЗПІЗНАВАННЯ СИМВОЛІВ (OCR); ПОРІВНЯЛЬНИЙ АНАЛІЗ; МАШИННЕ НАВЧАННЯ.

ЗМІСТ

Вступ	3
1. Інформаційний огляд	5
1.1 Поняття та види штучних нейронних мереж	5
1.2 Застосування штучних нейронних мереж для розпізнавання тексту	8
2. Вибір методів розв’язання задачі та аналіз	12
2.1 Вибір архітектури та алгоритму навчання	12
2.2 Обґрунтування вибору архітектури	15
2.3 Огляд та аналіз існуючих систем розпізнавання тексту: <i>Google Cloud Vision, Microsoft Azure OCR, Tesseract</i>	19
Висновки	28
Список використаних джерел	30

ВСТУП

Обґрунтування вибору теми роботи. У сучасному світі обробка та аналіз текстової інформації займає ключове місце у багатьох сферах діяльності. Ми можемо використовувати це від бізнесу та наукових досліджень до повсякденного життя. Зі зростанням обсягів інформації, яка генерується щодня, виникає потреба у автоматизації процесів обробки тексту. Це стимулює розвиток та впровадження новітніх технологій, зокрема штучного інтелекту та машинного навчання, для розв'язання задач розпізнавання тексту. Розпізнавання тексту – це завдання автоматичного перетворення зображень тексту в текстовий формат, який можна обробляти комп'ютером. Це складне завдання, яке потребує розуміння контексту, граматики та особливостей мови. ШНМ, завдяки своїй здатності навчатися на великих обсягах даних, дозволяють ефективно вирішувати цю задачу.

Однією з найбільш перспективних й ефективних технологій для розпізнавання тексту є штучні нейронні мережі (ШНМ), особливо *Convolutional Neural Networks (CNN)*. *CNN* широко застосовуються у різних задачах комп'ютерного зору та показують високі результати у розпізнаванні тексту на зображеннях. Ця технологія дозволяє автоматизувати процеси вилучення текстової інформації з документів, зображень та інших джерел, забезпечуючи високу точність та швидкість обробки.

Актуальність. Визначається тим, що розпізнавання тексту є важливою задачею в багатьох галузях, таких як обробка природної мови, машинне перекладення, комп'ютерний зір та робототехніка. З розвитком штучних нейронних мереж з'явилися нові можливості для значного покращення точності та ефективності розпізнавання тексту. ШНМ можуть навчатися на великих наборах даних тексту та зображень, що дозволяє їм виявляти складні закономірності та розпізнавати текст із високою точністю, навіть у складних умовах.

Метою даної кваліфікаційної роботи є дослідження теоретичних основ розпізнавання тексту з використанням штучних нейронних мереж і розробка інформаційної системи.

Завданнями роботи є:

1. Огляд та аналіз теоретичних основ розпізнавання тексту за допомогою штучних нейронних мереж.
2. Вибір та обґрунтування архітектури нейронної мережі для задачі розпізнавання тексту.
3. Аналіз практичних застосувань та оцінка ефективності існуючих систем.

Об'єкт дослідження – інформаційні системи розпізнавання тексту з використанням штучних нейронних мереж.

Предмет дослідження – методи та алгоритми розпізнавання тексту з використанням штучних нейронних мереж.

Новизна полягає у розкритті нових аспектів використання штучні нейронні мережі у розпізнаванні текстів.

Структура. Робота складається зі вступу, двох розділів, висновків, списку використаних джерел. Загальна кількість сторінок – 30.

1. Інформаційний огляд

1.1 Поняття та види штучних нейронних мереж

Нейронні мережі - це масивні паралельні розподілені процесори, що складаються з простих обчислювальних блоків, які мають природну тенденцію зберігати і робити доступними досвідчені знання.

Вони схожі на мозок у двох аспектах.

1) У процесі навчання мережа отримує знання з навколишнього середовища.

2) Сила зв'язків між нейронами, відома як синаптичні ваги, використовується для зберігання набутих знань. [12].

Штучні нейронні мережі – це комп'ютерні системи, засновані на біологічних нейронних мережах, які складають мозок тварин. ШНМ складається з великої кількості взаємопов'язаних штучних нейронів, які називаються вузлами. Кожен вузол обробляє вхідні сигнали і генерує вихідні сигнали. Зв'язки між вузлами мають ваги, які визначають вплив одного вузла на інший. [8]. ШНМ можуть навчатися виконувати різні завдання, аналізуючи дані та роблячи прогнози. Навчання відбувається за допомогою алгоритмів, які регулюють ваги зв'язків між вузлами. Існує багато різних алгоритмів навчання, кожен з яких має свої переваги та недоліки.

ШНМ працює наступним чином. Нейрон отримує та узагальнює сигнали. При цьому враховуються синаптичні ваги, тобто важливість окремих входів. Вхідні сигнали від одних нейронів передаються на входи інших нейронів. Вага кожного такого з'єднання може бути позитивною (збуджувальні зв'язки) або негативною (гальмівні зв'язки). Вони визначають обчислення нейронної мережі, а отже, її пам'ять і поведінку, рис. 1 [8].

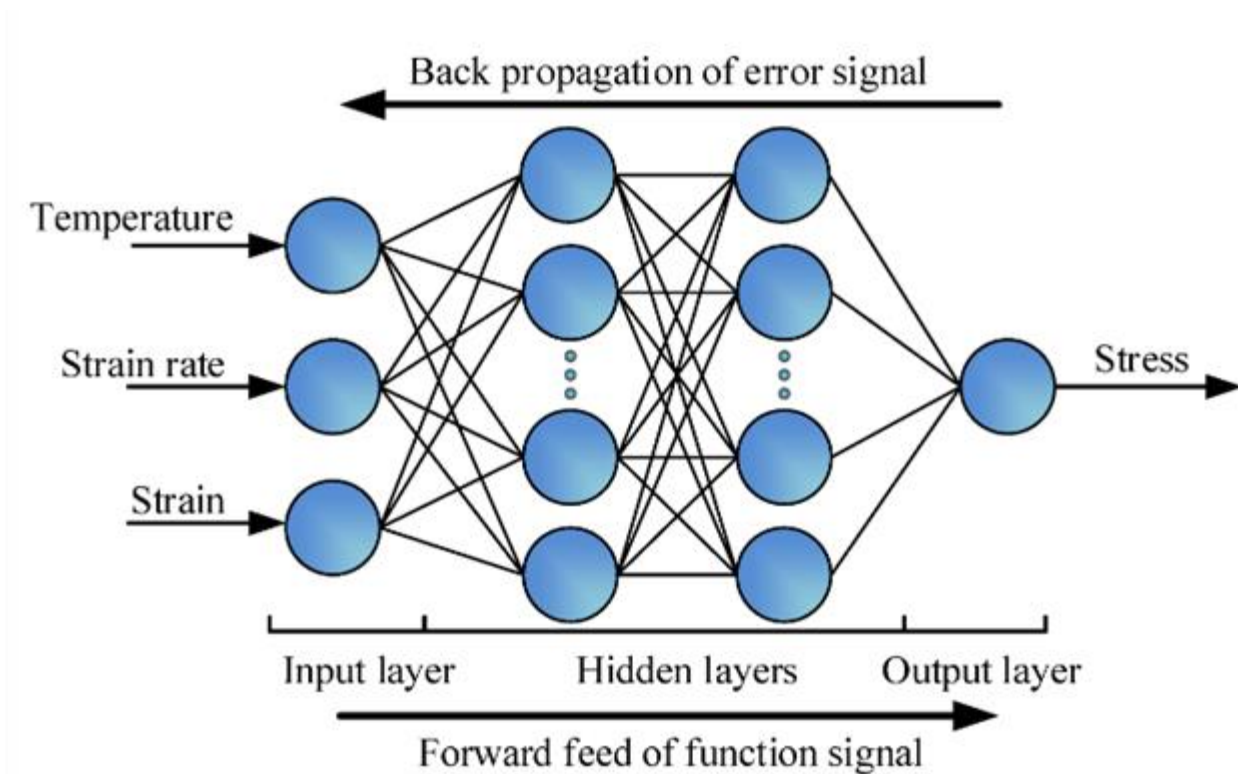


Рис.1

Для виконання складних операцій ШНМ потребують навчання.

Штучні нейронні мережі не програмується в традиційному розумінні; здатність до навчання є однією з головних переваг нейронних мереж над традиційними алгоритмами. Технічно кажучи, навчання полягає в знаходженні коефіцієнтів зв'язку між нейронами. У процесі навчання нейронні мережі здатні розпізнавати та узагальнювати складні взаємозв'язки між вхідними та вихідними даними.

Для навчання потрібна модель зовнішнього середовища, в якому працює нейронна мережа, інформація, необхідна їй для вирішення завдання. По-друге, потрібно вирішити, як змінювати вагові параметри мережі [8]. Здатність мережі навчатися на прикладах робить їх більш привабливими. Алгоритм навчання - це процедура, яка використовує правила навчання для налаштування вагових коефіцієнтів. Існує три парадигми навчання:

- з «вчителем»;
- без «вчителя» (самонавчання);
- змішана (комбінована).

За завданням:

- Нейронні мережі для класифікації: використовуються для класифікації даних за попередньо визначеними категоріями.
- Нейронні мережі для прогнозування: використовуються для прогнозування вихідних значень на основі вхідних даних.
- Нейронні мережі для обробки зображень: використовуються для обробки та аналізу зображень.
- Нейронні мережі для обробки природної мови: використовуються для обробки та аналізу природної мови [13].

Це лише деякі з найпоширеніших типів штучних нейронних мереж. Існує багато інших типів ШНМ, які мають свої унікальні характеристики і використовуються для вирішення різних завдань.

Таким чином, штучні нейронні мережі є потужним інструментом, який можна використовувати для вирішення складних завдань. Штучні нейронні мережі імітують функціонування біологічних нейронних мереж, що дозволяє створювати системи, які можуть навчатися на прикладах. Основними компонентами ШНМ є нейрони, зв'язки між ними (ваги) та функції активації, які перетворюють вхідні сигнали на вихідні. Однак їх розробка та використання може бути досить складним завданням.

1.2 Застосування штучних нейронних мереж для розпізнавання тексту

Розпізнавання тексту - це завдання автоматичного перетворення текстових зображень у текстовий формат, який може бути оброблений комп'ютером. Це складне завдання, яке вимагає розуміння контексту, граматики та особливостей мови. Штучні нейронні мережі (ШНМ), завдяки своїй здатності навчатися на великих обсягах даних, можуть ефективно вирішити це завдання.

Штучні нейронні мережі широко використовуються в задачах розпізнавання тексту завдяки своїй здатності навчатися на великих масивах даних і виявляти складні закономірності. Прикладом типового завдання є

розпізнавання друкованого тексту, а прикладом спеціалізованого - розпізнавання спеціальної символічної інформації. Існує багато програм, призначених для розпізнавання тексту. Наприклад, *FineReader*, *ScanSoft OmniPage* та інші.

Оптичне розпізнавання символів (*OCR*) - це машинне або електронне перетворення зображення рукописного, машинописного або друкованого тексту в набір кодів, які використовуються для представлення його в текстовій формі [5]. *OCR* часто використовується для автоматичного перетворення документів і книг в електронний формат або для публікації тексту на веб-сторінках. Оптичне розпізнавання дозволяє редагувати, форматовувати, аналізувати, перекладати в електронному вигляді, шукати і зберігати текст у різних форматах.

ШКМ використовуються для перетворення текстових зображень у цифровий текст. Це може бути корисно для таких завдань, як сканування документів, розпізнавання номерних знаків і автоматичне вилучення тексту із зображень.

Основний виклик *OCR* полягає у вирішенні проблеми, що існує багато способів написання одного і того ж символу різними шрифтами. Крім того, перед вибором алгоритму розпізнавання зображення необхідно попередньо обробити, щоб зробити його готовим до «читання». За правило, процес відбувається приблизно так:

1. Сканування або фотографування документів.
2. Використовується технологія оптичного розпізнавання символів.
3. Дані зберігаються у вигляді тексту з метаданими для пошуку.
4. Одержані дані використовуються для конкретних цілей.

Наразі ця технологія використовується у великій кількості програмних рішень, пов'язаних з розпізнаванням тексту. Основним завданням таких *OCR*-систем є присвоєння фрагменту текстового зображення відповідної символічної інформації. Останні дослідження в технології *OCR* базуються на застосуванні принципів роботи зорової системи людини, таких як [13]:

- принцип цілісності, згідно з яким об'єкт представляється як сукупність частин і просторових зв'язків між ними;

- принцип цілеспрямованості: будь-яка інтерпретація даних переслідує певну мету. Згідно з цим принципом, розпізнавання є процесом висунення гіпотез про цілий об'єкт і цілеспрямованої їх перевірки;

- принцип адаптивності передбачає, що система здатна до самонавчання, зберігаючи інформацію під час її обробки. Переваги системи розпізнавання, що працює відповідно до принципів СППР, очевидні: саме вони здатні забезпечити найбільш гнучку і осмислену поведінку системи. Наприклад, *FineReader* компанії *ABBYY's*, одна з найпоширеніших систем розпізнавання, працює за цими принципами на всіх етапах обробки документів [3].

Особливості використання ШМ для розпізнавання тексту:

- Гнучкість: можуть бути пристосовані до різноманітних завдань розпізнавання тексту, таких як розпізнавання тексту, машинний переклад і розпізнавання мовлення.

- Масштабованість: масштабуються для обробки великих обсягів даних.

Основні недоліки використання для розпізнавання тексту:

- Складність: Розроблення та навчання може бути складним завданням, яке вимагає знань про машинне навчання та обробку природної мови.

- Вартість: Підготовка може вимагати значних обчислювальних ресурсів і часу.

- Недостатньо вивчені: здатні бути «чорним ящиком», що ускладнює розуміння того, як вони приймають рішення.

Найпоширеніші застосування штучних нейронних мереж [8]:

У процесі розпізнавання та класифікації образів відбувається ідентифікація різних об'єктів, наприклад, текстових символів, зображень або звукових зразків, і віднесення їх до певних категорій. Під час навчання мережі надаються різні приклади зображень із зазначенням їхньої

класифікації. При пред'явленні мережі певного зображення один з її виходів повинен вказувати на те, що це зображення належить до відповідного класу, а інші виходи - на те, що зображення не відповідає цьому класу.

Ухвалення рішень та управління. Дана задача схожа на задачу класифікації. Класифікації підлягають ситуації, характеристики яких подаються на вхід нейронної мережі. Результат роботи мережі повинен показувати індикатор прийнятого нею рішення.

Групування (кластеризація). Кластеризація – це поділ множини вхідних сигналів на класи. Однак ні кількість, ні характеристики класів не відомі заздалегідь.

Передбачення. Спроможність нейронної мережі до прогнозування безпосередньо впливає з її здатності до узагальнення та виявлення прихованих залежностей між вхідними та вихідними даними. Навчившись, мережа здатна передбачити майбутнє значення послідовності на основі декількох попередніх значень або деяких чинників, що існують на даний момент.

Стиснення даних та асоціативна пам'ять. Можливість нейронних мереж виявляти взаємозв'язки між різними параметрами дає змогу більш компактно виражати багатовимірні дані, якщо вони тісно пов'язані між собою.

Загалом, штучні нейронні мережі є потужним інструментом для розпізнавання тексту, який має широкий спектр застосувань. Їх висока точність, гнучкість та масштабованість роблять їх цінним активом для багатьох галузей. Штучні нейронні мережі стали потужним інструментом для розпізнавання тексту.

Постановка задачі:

1. Дослідити теоретичні основи штучних нейронних мереж і їх застосування для розпізнавання тексту.

2. Проаналізувати сучасні підходи та технології розпізнавання тексту. Оглянути існуючі системи розпізнавання тексту: *Google Cloud Vision*, *Microsoft Azure OCR* та *Tesseract*, а також аналіз їх переваг та недоліків.

3. Провести порівняльний аналіз обраних систем розпізнавання тексту.

4. Розробити методику тестування систем розпізнавання тексту. А саме, визначити критерії оцінки ефективності систем та розробити тестові набори даних для проведення експериментів.

5. Проаналізувати отримані результати та зробити висновки.

Виконання поданих задач допоможе отримати глибоке розуміння сучасних технологій розпізнавання тексту з використанням штучних нейронних мереж, а також визначити найефективніші системи для різних застосувань.

2. Вибір методів розв'язання задачі та аналіз

2.1 Вибір архітектури та алгоритму навчання

Побудова інформаційних систем розпізнавання тексту з використанням штучних нейронних мереж починається з вибору відповідної архітектури нейронної мережі та алгоритму навчання. Важливим етапом є вибір архітектури нейронної мережі та алгоритму навчання. Від правильного вибору залежить точність, ефективність та масштабованість вашої системи.

Як працює штучна нейронна мережа? Загальновизнаного визначення, яке б задовольняло всіх, на жаль, на сьогодні не існує. Це пояснюється тим, що нейронні мережі вивчають фахівці з різних галузей науки, а методологічні та термінологічні бар'єри заважають взаєморозумінню. Якщо ми розглядаємо штучну нейронну мережу як певне середовище для обробки інформації, то її можна визначити, визначивши елементи цього середовища та правила їхньої взаємодії.

Інакше кажучи, штучні нейронні мережі - це обчислювальні парадигми, які реалізують спрощені моделі біологічних нейронних мереж (БНМ). Під ШНМ ми розуміємо локальні ансамблі нейронів, які об'єднані синаптичними зв'язками. Набір таких ансамблів утворює мозок з його різноманітними функціональними можливостями.

На сьогодні відома велика кількість нейронних структур та їх модифікацій, що орієнтовані на вирішення конкретного типу задач. Найбільш відомі типи таких структур показані на *рис.2.* [2]

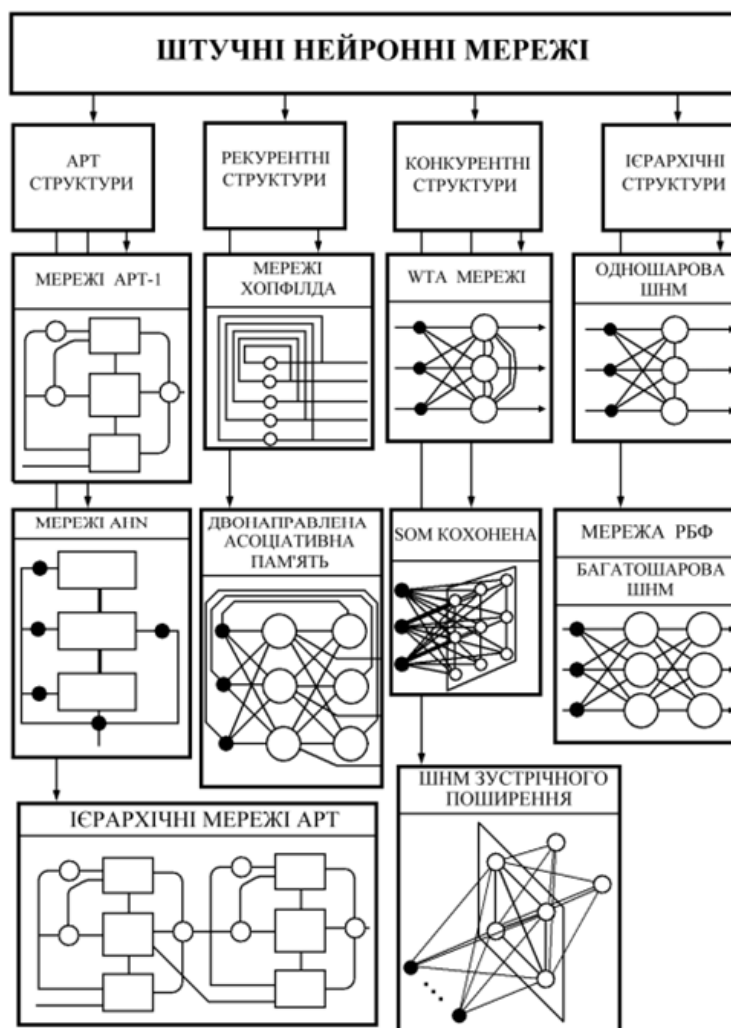


Рис.2

У моделях штучних нейронних мереж (ШНМ) і схемах адресації на основі вмісту є недоліки. Через слабку організацію вони можуть плутати різні об'єкти. Однак це є аналогом звикання, підвищення чутливості до асоціацій, які лежать в основі ментальних особливостей людини. Тобто будь-який комп'ютер, який претендує на звання «розумного», повинен мати такі особливості.

Створені штучні нейрони, які також називають нейронними клітинами, вузлами, модулями, моделюють структуру та функції біологічних нейронів. Структура та особливості штучних нейронних мереж, утворених нейронами, залежать від конкретних завдань, які будуть вирішуватися з їх допомогою.

Архітектури сучасних нейронних мереж найчастіше поділяють на три категорії:

- мережа з повним комплектом міжнейронних зв'язків;
- мережа з фіксованим індексом оточення;
- мережа з пошаровою структурою [2].

Рішення про побудову автомата на основі порогових елементів, подібних до нейронів (нервових клітин), які можуть виконувати логічні функції, було сформульовано понад півстоліття тому Мак-Каллохом і Піттсом. Втім, проблема проектування систем на основі порогових елементів викликала великі труднощі, і її рішення було знайдено лише через 20 років. Вона виявилася настільки складною, що практично виключала можливість синтезу автоматів, що складаються з більш ніж десятків нейронів.

Для розпізнавання тексту найчастіше використовуються наведені нижче архітектури нейронних мереж:

Згорткові нейронні мережі (*CNN*) – поширена архітектура для обробки зображень і розпізнавання тексту, зокрема у випадках, коли текст є частиною зображення. Мережі складаються зі згорткових шарів, шарів з частковою вибіркою та повністю з'єднаних шарів. Згорткові шари виділяють такі особливості, як контури і кути, а шари субдискретизації зменшують розмірність даних, що допомагає зменшити обчислювальну складність.

Архітектура *Transformers*, в тому числі такі моделі, як *BERT* і *GPT*, довела свою високу ефективність для задач обробки природної мови. У них використовується механізм уваги, який дозволяє моделі фокусуватися на різних частинах вхідних даних під час генерування вихідних даних, що дуже корисно для складних задач розпізнавання тексту.

Вибір архітектури та алгоритму навчання залежить від конкретних вимог і обмежень завдання. Наприклад, для розпізнавання тексту, яке може включати обробку зображень або послідовностей символів, можна розглянути можливість використання ШНМ для обробки зображень і трансформерів для обробки послідовностей.

2.2 Обґрунтування вибору архітектури

При створенні інформаційної системи розпізнавання тексту вибір архітектури є ключовим кроком, який значною мірою впливає на точність та ефективність системи. Рішення про використання згорткових нейронних мереж (*CNN*) як основної архітектури в цій роботі обґрунтовано кількома причинами, зокрема, їхньою здатністю ефективно обробляти та розпізнавати текст на зображеннях.

Згорткові нейронні мережі (*CNN*) є одним з найпоширеніших типів штучних нейронних мереж, що використовуються для розпізнавання тексту. Вони спеціально розроблені для обробки зображень і тексту, і можуть досягати високої точності розпізнавання символів, слів і речень.

Поданий різновид алгоритму глибокого навчання, розроблений для аналізу та розпізнавання зображень. Алгоритм обробляє вхідне зображення, присвоюючи значення у вигляді ваг і зсувів різним ознакам і об'єктам на зображенні. Порівняно з деякими іншими алгоритмами класифікації, *CNN* вимагає мінімальної попередньої обробки. Тоді як традиційні підходи створюють фільтри вручну, *CNN* може самостійно вивчати ці фільтри та ознаки. Ця особливість робить ШНМ дуже ефективними в задачах, пов'язаних з класифікацією та аналізом зображень. [9].

У згорткових нейронних мережах використовуються спеціалізовані шари, такі як згорткові, об'єднані та повністю зв'язані шари, що дозволяє їм автоматично виокремлювати важливі особливості та структури на зображеннях, зменшуючи при цьому кількість параметрів та обчислювальну складність моделі. Це робить ШНМ ідеальним вибором для задач оптичного розпізнавання символів (*OCR*), де потрібно обробляти великі обсяги текстових зображень з високою точністю.

Лише уявіть, що ви бачите фотографію папуги. Проте ви не бачите всю картину одразу. Натомість ваш мозок сканує зображення, зосереджуючись на

конкретних деталях, таких як краї, текстури та форми. Далі він порівнює ці деталі зі спогадами про домашніх улюбленців, яких ви бачили раніше, поступово вибудовуючи загальний образ яскравого ара, що стоїть перед вами.

Саме за таким принципом працюють згорткові нейронні мережі. Такий тип штучної нейронної мережі натхненний зоровою корою людського мозку і призначений для обробки сітчастих даних, таких як зображення, відео і навіть звукові хвилі. Нейрони людського мозку реагують на стимули в певних ділянках зорового поля, які називаються рецептивними полями. Кора головного мозку складається з багатьох таких рецептивних полів, які перекриваються, охоплюючи всю зорову кору. У дизайні *CNN* імітується ця структура, щоб ефективно аналізувати та розпізнавати візуальні патерни у вхідних даних.

Згорткові нейронні мережі володіють численними перевагами, які роблять їх оптимальним вибором для задач розпізнавання тексту. До основних переваг можна віднести наступні:

Екстракція ознак: ШНМ самостійно виділяють важливі особливості зображення, такі як контури, кути і текстури, які є важливими для розпізнавання тексту. Завдяки використанню шарів, що згортаються, модель може виявляти ці особливості на різних рівнях абстракції.

Скорочення розмірності: Шари субдискретизації (об'єднання) зменшують розмірність даних, зберігаючи при цьому суттєві ознаки. Таким чином зменшується обчислювальна складність і запобігається надмірне припасування моделі.

Незмінність до перетворень: ШНВ характеризуються властивістю інваріантності до масштабів, поворотів та інших афінних перетворень. Цей принцип важливий для розпізнавання тексту, оскільки текст може бути представлений у різних формах і розмірах.

Місцеві рецептивні поля дозволяють ШНМ ефективно обробляти зображення, фокусуючись на невеликих ділянках вхідного зображення. Так

мережа може виявляти деталі тексту на рівні пікселів, що підвищує точність розпізнавання.

Штучні мережі широко використовуються в задачах комп'ютерного зору і показують високу ефективність у розпізнаванні об'єктів, включаючи текст. Їх здатність обробляти великі масиви зображень робить їх ідеальними для побудови систем розпізнавання тексту.

Навчання штучних нейронних мереж відбувається на великій кількості маркованих текстових зображень за допомогою алгоритмів оптимізації, таких як *Adam* або *RMSprop*. Регуляризація та методи доповнення даних використовуються для запобігання надмірному пристосуванню та покращення загальної здатності моделі до узагальнення.

Наведемо приклади програм, які використовують архітектуру CNN для розпізнавання тексту:

Google Tesseract – одна з найпопулярніших систем оптичного розпізнавання символів, яка використовує архітектуру CNN для підвищення точності розпізнавання тексту. Нейронні мережі допомагають *Tesseract* розпізнавати текст різними мовами і шрифтами, а також справлятися зі складними текстовими зображеннями, які можуть містити шум і спотворення [11].

ABBYY FineReader – програма для розпізнавання тексту, яка також використовує CNN для ідентифікації тексту. *FineReader* точно розпізнає текст на відсканованих документах, включаючи складні багатомовні тексти, забезпечуючи високу якість цифрових копій паперових документів [7].

Microsoft Azure Cognitive Services забезпечує службу комп'ютерного зору, яка використовує CNN для різноманітних завдань розпізнавання зображень, серед яких і розпізнавання тексту. Служба дозволяє розробникам легко інтегрувати OCR у свої програми, забезпечуючи точне та швидке розпізнавання тексту.

Adobe Acrobat Reader підтримує CNN, щоб розширити можливості розпізнавання тексту в своїх продуктах. Використовуючи CNN, Adobe

Acrobat може точно розпізнавати текст у PDF-документах, включаючи рукописний текст і текст зі складних макетів, що полегшує роботу з цифровими документами.

CamScanner – мобільний додаток для сканування документів, який використовує згорткову мережу для розпізнавання тексту на зображеннях, знятих камерою смартфона. Користувачі можуть легко конвертувати фотографії з документів у текстові файли, забезпечуючи високу точність розпізнавання навіть при поганій якості зображення або на складних фонах [11].

Ці приклади демонструють, як *CNN* використовуються в різних програмах і сервісах для покращення точності та ефективності розпізнавання тексту, підтверджуючи їхню придатність для задач *OCR* та обробки зображень. Отже, використання цих мереж для побудови інформаційної системи розпізнавання тексту є оптимальним вибором, що забезпечує високу якість розпізнавання, ефективність обробки та широкий спектр можливих застосувань.

2.3 Огляд та аналіз існуючих систем розпізнавання тексту: *Google Cloud Vision, Microsoft Azure OCR, Tesseract*

Оптичне розпізнавання тексту – процес автоматичного перетворення текстових зображень у цифровий текст. Системи *OCR* широко використовуються в різних сферах, таких як обробка документів, сканування, переклад, забезпечення доступності та багато інших. У даному дослідженні розглядаються комерційні системи розпізнавання тексту з відкритим вихідним кодом: *Google Cloud Vision, Microsoft Azure OCR та Tesseract*.

Огляд систем

1. Google Cloud Vision забезпечує потужні можливості розпізнавання зображень, включаючи *OCR*.

Особливості:

- Підтримка множинних мов.
- Висока точність розпізнавання.
- Інтеграція з іншими сервісами *Google Cloud*.
- Можливість розпізнавання тексту на зображеннях, сканах, фотографіях.

2. *Microsoft Azure OCR* є частиною *Azure Cognitive Services*.

Особливості:

- Підтримка багатьох мов.
- Інтеграція з іншими сервісами *Azure*.
- Можливість обробки великих обсягів даних.
- Висока точність розпізнавання тексту.

3. *Tesseract* – це відкрита *OCR*-система, яка підтримує безліч мов і типів тексту. Вона активно розвивається спільнотою з відкритим вихідним кодом.

Особливості:

- Підтримка численних мов.
- Відкритий вихідний код.
- Можливість налаштування і розширення.
- Добре підходить для розробників і дослідників.

Критерії оцінки

Для оцінки ефективності систем *OCR* використовуються наступні критерії:

- Точність: Вимірюється як відсоток правильно розпізнаних символів або слів.
- Швидкість обробки: Час, необхідний для обробки певного обсягу даних.
- Підтримка мов: Кількість і різноманітність мов, які підтримує система.
- Можливість інтеграції: Легкість інтеграції з іншими системами та сервісами.

- Вартість: Витрати на використання системи.
- Гнучкість і налаштування: Можливість налаштування системи під специфічні потреби користувача.

Таблиця Порівняльний аналіз

	<i>Google Cloud Vision</i>	<i>Microsoft Azure OCR</i>	<i>Tesseract</i>
<u>Точність</u>	Висока точність розпізнавання тексту, особливо для друкованого тексту. Точність розпізнавання рукописного тексту залежить від якості зображення.	Схожа висока точність для друкованого тексту. Точність рукописного тексту також залежить від якості зображення.	Точність залежить від якості вхідних зображень та налаштувань. Для друкованого тексту точність висока, але для рукописного тексту може бути нижчою.
<u>Швидкість обробки</u>	Висока швидкість обробки, особливо при використанні хмарної інфраструктури Google.	Висока швидкість обробки завдяки хмарній інфраструктурі Azure.	Швидкість залежить від обчислювальних потужностей локальної машини. Може бути повільнішою в порівнянні з хмарними рішеннями.
<u>Підтримка мов</u>	Підтримує більше	Підтримує понад	Підтримує більше

	50 мов.	25 мов.	100 мов.
<u>Можливість інтеграції</u>	Легка інтеграція з іншими сервісами Google Cloud, API для розробників.	Легка інтеграція з іншими сервісами Azure, API для розробників.	Вимагає більше зусиль для інтеграції, але надає можливості для налаштування і розширення.
<u>Вартість</u>	Платний	Платний	Безкоштовний
<u>Гнучкість і налаштування</u>	Обмежені можливості для налаштування, орієнтовані на простоту використання.	Обмежені можливості для налаштування, орієнтовані на простоту використання.	Висока гнучкість і можливість налаштування, що робить його привабливим для дослідників і розробників.
<u>Використання в різних галузях</u>	<i>Бізнес-аналітика:</i> Автоматизація обробки документів, витягування тексту з рахунків, квитанцій тощо. <i>Охорона здоров'я:</i> Обробка медичних записів, розпізнавання тексту на	<i>Фінанси:</i> Автоматизація обробки фінансових документів, виписок, контрактів. <i>Юриспруденція:</i> Обробка юридичних документів, контрактів, судових справ.	<i>Дослідження та розвиток:</i> Використання в наукових дослідженнях, обробка архівних документів. <i>Невеликі підприємства:</i> Автоматизація обробки тексту для внутрішніх потреб,

	зображеннях медичних приладів. <i>Роздрібна торгівля:</i> Обробка інформації цінників, реклами, упаковок.	<i>Освіта:</i> Оцифрування навчальних матеріалів, тестів, анкет.	оцифрування документів. <i>Громадські організації:</i> Обробка тексту для некомерційних проєктів, архівування даних.
--	--	--	---

Графіки для порівняння систем OCR за точністю розпізнавання, швидкістю обробки та використанням пам'яті

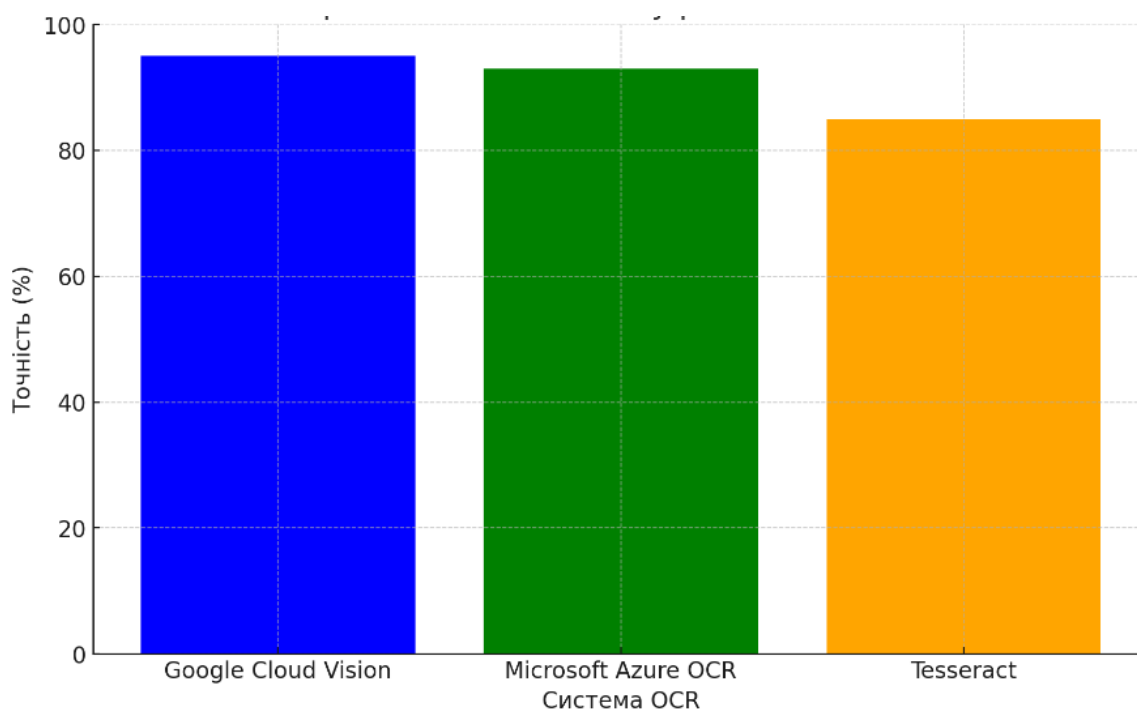


Рис3. Точність розпізнавання тексту

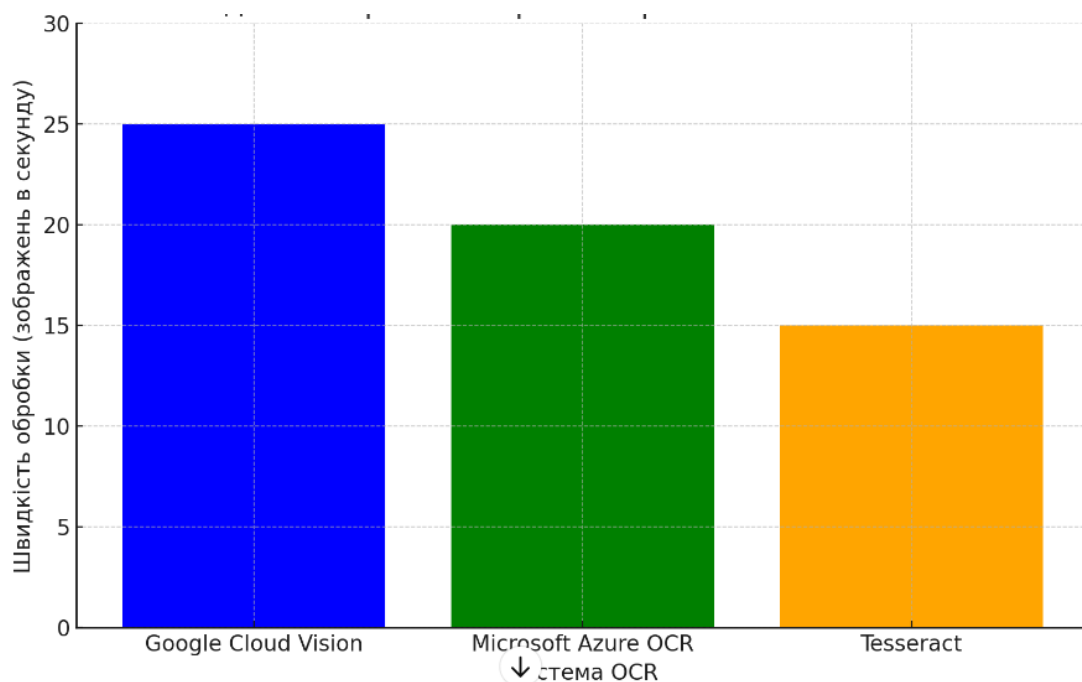


Рис.4 Швидкість обробки зображень

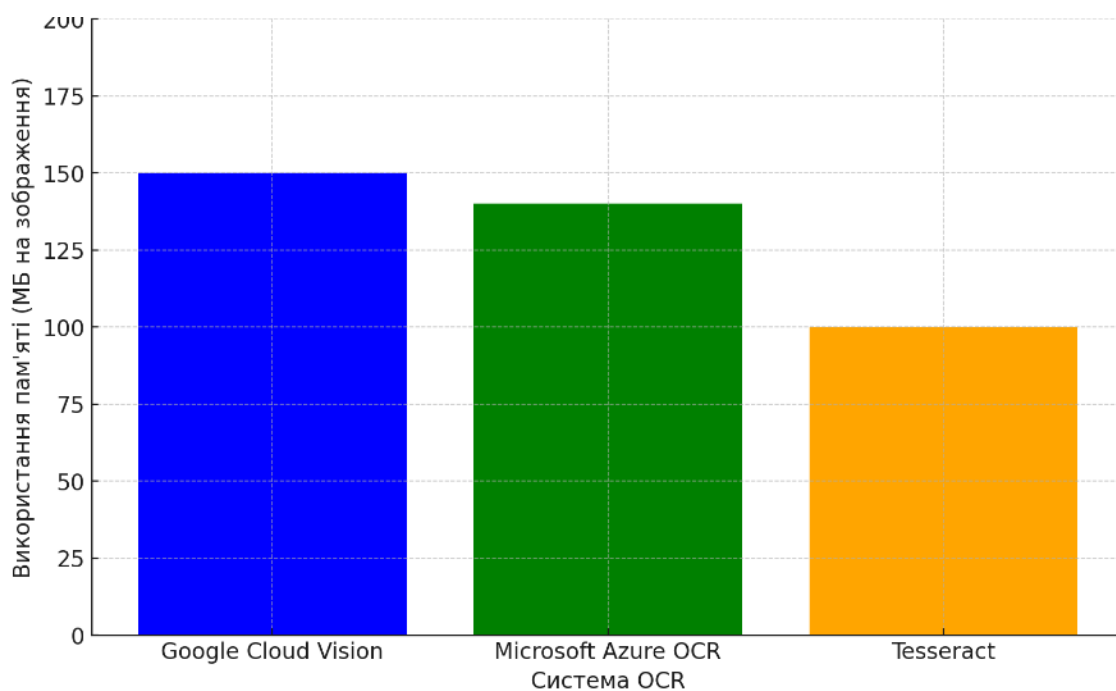


Рис.5 Використання пам'яті

Отже, *Google Cloud Vision* та *Microsoft Azure OCR* є потужними комерційними системами з високою точністю, швидкістю обробки та підтримкою багатьох мов. Вони добре інтегруються з іншими хмарними сервісами, але вимагають оплати за використання. *Tesseract* є безкоштовною й гнучкою системою з відкритим кодом, яка добре підходить для

дослідницьких проектів і невеликих підприємств, але може вимагати більше зусиль для інтеграції та налаштування.

ВИСНОВКИ

Припустімо світ, де машини можуть не лише бачити, але й по-справжньому розуміти те, що вони бачать. У ньому комп'ютери можуть аналізувати медичні знімки і діагностувати хвороби з надлюдською точністю, або вести безпілотний автомобіль по жвавих міських вулицях з точністю досвідченого водія. Саме це майбутнє обіцяють згорткові нейронні мережі - технології, яка стрімко змінює наш час.

Результати застосування згорткових нейронних мереж вже дійсно вражають. Вони покладені в основу програмного забезпечення для розпізнавання облич, розблокування наших смартфонів і забезпечення безпеки наших будинків. На них базується медична візуалізація, що дозволяє лікарям виявляти пухлини та інші аномалії з безпрецедентною точністю. Їх використовують навіть у мистецтві та музиці, розширюючи межі людського самовираження.

У процесі дослідження були вивчені та описані основні поняття штучних нейронних мереж, їх типи та застосування для задач розпізнавання тексту. Значну увагу приділено архітектурам, які ефективно працюють з текстовими даними, таким як *CNN*, *RNN* та *Transformers*. Це дозволило зрозуміти, які технології найбільше підходять для вирішення поставленої задачі.

Також було проведено детальний аналіз та вибір архітектури *CNN* як найбільш ефективної для розпізнавання тексту на зображеннях. ШНМ дозволяють виділити значущі ознаки, зменшити розмірність даних та досягти інваріантності до перетворень, що є критично важливим для задач розпізнавання тексту. Наведено сучасні алгоритми навчання, такі як *Adam* та *RMSprop*, які забезпечують стабільне та швидке навчання моделі.

У поданій роботі показано, що *CNN* можуть ефективно розпізнавати текст навіть на складних зображеннях, що підтверджується прикладами

успішних програм, таких як *Google Tesseract*, *ABBYY FineReader*, *Microsoft Azure Cognitive Services* та інші.

Крім того, було проаналізовано та порівняно системи розпізнавання тексту (OCR) *Google Cloud Vision*, *Microsoft Azure OCR* та *Tesseract*. Платні системи, такі як *Google Cloud Vision* та *Microsoft Azure OCR*, забезпечують високу точність, швидкість та інтеграцію, але вимагають внесення грошових внесків за використання. Система *Tesseract* з відкритим вихідним кодом пропонує високу гнучкість і економію коштів, але може вимагати більше зусиль для досягнення бажаної точності і швидкості.

Практика використання CNN в реальних додатках демонструє високу ефективність і точність розпізнавання тексту, що відкриває нові можливості для автоматизації документообігу, аналізу текстових даних та інших сфер. Подальше вдосконалення системи можливе за рахунок інтеграції з іншими сучасними технологіями, такими як *Transformers*, що дозволить розширити функціонал і підвищити точність розпізнавання.

Загалом, результати курсової роботи підтвердили, що згорткові нейронні мережі є потужним інструментом для розпізнавання тексту і можуть бути успішно застосовані для розробки ефективних інформаційних систем. Подальші дослідження та вдосконалення в цьому напрямку сприятимуть розвитку технологій розпізнавання тексту, підвищенню їх точності та розширенню можливостей.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Зайвелев Ю., Яковлева А. Застосування нейронних мереж для вирішення задачі розпізнавання емоційного забарвлення тексту. Грааль науки. 2021. С. 285–287. URL: <https://doi.org/10.36074/grail-of-science.19.11.2021.054> (дата звернення: 21.05.2024).
2. Новотарський М.А., Нестеренко Б.Б. Штучні нейронні мережі: обчислення // Праці Інституту математики НАН України. – Т50. – Київ: Ін-т математики НАН України, 2004. – 408 с.
3. Переяславська С., Шевченко В., Смагіна О. (2021). АНАЛІЗ ПІДХОДІВ ДО РОЗПІЗНАВАННЯ ТЕКСТОВОЇ ІНФОРМАЦІЇ У ТЕХНОЛОГІЇ OCR. Режим доступу: <https://ojs.ukrlogos.in.ua/index.php/interconf/article/view/9924>
4. Тема 15. Архітектура нейрокомп'ютерів. URL: https://e-tk.lntu.edu.ua/pluginfile.php/19335/mod_resource/content/0/Тема%2015.%20Архітектура%20нейрокомп'ютерів.pdf (дата звернення: 08.05.2024).
5. Учасники проєктів Вікімедіа. Оптичне розпізнавання символів – Вікіпедія. URL: https://uk.wikipedia.org/wiki/Оптичне_розпізнавання_символів (дата звернення: 21.05.2024).
6. Учасники проєктів Вікімедіа. Штучний нейрон – Вікіпедія. URL: https://uk.wikipedia.org/wiki/Штучний_нейрон (дата звернення: 21.05.2024).
7. Учасники проєктів Вікімедіа. ABBYY FineReader – Вікіпедія. URL: https://uk.wikipedia.org/wiki/ABBYY_FineReader (дата звернення: 21.05.2024).
8. Штучні нейронні мережі. Короткі теоритичні відомості. Державний університет "Житомирська політехніка" - Освітній портал. URL: <https://learn.ztu.edu.ua/mod/resource/view.php?id=17833>(дата звернення: 22.05.2024).

9. Що таке згорткові нейронні мережі (CNN, ConvNet)? | TheTransmitted. TheTransmitted. URL: <https://thetransmitted.com/adlucem/shho-take-zgortkovi-nejronni-merezhi-cnn-convnet/> (дата звернення: 21.05.2024).

10. CamScanner. CamScanner: text and image scanning and recognition, PDF to Word, document format conversion, online editor. URL: <https://www.camscanner.com/> (date of access: 14.05.2024).

11. Contributors to Wikimedia projects. Tesseract (software) - Wikipedia. Wikipedia, the free encyclopedia. URL: [https://en.wikipedia.org/wiki/Tesseract_\(software\)](https://en.wikipedia.org/wiki/Tesseract_(software)) (date of access: 21.05.2024).

12. Haykin S. S. Neural networks: a comprehensive foundation. 2nd ed. Upper Saddle River, NJ : Prentice Hall, 1999. 842 p.

13. Zdebska T., Andrunyk V., Kempnyk R., Chyhyra V. Optical Character Recognition. URL: http://ena.lp.edu.ua:8080/bitstream/ntb/52204/2/2020v2_Zdebska_T-Optical_Character_Recognition_90-100.pdf. (дата звернення: 20.05.2024)., с. 94