

**ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ІНФОРМАЦІЙНИХ
ТЕХНОЛОГІЙ
КАФЕДРА ШТУЧНОГО ІНТЕЛЕКТУ**

КВАЛІФІКАЦІЙНА РОБОТА

на тему: «Розробка інформаційної системи розпізнавання рукописних документів з використанням алгоритмів глибокого навчання»

на здобуття освітнього ступеня бакалавра

зі спеціальності 122 Комп'ютерні науки

освітньо-професійної програми Штучний інтелект

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

_____ Дмитро РАДКОВСЬКИЙ

Виконав: здобувач(ка) вищої освіти гр. ШІД-41
Дмитро РАДКОВСЬКИЙ

Керівник: Євген ЧИЧКАРЬОВ
*Доктор т.н,
професор*

Рецензент: _____

Київ 2024

**ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
Навчально-науковий інститут інформаційних технологій**

Кафедра Штучного інтелекту
Ступінь вищої освіти Бакалавр
Спеціальність Комп'ютерні науки
Освітньо-професійна програма Штучний інтелект

ЗАТВЕРДЖУЮ
Завідувач кафедрою Штучного інтелекту

Ольга ЗІНЧЕНКО
«___» _____ 20__ р.

**ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ**

Радковський Дмитро Олександрович

1. Тема кваліфікаційної роботи: Розробка інформаційної системи розпізнавання рукописних документів з використанням алгоритмів глибокого навчання

керівник кваліфікаційної роботи Євген ЧИЧКАРЬОВ Доктор технічних наук, професор.

затвержені наказом Державного університету інформаційно-комунікаційних технологій від «27» 02.2024р. № 36

2. Строк подання кваліфікаційної роботи «31» травня 2024 р.

3. Вихідні дані до кваліфікаційної роботи: науково-технічна література, параметри алгоритмів глибокого навчання

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)

Дослідження принципів алгоритмів глибокого навчання
Аналіз технологій машинного навчання та можливості застосування
Розвиток нейромережевої моделі

5. Перелік ілюстративного матеріалу:

1. Розвиток нейромережевої моделі
2. Характеристики нейромережевої моделі
3. Розпізнавання рукописних документів

6. Дата видачі завдання «27» лютого 2024 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1	Збір даних	26.02-22.03.24	Виконано
2	Обробка матеріалу	23.03-04.04.24	Виконано
3	Аналіз і навчання нейронні мережі	06.04-14.04.24	Виконано
4	Висновки за результатом аналізу	15.04-17.04.24	Виконано
5	Розробка телеграм бота	19.04-05.05.24	Виконано
6	Розробка презентації	06.05-10.05.24	Виконано
7	Передзахист	14.05-17.05.24	Виконано
8	Здача в деканат	25.05-01.06.24	Виконано

Здобувач(ка) вищої освіти _____

Дмитро РАДКОВСЬКИЙ

Керівник
кваліфікаційної роботи _____

Свген ЧИЧКАРЬОВ

РЕФЕРАТ

Текстова частина кваліфікаційної роботи на здобуття освітнього ступеня бакалавра: 58 стор., 20 рис., 10 джерел.

Мета роботи – створення системи, яка може розпізнавати рукописні документи, використовуючи сучасні алгоритми глибокого навчання.

Об'єкт дослідження – алгоритми глибокого навчання.

Предмет дослідження – розпізнавання рукописних документів.

Короткий зміст роботи: У роботі проведено огляд існуючих технологій розробки апаратно–програмного комплексу розпізнавання рукописних документів, використовуючи сучасні алгоритми глибокого навчання.

Проведено аналіз статистики використання месенджерів на смартфонах та обрано для реалізації месенджер телеграм, для реалізації програмного продукту мови програмування вибрано мову Python;

Використано telebot, Google Cloud Vision API. Для управління системою необхідно мати ПК з операційною системою Windows, а щоб скористатися системою потрібен лише телефон с телеграмом.

КЛЮЧОВІ СЛОВА: ШТУЧНА НЕЙРОННА МЕРЕЖА; ОПТИЧНЕ РОЗПІЗНАВАННЯ СИМВОЛІВ; ТЕЛЕГРАМ-БОТ; НАБІР ДАНИХ MNIST; ЗГОРТКОВА НЕЙРОННА МЕРЕЖА

ЗМІСТ

ВСТУП	9
1.1 Загальні відомості про задачу розпізнавання текст	12
1.2 Аналіз літературних джерел	13
1.2.1 Офлайн розпізнавання цифр в англійській мові з використанням кореляційного методу	13
1.2.2 Розпізнавання рукописних символів хінді за допомогою нейронної мережі зі зворотним поширенням	14
1.2.3 Розпізнавання символів деванагарі за допомогою нейронних мереж	14
1.2.4 Система інтелектуальна нелінійного розпізнавання символів	14
1.2.5 Система розпізнавання символів рукописного тексту на основі нечітких обчислень	14
1.2.6 Використання сегментації та штучних нейронних мереж для розпізнавання почерку деванагарі	15
1.2.7 Система розпізнавання символів рукописного тексту Gurumukhi	15
1.2.8 Попередня обробка зображень для оптичного розпізнавання символів за допомогою нейронних мереж	15
1.2.9 Розпізнавання рукописних англійських літер	16
1.2.10 Технологія розпізнавання рукописних цифр за допомогою послідовної згорткової нейронної мережі	16
1.3 Огляд існуючих рішень	16
1.3.1 Istio.com Handwriting Input	16
1.3.2 MyScript Nebo	17
1.3.3 Img2txt.com	18
2. ПРОЕКТУВАННЯ СИСТЕМИ	21
2.1. Огляд предметної області	21
2.2 Постановка задач і цілей	22
2.3 Опис функціональності системи	22
2.4 Опис нейронної мережі	23
2.4.1 Спосіб первинної обробки	26
2.4.2 Сегментація	28
2.5 Передумови	29
2.6 Діаграми послідовності	30
2.7 Висновки до розділу 2	35
3 ОПИС РОЗРОБЛЕНОЇ СИСТЕМИ	37
3.1 Вибір технологій та їх обґрунтування	37
3.1.1 Вибір платформи для системи	37

3.1.2 Обґрунтування вибору месенджера Telegram	37
3.1.3 Пошук мови для програмування	39
3.1.4 Вибір допоміжних бібліотек	40
3.2 Структурні модулі розробленої системи включають наступні компоненти:	41
3.3. Нейромережева модель	42
3.4 Опис графічного інтерфейсу користувача системи	44
3.5 Висновки до розділу 3	47
ПЕРЕЛІК ПОСИЛАНЬ	49
Додаток А	51
ДЕМОНСТРАЦІЙНІ МАТЕРІАЛИ	54

ВСТУП

Розпізнавання рукописного тексту - це нелегке завдання, для якого треба навчити нейронну мережу аналізу та виявлення рукописних символів. Для початку треба дати можливість штучному інтелекту зрозуміти рукописний текст так само, як і людині, щоб вона могла читати, писати і працювати з текстом. Розпізнавання рукописного тексту вважається одним з найскладніших і найцікавіших напрямків в цій області. Воно має великий вплив на розвиток автоматизації та покращення взаємодії людини з комп'ютером у багатьох сферах.

Існує два головних типи для розпізнавання тексту це онлайн і офлайн. При онлайн розпізнаванні дані пишуться прямо на сайті або в програмі в реальному часі, коли вони пишуться за допомогою спеціального пера в спеціальному окні. Офлайн-розпізнавання використовується вже для відсканованих документів або фото, які зберігаються в цифровому форматі.

Онлайн-розпізнавання тексту використовується переважно для рукописних символів і дозволяє вводити текст на екрані, використовуючи рукописний почерк. Це важливо для покращення взаємодії між людиною та комп'ютером. З іншого боку, офлайн-розпізнавання широко використовується для обробки документів у офісах, на поштах та інших сферах, де потрібно розпізнавати рукописний текст.

Багато додатків, що використовуються в різних галузях, від сортування пошти до обробки даних у офісах, вимагають розпізнавання рукописного тексту в режимі офлайн.

Впровадження технології розпізнавання рукописного тексту відкриває нові горизонти в багатьох сферах діяльності. Наприклад, в освіті розпізнавання рукописного тексту може допомогти автоматично оцінювати роботи студентів, зменшуючи тим самим навантаження на викладачів. Це дозволяє їм зосередитися на навчальному процесі, а не на рутинній роботі. Системи OCR також можна інтегрувати з текстовими процесорами та інструментами редагування, що полегшує створення цифрових копій рукописних нотаток.

Також у медичній галузі розпізнавання рукописного тексту може значно покращити управління медичними записами та картами. Лікарі часто використовують рукописні нотатки для зберігання історії хвороби пацієнта, а можливість автоматичного оцифрування цих нотаток може значно спростити доступ до інформації та покращити обслуговування пацієнтів.

Автоматичне розпізнавання тексту також може зменшити кількість помилок, пов'язаних з нерозбірливим почерком, що має вирішальне значення в медичній сфері. У юридичній сфері технологія розпізнавання рукописного тексту допомагає обробляти великі обсяги документів, значно прискорюючи роботу юристів і знижуючи ризик втрати важливих документів. OCR дає можливість швидко оцифрувати рукописні контракти, угоди та інші юридичні документи, забезпечуючи їх безпечне зберігання та легкий доступ до них. Варто також згадати про потенційне використання розпізнавання рукописного тексту у сфері архівування та збереження історичних документів.

Оцифрування рукописних документів дозволяє зберігати їх у цифровому вигляді, роблячи їх доступними для майбутніх поколінь. Це особливо важливо для збереження культурної спадщини та історичних знань. Не менш важливим є використання технології розпізнавання рукописного тексту в повсякденному житті. Наприклад, багато сучасних додатків для мобільних телефонів використовують цю технологію для зручного введення тексту. Це дозволяє користувачам швидко і легко вводити текст, використовуючи свій почерк, що значно покращує зручність використання цих додатків.

Загалом, розпізнавання рукописного тексту - важлива технологія, яка має широкий спектр застосувань у різних сферах. Вона допомагає автоматизувати рутинні процеси, підвищує якість роботи з текстовими даними і відкриває нові можливості для взаємодії людини з комп'ютером. Подальший розвиток цієї технології, безсумнівно, матиме значний вплив на наше повсякденне життя і роботу. Однак варто зазначити, що розробка технології розпізнавання рукописного

тексту вимагає значних ресурсів і зусиль. Для навчання нейронних мереж потрібна велика кількість даних, включаючи різні зразки почерку, різні стилі письма та мови.

Також необхідно враховувати особливості мови, якою написаний текст, і символи, що використовуються в різних алфавітах. Це робить процес навчання нейронних мереж складним і багаторівневим. Однією з найбільших проблем є необхідність обробки великих обсягів даних для забезпечення високої точності розпізнавання. Це вимагає значних обчислювальних ресурсів і часу. Також важливо забезпечити надійність і точність роботи системи в різних умовах, наприклад, при поганій якості зображення або нерозбірливому почерку. Незважаючи на ці виклики, досягнення в цій галузі продовжують зростати і надають нові можливості для використання технології розпізнавання рукописного тексту.

Отже, розпізнавання рукописного тексту є важливою складовою сучасної інформатики та штучного інтелекту, що сприяє розвитку автоматизації та покращенню взаємодії людини з комп'ютером. Вона має величезний потенціал для застосування в різних сферах, від освіти і медицини до бізнесу і особистого використання. Подальший розвиток цієї технології матиме значний вплив на наше життя, зробивши його більш комфортним та ефективним

1 ОГЛЯД ІСНУЮЧИХ РІШЕНЬ

1.1 Загальні відомості про задачу розпізнавання текст

Цей розділ присвячений огляду існуючих рішень та підходів до розпізнавання рукописного тексту. Основна мета нашого дослідження полягає в тому, щоб знайти нові пошукові метод для розпізнавання рукописного тексту різних шрифтів та стилів шляхом вдосконалення традиційних штучних нейронних мереж (ШНМ). ШНМ успішно використовуються для різноманітних завдань, таких як розпізнавання образів, асоціації та класифікації, прогнози дослідження та керування додатками.

Оптичне розпізнавання символів (OCR) - це електронний процес перекладу зображень друкованого, машинописного або рукописного написання в текстовий формат. Ця технологія вже повністю експлуатуються і використовується для розпізнавання образів, штучного інтелекту та машинного зору. Однак пряме застосування оптичного розпізнавання тексту рукописних символів залишається проблемою через високі спотворення, що призводить до низької точності читання.

Розпізнавання рукописного тексту онлайн передбачає автоматичне перетворення рукописного тексту в текстовий формат в реальному часі за допомогою спеціального цифрового перетворювача. Такий підхід дозволяє отримати текстові дані, які можна використовувати для обробки тексту в програмах.

Наслідок сканування, сегментації та класифікації рукописних документів призводить до складних проблем розпізнавання рукописного тексту офлайн. Різні люди мають різні стилі рукописного вводу, що ускладнює процес розпізнавання. В минулих дослідженнях штучні нейронні мережі виявилися ефективними для розпізнавання друкованих символів та рукописних цифр, але у розпізнаванні рукописних слів досягнення не завжди були задовільними.

Штучна нейронна мережа (рис 1.1) (ШНМ) моделює біологічну нервову систему, ідею якої запозичено з мозку людини. Ця модель має велику кількість взаємопов'язаних обробляючих елементів, що працюють разом для вирішення конкретних завдань. ШНМ може навчатися на прикладі і використовувати отримані знання для розв'язання нових задач. Переваги ШНМ включають адаптивне навчання, самоорганізацію, операції в реальному часі та відмовостійкість через зайве кодування інформації.

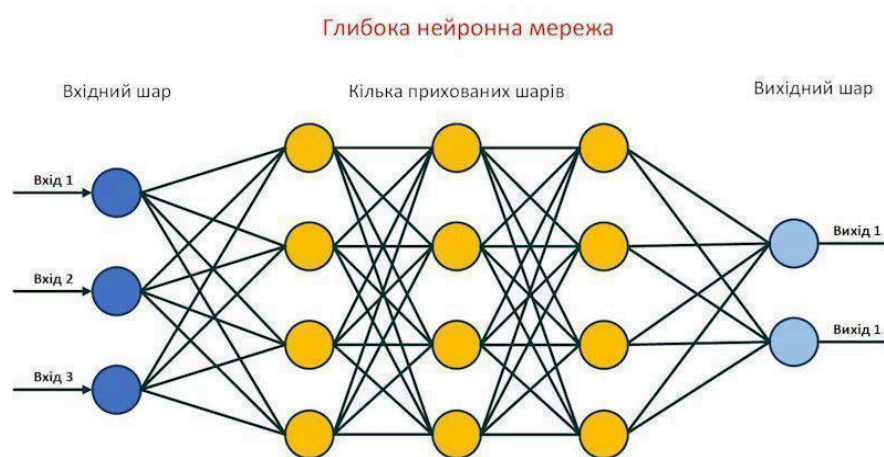


Рисунок 1.1 – Приклад глибокої нейронної мережі

У подальших розділах ми розглянемо методи, як вдосконалити розпізнавання рукописного тексту з використанням ШНМ та інших підходів.

1.2 Аналіз літературних джерел

1.2.1 Офлайн розпізнавання цифр в англійській мові з використанням кореляційного методу

У цьому дослідженні розглядається вдосконалення системи розпізнавання рукописних цифр, що передбачає вищу точність порівняно з попередніми роботами. Також розглядається проблема розпізнавання декількох чисел одночасно, що є важливим аспектом для підвищення ефективності системи.[1]

1.2.2 Розпізнавання рукописних символів хінді за допомогою нейронної мережі зі зворотним поширенням

У статті запропоновано автономну систему розпізнавання рукописного тексту гінді на основі нейронних мереж. Зазначено, що нейронні мережі можуть

ефективно розпізнавати рукописні символи завдяки своїй нечутливості до пропущених даних.[2]

1.2.3 Розпізнавання символів деванагарі за допомогою нейронних мереж

У цьому дослідженні обговорюється підхід як розпізнати символи деванагарі за допомогою складного багат шарового персептрона з прихованим шаром. Підкреслюється складність розпізнавання символів через їхню різноманітну форму та структуру.[3]

1.2.4 Система інтелектуальна нелінійного розпізнавання символів

Стаття присвячена всебічному огляду існуючих робіт з розпізнавання символів рукописного тексту на основі м'яких обчислень. Підкреслюється важливість нелінійного підходу до цієї проблеми.[4]

1.2.5 Система розпізнавання символів рукописного тексту на основі нечітких обчислень

У статті запропоновано метод нечіткого розпізнавання символів. Нечіткі множини та логіка використовуються як основа для представлення та розпізнавання символів. У статті описано нечіткий алгоритм, який спочатку сегментує символ, потім використовує нечітку систему для надання можливих символів, які відповідають заданому вхідному сигналу, а потім використовує систему фазового зсуву для остаточної ідентифікації символів.[5]

1.2.6 Використання сегментації та штучних нейронних мереж для розпізнавання почерку деванагарі

Розпізнавання рукописного тексту - це здатність комп'ютера отримувати та розуміти рукописні дані з різних джерел, таких як паперові документи, фотографії та сенсорні екрани. Відомо, що рукописні символи маратхі складніше розпізнавати, ніж англійські, через їх різноманітний стиль та форму. Метою цього дослідження є впровадження нового методу розпізнавання рукописних символів деванагарі за допомогою сегментації та штучних нейронних мереж. Цей процес складається з двох основних етапів: сегментація символів на рядки, слова та окремі символи, а потім їх розпізнавання за допомогою нейронної мережі.[6]

1.2.7 Система розпізнавання символів рукописного тексту Gurumukhi

У цій статті описано систему розпізнавання рукописних символів Гурмухі, що базується на використанні статистичних ознак, таких як щільність площі, гістограма проекції та вісім спрямованих функцій щільності площі, а також геометричних ознак, таких як площа, периметр та ексцентриситет. Методи бінаризації та орфографічні операції, такі як ерозія та дилатація, використовуються для позбавлення від шуму та подальшої сегментації на окремі символи. Зазначається, що найвища точність досягається завдяки цим ознакам та класифікатору зворотного поширення і становить 98%.[7]

1.2.8 Попередня обробка зображень для оптичного розпізнавання символів за допомогою нейронних мереж

Основною метою даного дослідження є створення теоретичної та практичної бази для попередньої обробки друкованого тексту для подальшого оптичного розпізнавання символів за допомогою нейронної мережі. Було розроблено демонстраційний додаток та налаштовано його параметри на основі результатів експериментів.[8]

1.2.9 Розпізнавання рукописних англійських літер

Розпізнавання символів - одна з найцікавіших і найскладніших областей досліджень в обробці зображень. Розпізнавання англійських символів інтенсивно вивчалось в останні півстоліття, оскільки воно має багато застосувань, таких як перевірка документів, читання банківських депозитів тощо. У цій статті розглянуто різні способи виявлення рукописних англійських літер, беручи до уваги їхній різноманітний стиль і форму, що ускладнює завдання.[9]

1.2.10 Технологія розпізнавання рукописних цифр за допомогою послідовної згорткової нейронної мережі

У цьому дослідженні для розпізнавання рукописного тексту використовується згорткова нейронна мережа з п'ятишаровою архітектурою. Для навчання використовується набір даних з бази даних MNIST, що містить багато зображень рукописних цифр. Перед навчанням дані готуються шляхом

нормалізації та кодування в унітарний кодовий формат. Після навчання досягнута точність розпізнавання становить 99,7%. [10]

1.3 Огляд існуючих рішень

1.3.1 Istio.com Handwriting Input

Istio.com Handwriting Input, інструмент, доступний виключно на сайті, перекладає рукописні нотатки коли ви загрузаєте свій файл на сайті. Після загрузки файла з'являться кілька панелей налаштувань, де ви зможете вибрати мову, що дозволить використовувати інструмент з іншими програмами для введення тексту. Окрім підтримки понад 100 мов, програма дозволяє легко вводити ідеографічні та голосові символи, а також може перевірити ваш текст і аналізувати ваш текст. Приклад вікна введення показано на рисунку 1.2.

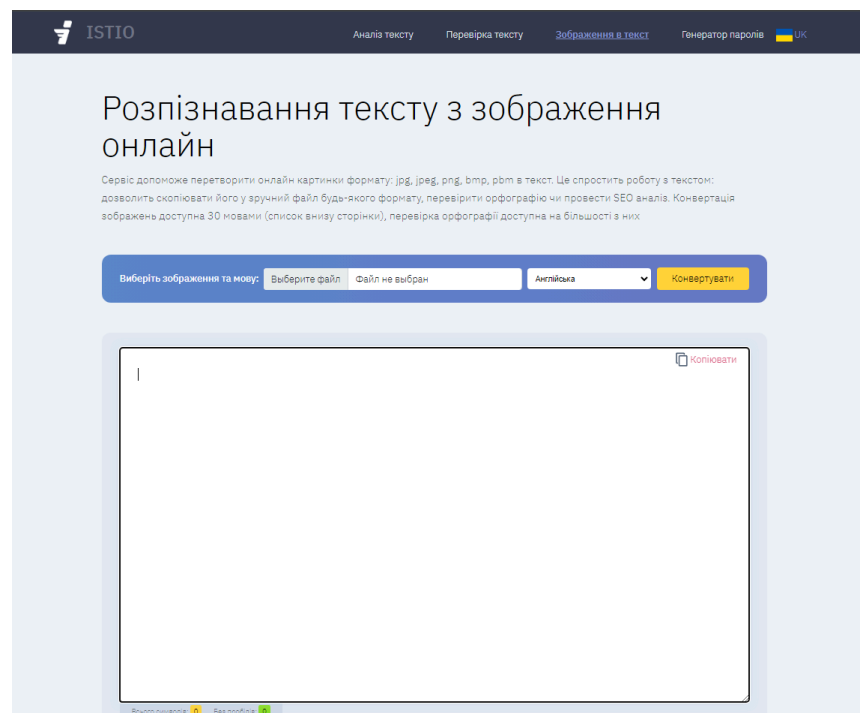


Рисунок 1.2 – Приклад вікна вводу на сайті

Додаток розуміє навіть найскладніший сценарій і пропонує прогноз тексту у верхній частині вікна, щоб можна було налаштувати інтерпретацію - або виправити почерк безпосередньо під час введення.

Додаток надсилає інформацію на свої сервери для розшифрування почерку, щоб покращити систему розпізнавання тексту, і ви не можете відмовитися від цього.

1.3.2 MyScript Nebo

Sider.AI - це сучасний сервіс, який надає різноманітні інструменти для роботи з текстом і даними. Незалежно від того, що ви пишете, Sider.AI може підвищити вашу продуктивність до максимуму. Завдяки своїм алгоритмам Sider.AI може автоматично витягувати інформацію з таких джерел, як зображення, рукописні нотатки та текстові файли. Просто треба завантажити зображення або файл в додаток Sider.AI швидко перетворить його на текст. Приклад вікна введення показано на рисунку 1.3.

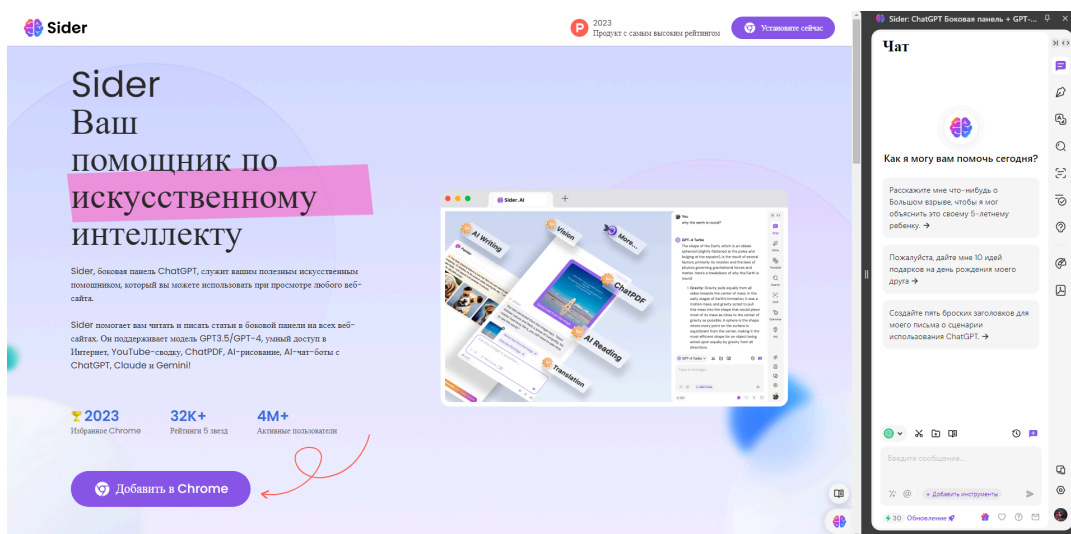


Рисунок 1.3 – Приклад вікна вводу в сайті

Крім того, Sider.AI має багато версій штучних інтелектів таких як gpt, claude, gemini, Llama, які можуть допомогти вам аналізувати, узагальнювати, пояснювати або перекладати текст для ваших проєктів.

Ви можете працювати з Sider.AI не через веб-інтерфейс, а через розширення для браузера, що дає змогу отримати доступ до всіх його функцій під час роботи в Інтернеті. Це дозволяє швидко аналізувати та обробляти інформацію без перемикання вкладок.

1.3.3 Img2txt.com

Img2txt.com - це інноваційний онлайн-інструмент, який дозволяє легко і швидко витягувати текстові файли з будь-яких зображень. Незалежно від того, чи це зображення з документа, скан або будь-яке інше зображення з текстом, Img2txt.com забезпечить точне і чітке вилучення тексту.

Для того щоб скористатися їм: завантажте зображення з текстом і надішліть його на обробку. Спеціальний алгоритм розпізнає текст на зображенні і перетворює його в редагований текстовий формат. Це дозволить легко редагувати зображення, копіювати та вставити отриманий текст у зручний для користувача спосіб. Приклад вікна введення показано на рисунку 1.4.

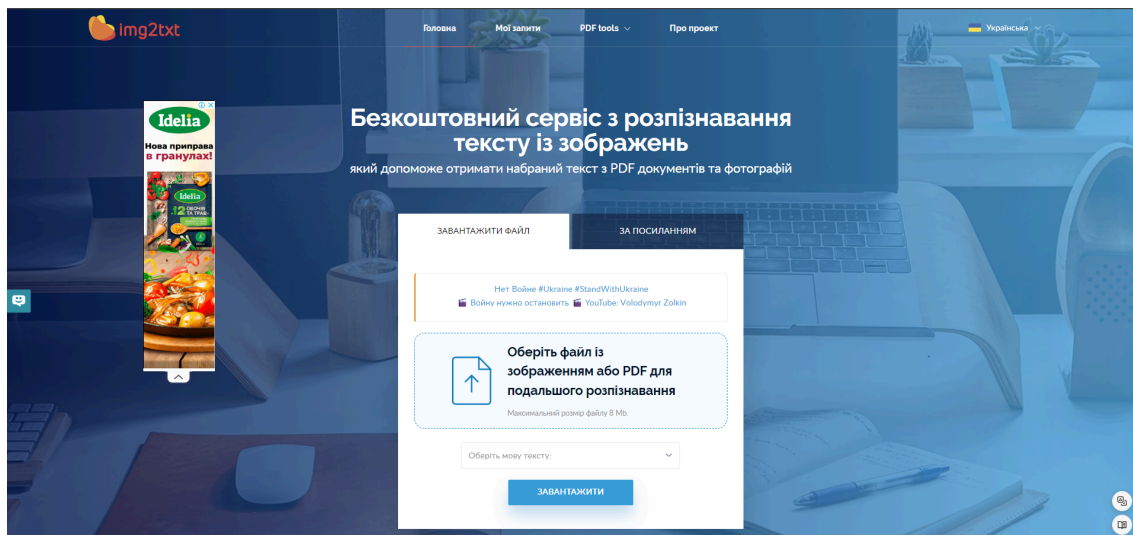


Рисунок 1.4 – Приклад вікна вводу в сайті

Однією з головних переваг цього сайту є його універсальність. Сервіс підтримує розпізнавання тексту з зображень найпопулярнішими мовами, що дозволяє користувачам з різних країн використовувати його відповідно до своїх потреб. Крім того, його можна використовувати на будь-якому пристрої з доступом до Інтернету, без необхідності встановлювати будь-які додатки або програмне забезпечення. За допомогою цього сайту у вас завжди буде простий і ефективний спосіб отримати текст із зображень, що робить його ідеальним інструментом для студентів, викладачів, і всіх, кому потрібно отримати текст із зображень якнайшвидше.

1.3.4 Висновок до розділу 1

1. Розглянуто різні реалізації алгоритмів розпізнавання рукописного тексту, а також розглянули існуючі додатки для цієї мети на платформах Android, iOS та Windows. Встановлено, що кожна з них має свої переваги та недоліки, і аналіз цих аспектів дозволяє визначити оптимальні вимоги до функціональності, щоб

задовольнити основні потреби користувача та забезпечити максимальне задоволення при використанні додатку.

2. Аналіз існуючих реалізацій алгоритмів розпізнавання рукописного тексту привів до визначення необхідних методів та інструментів для їх реалізації. Зокрема, найбільш ефективним методом було визнано символічне розпізнавання тексту з використанням штучних нейронних мереж.

3. Сформульовано функціональні вимоги до системи, а саме
Розпізнавання рукописних англійських символів на зображенні.
Сповіщення користувача про помилки у форматі вхідних даних.
Забезпечення якості обслуговування користувачів.
Висока точність розпізнавання.

4. Для реалізації користувацького інтерфейсу доцільно використати відкритий API під назвою Telegram Bot. Це було зумовлено зростаючою популярністю додатку для обміну повідомленнями серед користувачів та простотою використання бота, який не потребує встановлення додаткових додатків.

2. ПРОЕКТУВАННЯ СИСТЕМИ

2.1. Огляд предметної області

Дивно, що багато людей легко впізнають ряд чисел на рисунку 2.1 як 504192. Однак це враження спростовується при ближчому розгляді. Завдяки сотням мільйонів нейронів і десяткам мільярдів зв'язків наш мозок є найскладнішим процесором візуальної інформації з усіх існуючих, але більша частина цього процесу відбувається без нашого усвідомлення. Тож навіть якщо ми добре розуміємо те, що бачимо, ми не завжди усвідомлюємо, наскільки складним є цей процес.



Рисунок 2.1 – Рукописна послідовність чисел

До завдання розпізнавання рукописного тексту нейронні мережі підходять по-різному. Основна ідея полягає у використанні великої кількості рукописних символів як навчальних прикладів для розробки системи, яка може навчатися на цих прикладах. Тобто нейронна мережа використовує ці приклади для автоматичної генерації правил розпізнавання рукописного тексту. Збільшуючи кількість навчальних прикладів, мережа може дізнатися більше про рукописний текст і поліпшити точність розпізнавання.

Для цього завдання часто використовують згорткову нейронну мережу (CNN), яка спеціально розроблена для розпізнавання та обробки зображень, що використовуються в розпізнаванні рукописного тексту. Нейрони в CNN влаштовані так, щоб більше нагадувати нейрони, відповідальні за обробку зорових стимулів у людському мозку, що дозволяє уникнути проблем часткової обробки зображень, характерних для звичайних нейронних мереж.

CNN застосовує систему, подібну до багаторівневого персептрона, але з оптимізованою структурою, яка зменшує вимоги до обробки. Ця система складається з вхідних, вихідних і прихованих шарів, включаючи шари згортки, монтажні шари, шари повного з'єднання і шари нормалізації. Результатом є більш

ефективна і легка в навчанні система, призначена виключно для розпізнавання зображень, з меншою кількістю обмежень і вищою ефективністю обробки зображень.

2.2 Постановка задач і цілей

Основними вимогами до цього додатку є наступні:

- Розроблена система повинна бути здатна розпізнавати рукописні англійські символи на зображенні.

- Система повинна виводити користувачеві повідомлення про помилку, якщо введене зображення не відповідає заданому формату.

- Система повинна забезпечувати якісне обслуговування користувачів.

- Система повинна гарантувати точність розпізнавання.

Основними функціональними можливостями цього додатку є

- надсилання зображень на розпізнавання через чат з ботами в додатку Telegram.

- отримання результату розпізнавання в повідомленні від бота в додатку Telegram.

- розпізнати рукописний текст на зображенні.

- проаналізували зображення, щоб надати інформацію про розпізнавання тексту.

2.3 Опис функціональності системи

Система повинна дозволяти користувачам виконувати наступні операції:

- надати зображення почерку для розпізнавання системою, надіславши його боту приватним повідомленням.

- отримувати результати розпізнавання почерку від бота у приватному повідомленні.

- Надати зображення почерку для розширеного аналізу розпізнавання.

- Отримувати результати аналізу почерку від бота у приватному повідомленні.

2.4 Опис нейронної мережі

Типова система розпізнавання рукописного тексту складається з декількох ключових етапів: Попередня обробка зображення, сегментація, виділення ознак, класифікація та постобробка. Загальна базова схема показана на Рисунку 2.2.



Рисунок 2.2 – Загальна схема розв’язання для вирішення задачі розпізнавання.

1. Отримання зображення

Під час отримання зображення система розпізнавання отримує відскановане зображення як вхідне зображення. Зображення має бути у певному форматі, наприклад, JPEG, BMP, PNG тощо. Це зображення можна отримати за допомогою

сканера, цифрової камери або будь-якого іншого відповідного пристрою цифрового введення

2. Попередня обробка

Попередня обробка передбачає низку операцій, що застосовуються до відсканованого вхідного зображення для його суттєвого покращення та підготовки до подальшої сегментації. Послідовність цих операцій показано на рисунку 2.2. Сегментація виконується шляхом перетворення чорно-білого зображення в бінарне за допомогою методу глобального порогу. Краї бінарного зображення розширюються за допомогою методу Собела, і останніми кроками є розширення та заповнення проміжків, щоб отримати попередньо оброблене зображення, готове до подальшої сегментації.

3. Сегментація

На стадії сегментації зображення послідовності символів розбивається на підсторінки з окремими символами. У пропонованій системі попередньо оброблене вхідне фото сегментується на індивідуальні знаки, і кожному знаку надається номер за допомогою процесу маркування. Це позначення надає відомості про число символів у зображенні.

4. Отримання ознак

Після сегментації кожного символу на зображенні виділяються його особливості. Ці ознаки можуть базуватися на різних аспектах, таких як форма і текстура символу. Форму символу можна описати за допомогою геометричних параметрів, таких як площа, периметр і співвідношення сторін. Текстуру символу, в свою чергу, можна охарактеризувати статистичними показниками, такими як середня інтенсивність пікселів та дисперсія пікселів.

5. Класифікація та розпізнавання

На етапі класифікації, який є частиною процесу прийняття рішень системою розпізнавання, використовується згорткова нейронна мережа для класифікації та розпізнавання рукописних символів. На вхід мережі подається сегментоване зображення, отримане на попередньому етапі. Вихідний шар має 26 нейронів, оскільки система призначена для розпізнавання англійського алфавіту.

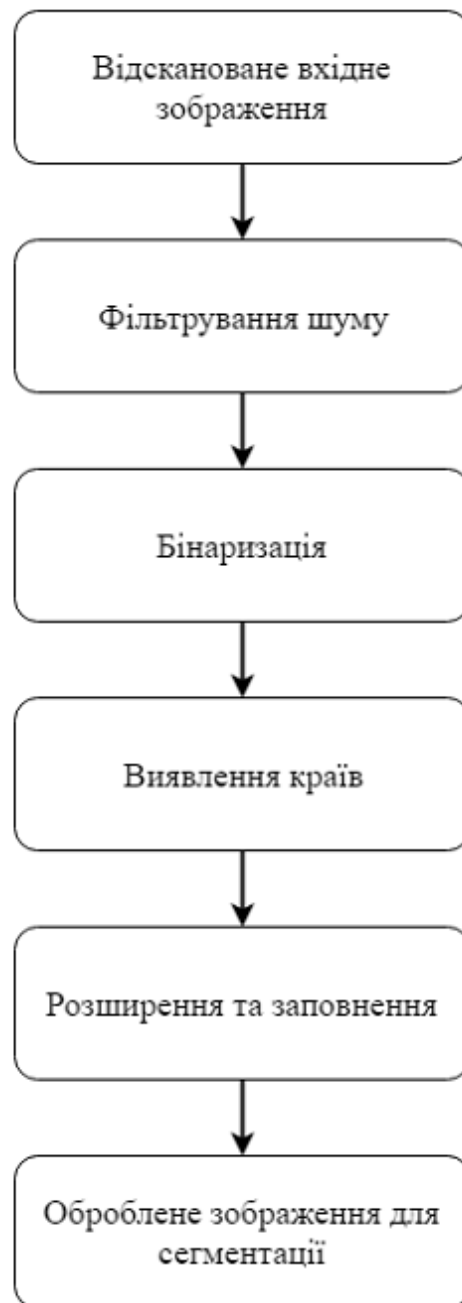


Рисунок 2.3 – Алгоритм передпроцесінгу зображення.

6. Пост-обробка

Етап постобробки є останнім етапом пропонованої системи ідентифікації. На цьому етапі всі розпізнані символи витягуються у вигляді структурованого тексту шляхом розрахунку еквівалентних ASCII-значень з урахуванням індексу розпізнавання тестових прикладів. Алгоритм передпроцесінгу показано на рисунку 2.3.

2.4.1 Спосіб первинної обробки

Сирі дані проходять кілька етапів попередньої обробки перед тим, як їх буде використано на описових етапах аналізу символів. Метою попередньої обробки є отримання даних, які можуть бути легко використані для систем візуального розпізнавання символів. Основні завдання попередньої обробки полягають у наступному:

- Фільтрація шуму. Коли ви скануєте документ, відскановані зображення можуть бути забруднені адитивним шумом, і ці неякісні зображення вплинуть на наступний етап обробки документа. Тому, щоб поліпшити якість зображень, потрібно виконати попередню обробку, перш ніж передавати їх на наступний етап обробки документа. Шум може призвести до виникнення віддалених фіктивних сегментів, значних пропусків між строчками тощо, і дуже важливо прибрати всі ці похибки, щоб зробити отриману інформацію більш якісною. Існує багато типів шуму на зображеннях. Адитивний шум - це чорно-білі точки, розкидані по всьому зображенню, які можна знайти майже в усіх документах. Методи зниження шуму можна розділити на дві основні групи, а саме

Фільтрація - має на меті видалити шум і зменшити кількість помилкових точок, зазвичай спричинених завидною поверхнею запису та/або низькою частотою дискретизації пристрою збору даних.

Морфологічні операції широко використовуються в обробці зображень для вилучення компонентів, важливих для представлення та опису форми області. Вони також ефективно застосовуються для видалення шуму на зображеннях документів, спричиненого поганою якістю паперу та чорнила або нерегулярними рухами рук.

- Бінарзація. Важливим кроком у розпізнаванні символів в автономному режимі є бінарзація чорно-білих зображень. Якщо процес бінарзації виконано правильно, це полегшує подальшу сегментацію і розпізнавання символів. Бінарзація - це перетворення чорно-білого зображення в двійкове.

- Виявлення країв. Виявлення країв є важливим етапом у обробці зображень, оскільки вони вказують на межі об'єктів і допомагають у їх сегментації, реєстрації

та ідентифікації. Цей процес значно зменшує обсяг даних і фільтрує надлишкову інформацію, зберігаючи важливі структурні особливості зображення. Існує кілька методів для виявлення країв, але вони можуть бути узагальнені до двох основних категорій: градієнтний метод і метод Лапласа. Градієнтний метод виявляє краї, шукаючи максимальні та мінімальні значення першої похідної зображення, тоді як метод Лапласа шукає нульові перетини другої похідної для визначення країв.

- Сканування та заливка. Виявлення спотворень - процес сканування може спричинити спотворення зображення. Існує кілька поширених методів виявлення викривлення сторінок; деякі з них базуються на виявленні з'єднаних компонентів і знаходженні середніх кутів між їхніми центрами. Асиметрію слід усунути, оскільки вона знижує точність документа. Обчислюється кут нахилу, і викривлені лінії вирівнюються по горизонталі за допомогою кута нахилу.

- Оцінка та нормалізація градієнта. Рукописний текст зазвичай характерний градієнтом символів. Символи бувають зміщені з правої на ліву сторону або навпаки. Крім того, різні нахили можуть виникати не лише в межах тексту, а й в рамках одного й того ж слова. Корекція нахилу не змінює порядок слів, і отримані слова виглядають природно. Градієнтна нормалізація застосовується для приведення всіх символів до стандартного вигляду. Найпоширенішим способом визначення градієнта є розрахунок середнього кута нахилу сусідніх вертикальних елементів.

2.4.2 Сегментація

У методах розпізнавання символів сегментація є найважливішим процесом. Сегментація виконується для того, щоб розділити зображення на окремі символи. Сегментувати рукописне слово на різні зони (верхню, середню і нижню) і символи складніше, ніж у друкованому документі. Це пов'язано, головним чином, з мінливістю інтервалів між символами, нахилу, розміру, тобто почерку. Іноді компоненти двох послідовних символів можуть торкатися або накладатися один на одного, і така ситуація ускладнює завдання сегментації. В індійських мовах такий контакт або накладання часто відбувається через варіації символів у верхній і нижній частинах. Сегментація є важливим етапом, оскільки ступінь розділення

слів, рядків або символів безпосередньо впливає на швидкість розпізнавання тексту.

Існує два типи сегментації:

Зовнішня сегментація: розділяє макет сторінки на логічні одиниці. Зовнішня сегментація - це відокремлення різних блоків тексту, таких як абзаци, речення або слова. Це найважливіша частина аналізу документа. Метою аналізу та розпізнавання документів є автоматичне вилучення інформації з паперу, а його первинною метою є розуміння документа людиною. Сегментація сторінок є одним з найважливіших кроків в аналізі презентації і є особливо складною, коли йдеться про складні презентації. Аналіз презентації відбувається у два етапи: перший - структурний аналіз, який передбачає сегментацію зображення на блоки елементів документа (абзац, рядок, слово тощо); другий - функціональний аналіз, який вказує на функціональний зміст елементів документа за допомогою позиції, розміру та різних правил розмітки. Потім сторінка сегментується шляхом пошуку структурованих областей на чорно-білих або кольорових зображеннях.

Внутрішня сегментація - це операція, яка намагається розбити зображення послідовності символів на підмножини окремих символів. Хоча ці методи значно вдосконалилися за останнє десятиліття і з'явилися різні техніки, сегментація рукописних символів залишається невирішеною проблемою.

2.5 Передумови

Після того як ми визначили функціональність системи та нейронної мережі, було створено діаграму сценаріїв використання для візуалізації та представлення сценаріїв використання системи. Детальна схема ієрархії варіантів використання показана на рисунках 2.4- 2.6.



Рисунок 2.4 - Структурна схема пріоритетності для користувача



Рисунок 2.5 - Структура варіантів використання для телеграм-бота

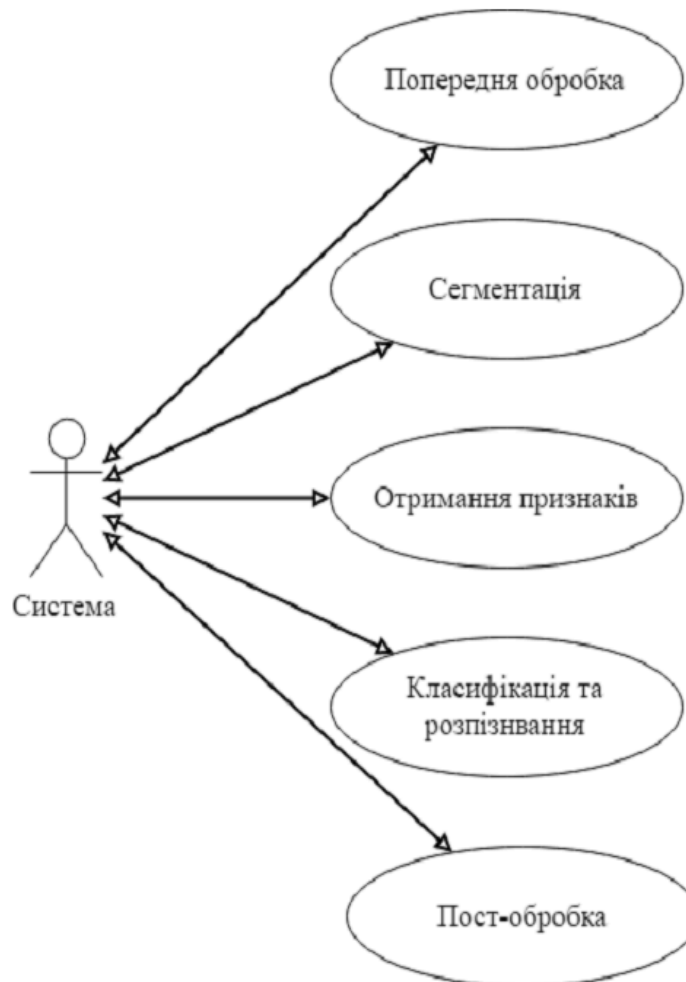


Рисунок 2.6 - Структура варіантів використання для моделі системи

2.6 Діаграми послідовності

У розділі 2.6 обговорюються діаграми послідовності, які є графічним представленням сценаріїв взаємодії між об'єктами в часовій послідовності. Вони показують, які дії відбуваються першими, а які - пізніше. Ці діаграми допомагають визначити ролі об'єктів і надають інформацію для визначення обов'язків класів та інтерфейсів. Діаграми послідовності показані на рисунках 2.7-2.11.



Рисунок 2.7 - Схема процесу взаємодії користувача з системою

На рисунку 2.7 показана схема взаємодії між користувачем, телеграм-роботом і системою в процесі розпізнавання зображень. Користувач надсилає зображення через Telegram-бота, який надсилає його системі для обробки. На схемі показано послідовність дій та обміну повідомленнями між трьома учасниками процесу. Після того, як користувач надсилає зображення, Telegram-робот інформує користувача про те, що розпізнавання розпочато. Після цього телеграм-робот надсилає зображення в систему для обробки. Система обробляє зображення, розпізнає текст і надсилає результат назад Telegram-боту. Нарешті, Telegram-робот надсилає користувачеві повідомлення з результатами розпізнавання тексту.

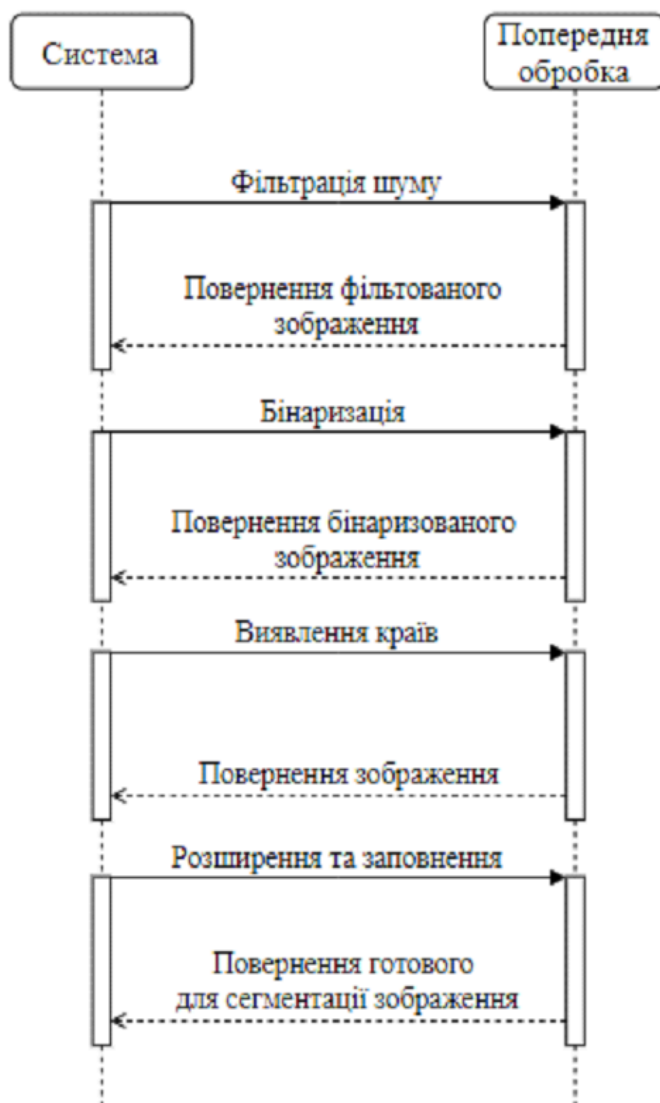


Рисунок 2.8 - Схема процесу попередньої обробки зображень

На рисунку 2.8 показано процес попередньої обробки зображення, який складається з декількох кроків. Спочатку виконується фільтрація шуму для видалення небажаних шумів, що покращує якість зображення. Далі виконується бінаризація для перетворення зображення у двоколірний формат для полегшення подальшої обробки. Наступний крок - виявлення країв для пошуку контурів об'єктів на зображенні. Останні кроки включають розширення і заливку для посилення контурів і усунення прогалин, після чого оброблене зображення готове до сегментації.



Рисунок 2.9 - Схема процесу сегментації зображення

Рисунок 2.9 ілюструє процес сегментації зображення. Після попередньої обробки оброблене зображення передається на етап сегментації, де воно ділиться на окремі частини або об'єкти. Це дозволяє виділити необхідні елементи для подальшого аналізу. Результат сегментації повертається в систему, де може бути використаний для подальших розрахунків або обробки.

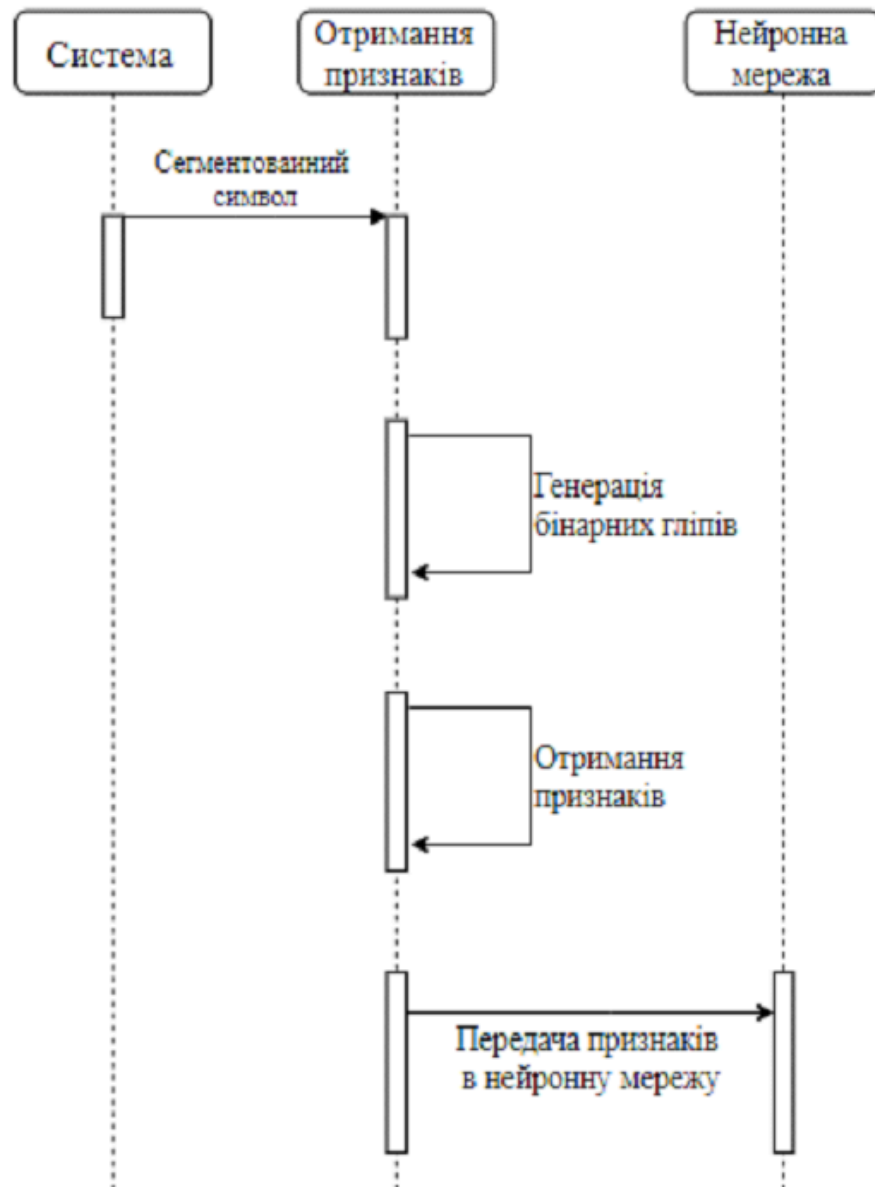


Рисунок 2.10 - Схема процесу виділення ознак на зображенні

На рисунку 2.10 показано процес вилучення ознак з сегментованих символів. Спочатку сегментовані символи передаються для створення бінарних гліфів, які представляють символи у вигляді двійкових кодів. Потім з цих гліфів виділяються ознаки, які є характеристиками, що описують символи. Виділені ознаки надсилаються до нейронної мережі для подальшого навчання.



Рисунок 2.11 - Схема процесу роботи нейронної мережі

На рисунку 2.11 показано процес навчання та тестування нейронної мережі. Спочатку ознаки, отримані на попередньому кроці, передаються нейронній мережі разом з інформацією про кількість ітерацій, необхідних для навчання. Потім нейронна мережа навчається, що завершує етап навчання. Потім нейронна мережа тестується для перевірки якості навчання. Нарешті, система отримує розпізнаний символ, який є результатом роботи нейронної мережі.

2.7 Висновки до розділу 2

1. У цьому розділі було окреслено обсяг проблеми, сформульовано основні вимоги та функціональні можливості системи розпізнавання рукописного тексту.
2. Розроблено робочу модель нейронної мережі.

3. Визначено і основні етапи розпізнавання тексту, а саме:

- попередня обробка зображення
- сегментація
- виділення ознак
- класифікація та розпізнавання

Визначено принципи роботи та функціонал кожного етапу.

4. Визначено основні пріоритетні кейси та розроблені діаграми варіантів використання.

5. На основі створених схем, функцій та вимог до системи були розроблені діаграми послідовності для основних функцій системи розпізнавання, що відображають взаємодію між об'єктами в часовій відповідності до їхньої послідовності.

3 ОПИС РОЗРОБЛЕНОЇ СИСТЕМИ

3.1 Вибір технологій та їх обґрунтування

3.1.1 Вибір платформи для системи

Розроблена система розпізнавання рукописного тексту має на меті надати користувачам зручний і доступний інструмент для розпізнавання тексту з різних джерел, таких як фотографії або відскановані документи. Однією з ключових вимог до системи є її універсальність та доступність для широкого кола користувачів, включаючи як власників мобільних пристроїв, так і користувачів ПК. У цьому розділі буде розглянуто вибір платформи для системи та обґрунтування вибору використовуваних технологій.

Однією з ключових вимог до платформи розгортання системи є її універсальність та доступність для користувачів на різних типах пристроїв. Враховуючи, що кількість користувачів мобільних та ПК постійно зростає, важливо обрати платформу, яка зможе задовольнити потреби обох категорій користувачів.

При виборі платформи для системи були враховані наступні критерії:

- Доступність: Платформа має бути доступною для користувачів з різних країн та регіонів світу.
- Гнучкість: Платформа повинна підтримувати різні типи пристроїв, такі як мобільні телефони, планшети та персональні комп'ютери.
- Простота використання: Платформа повинна мати дружній та інтуїтивно зрозумілий інтерфейс для користувачів.

3.1.2 Обґрунтування вибору месенджера Telegram

Проаналізувавши різні платформи, ми вирішили обрати Telegram як основну платформу для розміщення системи розпізнавання текстів внаслідок наступних особливостей:

- Широка популярність: Telegram - один з найпопулярніших месенджерів у світі, який налічує мільйони користувачів у різних країнах. Це забезпечує широке охоплення аудиторії для нашої системи.
- Кросплатформеність: Telegram підтримує різні операційні системи,

включаючи Android, iOS, Windows, macOS та Linux. Це означає, що користувачі зможуть використовувати нашу систему на будь-якому пристрої, який підтримує Telegram.

- Відкритий API для розробників: Telegram надає повністю відкритий API для створення чат-ботів. Це дозволяє розробникам легко і швидко інтегрувати свої системи з месенджером Telegram і створювати чат-ботів з різними функціями.

- Безпека та конфіденційність: Telegram відомий своїм високим рівнем безпеки та конфіденційності даних користувачів. Він використовує надійне шифрування та інші заходи безпеки для захисту особистої інформації користувачів.

Тому, обравши Telegram як платформу для нашої системи розпізнавання, ми гарантуємо її доступність, гнучкість і зручність для широкого кола користувачів.

Наша система розпізнавання рукописного тексту базується на платформі Telegram Bot API та бібліотеці Google Cloud Vision API для Python. Обґрунтуванням вибору цих технологій є їхня ефективність, простота використання та широкі функціональні можливості.

Telegram Bot API: Ця платформа була обрана для взаємодії з користувачами через месенджер Telegram з наступних причин:

- Популярність: Telegram - один з найпопулярніших месенджерів, що забезпечує широке охоплення аудиторії для нашої системи.

- Багатоплатформеність.

- Легка інтеграція.

Google Cloud Vision API: Ця бібліотека Python обрана для розпізнавання тексту на зображеннях з таких причин:

- Продуктивність: Google Cloud Vision API забезпечує швидке та ефективне розпізнавання тексту, надаючи користувачам швидку відповідь.

- Мовна підтримка.

- Простота використання.

В цілому, вибір API Telegram Bot і бібліотеки Google Cloud Vision API дозволяє нам створити ефективну і зручну систему розпізнавання рукописного тексту, яка відповідає потребам наших користувачів.

3.1.3 Пошук мови для програмування

Сьогодні існує багато мов програмування, які дозволяють створювати та навчати нейронні мережі. Але оскільки вам потрібно розробити дві взаємозалежні частини системи - нейронну мережу та користувацький інтерфейс у вигляді Telegram-бота - потрібно обрати мову програмування, яка зможе реалізувати і те, і інше.

Python був обраний як найбільш підходящий вибір для побудови цієї системи з кількох причин:

- Простота вивчення мови та її користування: Python відомий своєю простотою вивчення та зрозумілим синтаксисом, що робить його ідеальним вибором для швидкого розвитку та розробки.

- Широкий вибір бібліотек для роботи з нейронними мережами: Python має величезну кількість різноманітних бібліотек, таких як TensorFlow, Keras, PyTorch та інші, які дозволяють легко створювати, навчати та використовувати нейронні мережі.

- Підтримка API Telegram Bot: В Python є вже готова бібліотека для роботи з Telegram ботом, що спрощує процес створення та управління чат-ботом в Telegram.

- Популярність і спільнота розробників: Python є однією з найпопулярніших мов програмування, що забезпечує наявність великої та активної спільноти розробників, яка може надати допомогу та підтримку.

- Кросплатформенність: Python працює на різних операційних системах, що дозволяє розгорнути систему на будь-якому пристрої без проблем з сумісністю.

Щодо мов програмування Java і C#, обидві є потужними інструментами для розробки програмного забезпечення, проте для нашої конкретної системи розпізнавання рукописного тексту вони не є найкращими варіантами. Нижче наведено додаткові аргументи, які виправдовують відмову від цих мов:

Java:

Вимоги до пам'яті: Java споживає значно більше пам'яті порівняно з Python. У випадку обробки великих обсягів даних, це може призвести до неефективності роботи системи.

Проблема множинних наслідувань: В Java неможливо використовувати множинне наслідування, що може ускладнити розробку інтерфейсу користувача, особливо якщо потрібно реалізувати різні функції та можливості.

Швидкодія: В порівнянні з Python, Java може бути менш продуктивною через необхідність інтерпретування програмного коду.

C#:

Обмежена кросплатформеність: Хоча C# є потужною мовою програмування, вона переважно використовується для розробки програмного забезпечення під платформу Windows, що може обмежити нашу систему щодо розгортання на різних операційних системах.

Специфічний екосистеми: C# має свою власну екосистему, яка не така широка, як у Python, зокрема, що стосується бібліотек для роботи з нейронними мережами та API для розробки чат-ботів для Telegram.

Обмежена спільнота розробників: Хоча спільнота розробників C# є активною, вона не така широка, як у Python, що може ускладнити пошук рішень на форумах та знаходження допомоги в разі виникнення проблем.

Отже, на мою думку Python є найкращим варіантом для нас, вона є ефективною, простою у використанні та забезпечує всі необхідні інструменти для реалізації нашої системи розпізнавання рукописного тексту.

3.1.4 Вибір допоміжних бібліотек

На додаток до бібліотеки Google Cloud Vision API для реалізації системи розпізнавання рукописного тексту було обрано додаткові допоміжні бібліотеки для забезпечення більшої ефективності та функціональності. Ці бібліотеки описані нижче:

1. OpenCV

OpenCV (Open Source Computer Vision Library) - бібліотека з відкритим вихідним кодом для комп'ютерного та машинного зору. Вона містить великий набір алгоритмів для обробки зображень і відео, включаючи обробку зображень, відстеження об'єктів, розпізнавання облич, калібрування камер і багато іншого. OpenCV можна використовувати для попередньої обробки зображень перед подальшою обробкою за допомогою Google Cloud Vision API, що може підвищити точність розпізнавання тексту.

2. NumPy

NumPy - це базовий пакет для наукових обчислень на мові Python. Він підтримує масиви і матриці та дозволяє проводити швидкі обчислення з числовими даними. NumPy можна використовувати для обробки та перетворення зображень перед передачею їх до бібліотеки Google Cloud Vision API, що допомагає оптимізувати продуктивність системи.

3. Matplotlib

Matplotlib - це графічна бібліотека Python. Вона дозволяє створювати широкий спектр графіків, включаючи гістограми, лінійні діаграми та кругові діаграми, які можуть бути корисними для візуалізації результатів розпізнавання зображень.

Використання цих допоміжних бібліотек доповнює можливості бібліотеки Google Cloud Vision API і забезпечує ширшу функціональність та ефективність системи розпізнавання рукописного тексту.

3.2 Структурні модулі розробленої системи включають наступні компоненти:

1. Модуль попередньої обробки зображень.
2. Модуль сегментації зображень.
3. Модуль виділення ознак на сегментованому зображенні.
4. Модуль розробки моделі навчання нейронної мережі.
5. модуль для тестування нейронної мережі.
6. Модуль для реалізації Telegram-робота для взаємодії з користувачем.

3.3. Нейромережева модель

Нейромережева модель системи розпізнавання рукописного тексту містить згорткову нейронну мережу (CNN), рекурентну нейронну мережу (RNN) і, нарешті, шари класифікатора часових зв'язків (TLC). Схему нейромережевої моделі показано на рисунку 3.1.

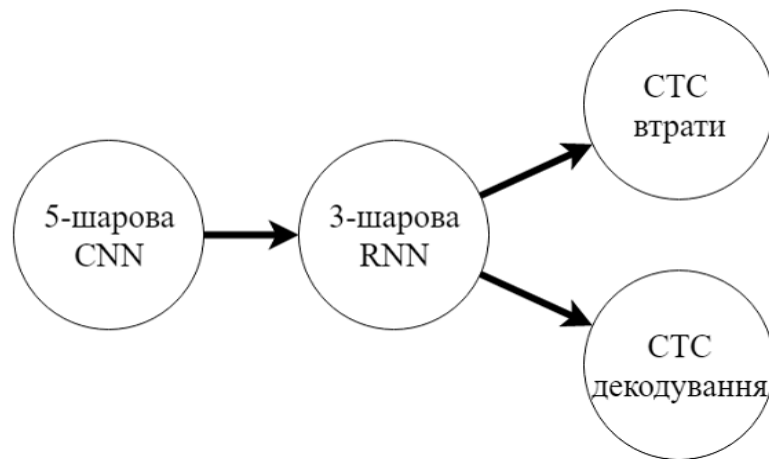


Рисунок 3.1 - Структурна схема нейромережевої моделі для розпізнавання тексту

Такий підхід дозволяє розпізнавати текст на рівні символів, враховуючи можливість розпізнавання слів або фраз, які можуть бути відсутніми в навчальних даних.

CNN (Convolutional Neural Network):

Вхідне зображення пропускається через шари згорткової нейронної мережі для виділення ознак. Кожен шар включає операції згортки, за якими слідує нелінійна функція активації ReLU та операція згортки. Ці шари навчені виокремлювати важливі характеристики зображень. Результатом є послідовність ознак довжиною 32 і розміром 32 x 256. (рисунок 3.2)

Recurrent Neural Network (RNN):

RNN передає інформацію через послідовність ознак, яка містить 256 ознак на кожному кроці. Вона використовує модифікацію RNN з довготривалою пам'яттю для кращого зберігання та передачі інформації на великі відстані. Вихідна послідовність являє собою матрицю 32 x 80.

СТС (Classifier of Temporal Connection):

Класифікатор часових зв'язків використовується для обчислення втрат під час навчання і декодування отриманої матриці ШНМ в остаточний текст під час розпізнавання. Довжина правильного і розпізнаного тексту не може бути більше 32 символів.

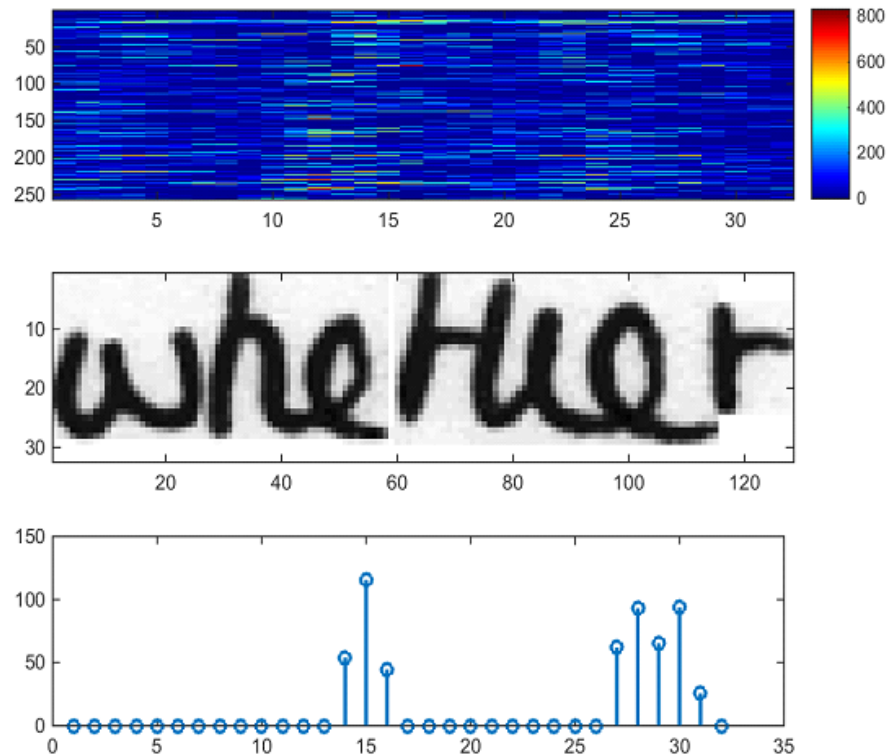


Рисунок 3.2 - Отриманий результат після CNN шару

На рисунку 3.3 зображено вихідну матрицю ШНМ для зображення з текстом «little». Матриця містить оцінки символів, включаючи порожній тег CTC як останній вхідний символ. Символи показано в порядку зверху вниз.

“!”#&’()*+,-./0123456789:;?ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz”

Більшість символів передбачаються точно в тому місці, де вони з'являються на зображенні. Наприклад, можна порівняти положення "i" на зображенні і на графіку. Лише останній символ "e" не вирівняний, але це нормально, оскільки операція CTC не враховує абсолютні позиції.

На нижньому графіку показані бали для символів "l", "i", "t", "e" і порожньої мітки CTC. Текст може бути легко декодований, приймаючи найбільш вірогідний

символ з кожного такту, що утворює так званий кращий шлях. Потім видаляються повторювані символи і всі пропуски, щоб отримати кінцевий текст "little".

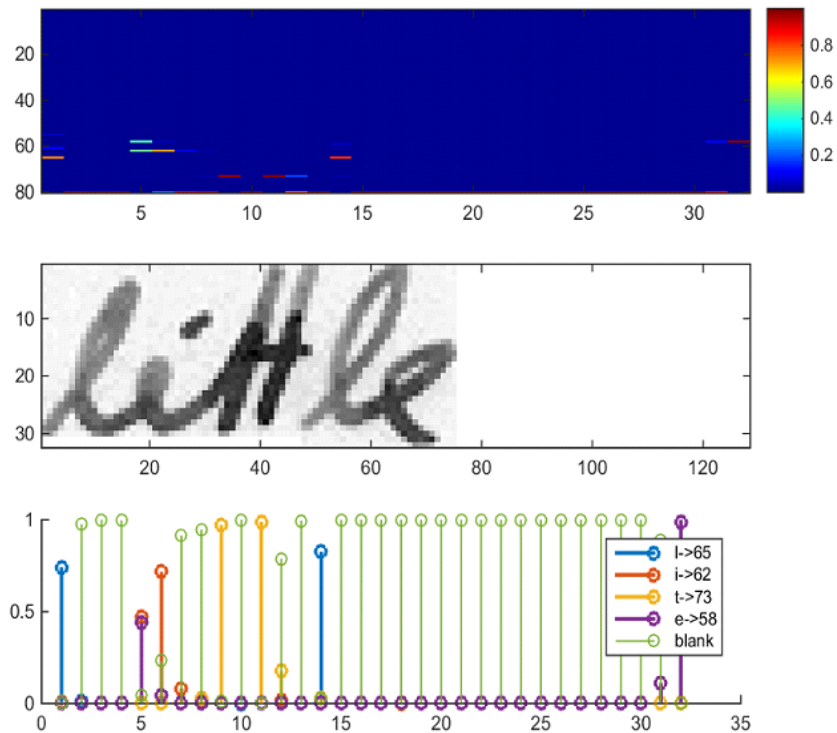


Рисунок 3.3 - Результати ідентифікації після RNN-шару

3.4 Опис графічного інтерфейсу користувача системи

Графічний інтерфейс для цієї системи розпізнавання має бути якомога поблажливим і зрозумілим у використанні, тому для реалізації нашого графічного інтерфейсу ми будемо використовувати функціонал python скіпта, а саме TeleBot API. Адже в сучасних умовах майже всі люди використовують смартфони або планшети, усі вони використовують месенджери для комунікації, обміну фото та багато іншого. Тому ми будемо використовувати лідера серед месенджерів, а саме Telegram. Основний плюс цього месенджера є те що в ньому дуже легко зробити свого бота і відкритий API. На рисунках 3.4 – 3.6 зображено вигляд графічного інтерфейсу користувача для роботи з системою розпізнавання рукописних текстів.

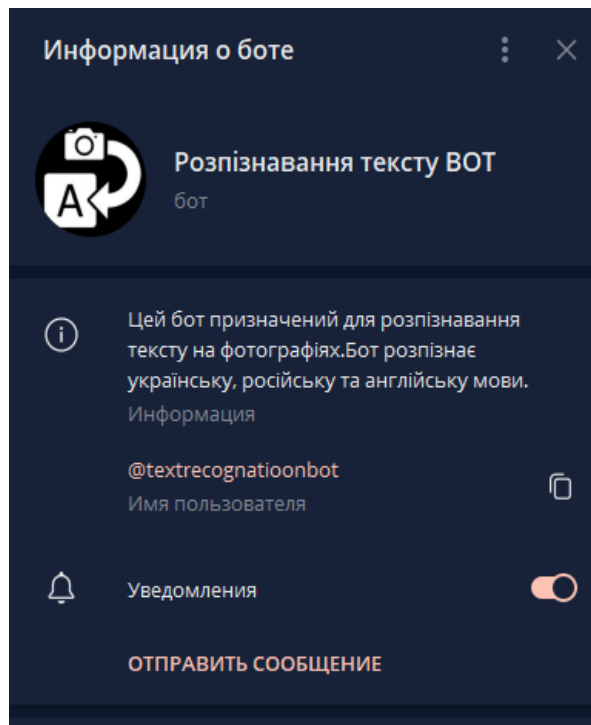


Рисунок 3.4 - Розташування даних про бота

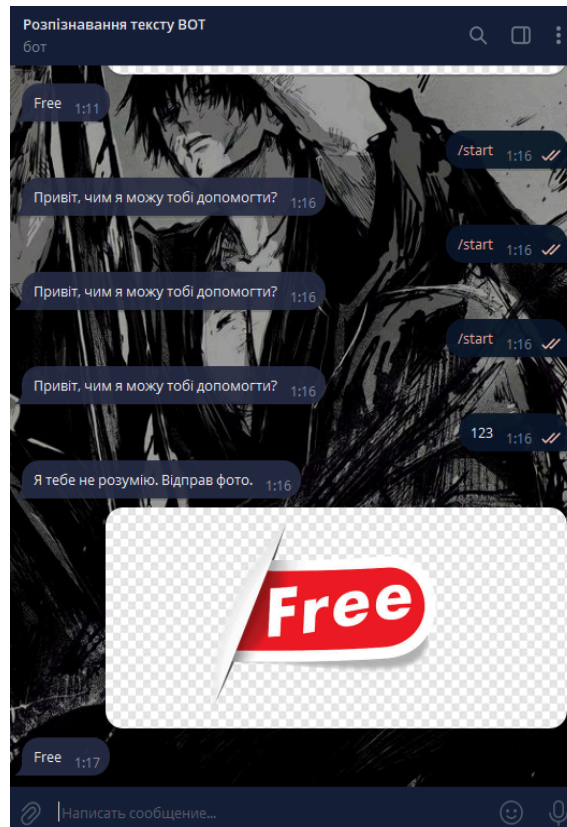


Рисунок 3.5 - Чат з ботом

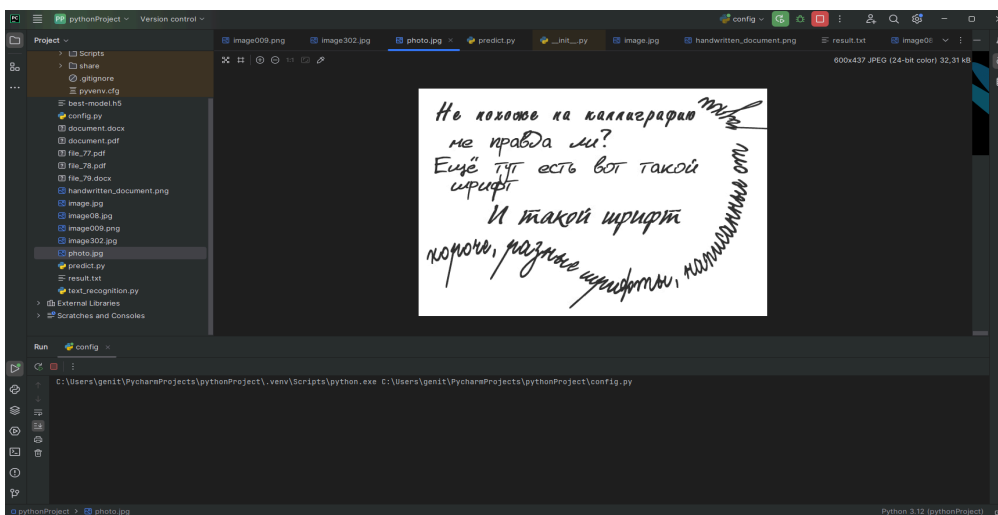


Рисунок 3.6 - Зберігання фото для аналізу тексту

3.5 Висновки до розділу 3

У розділі 3 ми розглянули процес розробки системи розпізнавання рукописного тексту.

1. Для взаємодії з користувачем ми обрали платформу Telegram Bot API, що надає зручний інтерфейс через месенджер Telegram, та бібліотеку Google Cloud Vision API для Python для розпізнавання тексту на зображеннях.

2. Обрали мову програмування Python через її високу продуктивність, багаті бібліотеки для роботи з даними та легкість у вивченні.

3. Для реалізації моделі нейронної мережі використали бібліотеки для глибокого навчання, такі як TensorFlow або PyTorch.

4. У результаті вибраних технологій та методів розробили ефективну систему для розпізнавання рукописного тексту, придатну для різних сфер застосування

ВИСНОВКИ

1. Метою кваліфікаційної роботи була розробка та аналіз системи розпізнавання рукописного документів з використанням алгоритмів глибокого навчання.

2. Однією з основних цілей було створення програмного забезпечення для автоматичного розпізнавання та інтерпретації рукописного тексту. Під час роботи над проектом досліджено актуальність проблеми та визначені вимоги до системи, яка б відповідала потребам користувачів.

3. Ключовим компонентом нашої системи є нейронна мережа, що відповідає за аналіз і розпізнавання рукописного тексту. Важливим моментом був вибір оптимальної архітектури мережі, яка гарантує високу точність розпізнавання при мінімальному споживанні ресурсів. Було розглянуто кілька підходів до архітектури нейронної мережі та проведено експерименти з різними архітектурами для визначення найефективнішого варіанту.

4. Розглянуто можливість використання Telegram-бота в якості користувацького інтерфейсу для нашої системи. Такий підхід забезпечив зручну та ефективну взаємодію з додатком, дозволяючи користувачам завантажувати зображення з рукописним текстом та отримувати результати розпізнавання безпосередньо в месенджері.

5. Враховуючи вимоги до системи та результати наших досліджень, було обрано мову програмування Python для реалізації проекту. Python відомий своєю простотою та широким спектром доступних бібліотек для роботи з нейронними мережами, що робить його ідеальним вибором для нашої задачі.

6. В результаті нашої роботи створено систему, яка дозволяє ефективно і з високою точністю розпізнавати рукописний текст. Наша система може бути використана в широкому спектрі галузей, включаючи освіту, документообіг та інші сфери, де важлива швидкість і точність обробки текстової інформації.

ПЕРЕЛІК ПОСИЛАНЬ

1. Вац І. Офлайн розпізнавання рукописних англійських числівників з використанням кореляційного методу / І. Вац, С. Сінгх. // Міжнародний журнал інженерних досліджень та технологій. - 2014. - №6. - С. 1626-1629.
2. Сінгх Г. Розпізнавання рукописних символів хінді з використанням нейронної мережі зворотного поширення / Г. Сінгх, С. Лері. // Міжнародний журнал комп'ютерних наук та інформаційних технологій. - 2012. - №4. - С. 4892-4895.
3. Розпізнавання символів Devnagiri з використанням нейронних мереж / [S. S. Sayuad, A. Jadhav, M. Jadhav та ін.]. // Міжнародний журнал передових технологій та інженерії. - 2013. - №1. - С. 476-480.
4. Патель К. Офлайн розпізнавання рукописного тексту: A Review / К. Patel, М. Gandhi. // Міжнародний журнал науково-технічних досліджень. - 2016. - №5. - С. 193-196.
5. Система розпізнавання рукописних символів на основі нечіткої логіки / С. Борде, Е. Шах, П. Рават, В. Патіл. // Міжнародний журнал інженерних досліджень та застосувань. - 2012. - С. 151-154.
6. Донгаре С. А. Розпізнавання рукописних символів деванагарі з використанням нейронної мережі / С. А. Донгаре, Д. Б. Кширсагар, С. В. Вагчауре. // Журнал комп'ютерної інженерії IOSR. - 2014. - №14. - С. 74-79
7. Патіл М. Б. Розпізнавання рукописних символів деванагарі за допомогою сегментації та штучних нейронних мереж / М. Б. Патіл, В. Нараваде. // Міжнародний журнал інженерних досліджень та технологій. - 2012. - №1. - С. 1-15.
8. Каур М. Система розпізнавання рукописних символів гурмухі / М. Каур, С. Кумар. // Міжнародний журнал інженерних досліджень та технологій. - 2012. - №1. - С. 1-5.

9. Ногай М. Попередня обробка зображень для оптичного розпізнавання символів з використанням нейронних мереж / М. Ногай, Р. Джака. // Journal of Patter Recognition Research. - 2011. - С. 1-11.
10. Шарма Н. Розпізнавання рукописних англійських літер: A Review / N. Sharma, T. Patnaik, K. Bhupendra. // Міжнародний журнал передових технологій та інженерії. - 2013. - №2. - С. 318-320.

Додаток А

Код телеграм-бота

```
import os
import telebot
from google.cloud import vision

# Устанавливаем переменные окружения
os.environ["GOOGLE_APPLICATION_CREDENTIALS"] =
"C:/pelagic-diorama-422818-n2-50cfa92e10ef.json"

# Создаем клиент для Google Cloud Vision API
client = vision.ImageAnnotatorClient()

# Инициализируем бота
bot =
telebot.TeleBot(token='7149884705:AAF-yboq44Kj0Q8i761KuiLo0T4vApM5wm4')

# Функция для обработки изображений
def process_image(file_path):
    with open(file_path, 'rb') as image_file:
        content = image_file.read()

    # Передаем изображение на обработку в Google Cloud Vision API
    image = vision.Image(content=content)
    response = client.text_detection(image=image)
    texts = response.text_annotations

    # Извлекаем текст из первого элемента ответа или возвращаем сообщение
    "Текст не найден"
```

```
if texts:
    extracted_text = texts[0].description.strip()
else:
    extracted_text = "Текст не знайдено"

return extracted_text

# Обработчик для команды /start
@bot.message_handler(commands=['start'])
def handle_start(message):
    bot.reply_to(message, "Привіт! Надішліть мені фотографію з рукописним текстом, і я спробую його розпізнати.")

# Обработчик для повідомлень, що не є фотографією
@bot.message_handler(func=lambda message: True)
def handle_text(message):
    bot.reply_to(message, "Надішліть, будь ласка, фотографію з рукописним текстом.")

# Обработчик для повідомлень з фотографіями
@bot.message_handler(content_types=['photo'])
def handle_photo(message):
    # Получаем информацию о фотографии
    file_id = message.photo[-1].file_id
    file_info = bot.get_file(file_id)
    file_path = file_info.file_path

    # Скачиваем фотографию
    downloaded_file = bot.download_file(file_path)
```

```
file_name = 'photo.jpg'

# Сохраняем фотографию на диск
with open(file_name, 'wb') as new_file:
    new_file.write(downloaded_file)

# Обрабатываем фотографию и отправляем результат пользователю
extracted_text = process_image(file_name)
bot.reply_to(message, f'Розпізнаний текст:\n{extracted_text}')

# Запускаем бота
bot.polling()
```

ДЕМОНСТРАЦІЙНІ МАТЕРІАЛИ

ДЕРЖАВНИЙ УНІВЕРСИТЕТ ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
КАФЕДРА ШТУЧНОГО ІНТЕЛЕКТУ

КВАЛІФІКАЦІЙНА РОБОТА на тему: «Розробка інформаційної системи розпізнавання рукописних документів з використанням алгоритмів глибокого навчання»

Виконав: здобувач вищої освіти гр.
ШПД-41 Дмитро Радковський
Керівник: Доктор технічних наук,
професор Євген ЧИЧКАРЬОВ

2024 рік

Мета роботи: удосконалення методів оптичного розпізнавання символів для ефективного перетворення тексту з зображень у редагований формат з високою точністю та швидкістю.

Об'єкт дослідження: процес розпізнавання тексту на зображеннях.

Предметом дослідження: вивчення різних методів, методик та алгоритмів розпізнавання тексту на зображеннях.

Основні завдання кваліфікаційної роботи:

- Розробка алгоритмів розпізнавання рукописних документів, включаючи оптимізацію нейронних мереж або інших моделей машинного навчання
- Збір та підготовка наборів даних, включаючи відбір та обробку достатньої кількості рукописних даних для навчання моделей
- Впровадження інформаційної системи, включаючи розробку телеграм боту для взаємодії з навченою моделлю та створення користувацького інтерфейсу.
- Оцінка продуктивності системи, включаючи тестування точності, швидкості та надійності розпізнавання на реальних даних.

АНАЛІЗ ЗАПРОПОНОВАНОГО СПОСОБУ

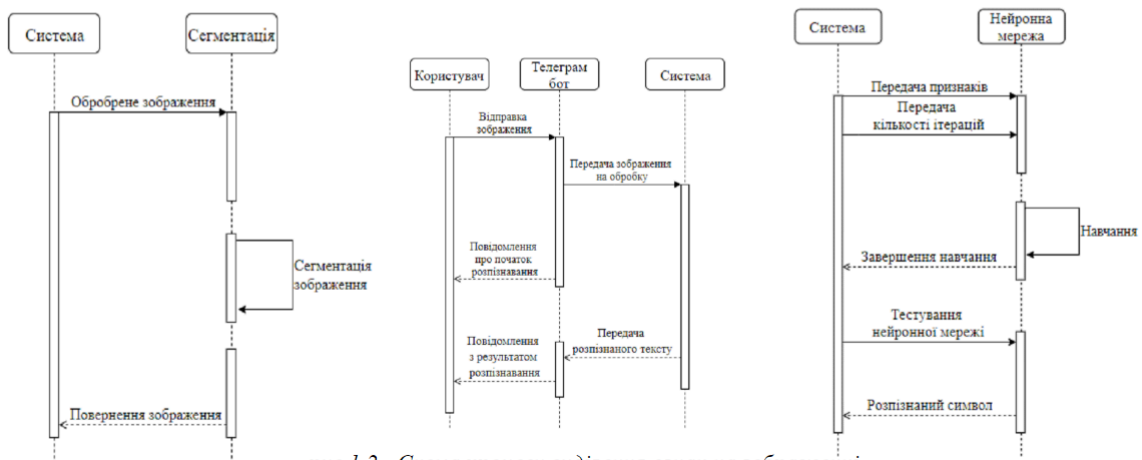


рис 1.1 - Алгоритм
передпроцесінгу зображення

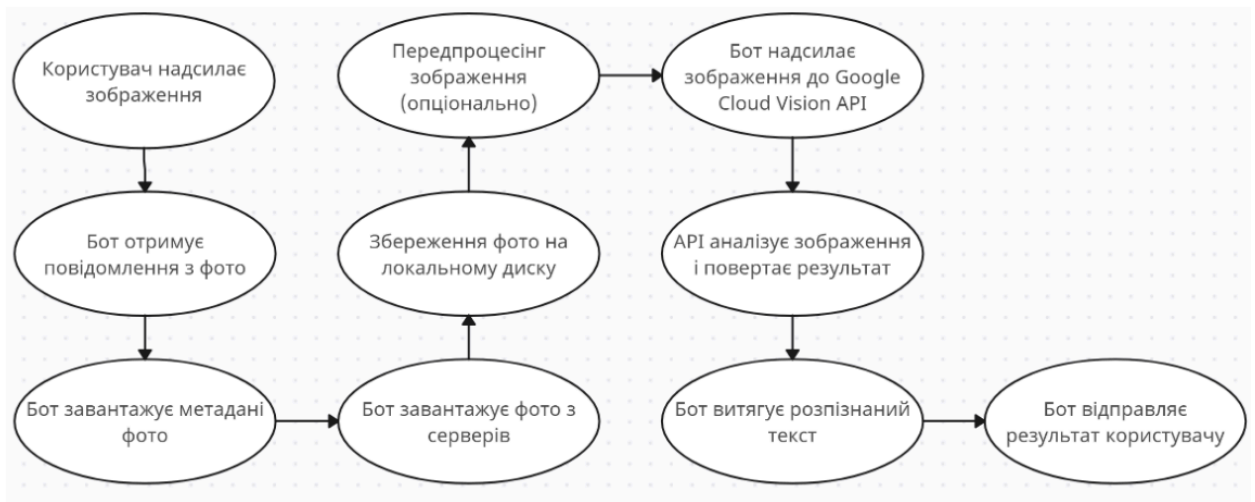
Огляд існуючих рішень

Назва CNN	Особливості	Застосування
VGG	Глибока архітектура з невеликою кількістю параметрів	Класифікація зображень, детекція об'єктів
ResNet	Використовує блоки зі з'єднаннями ("skip connections"), які полегшують тренування та запобігають виникненню градієнту	Класифікація виявлення об'єктів, семантична сегментація
Inception	Використовує "Inception" блоки, які дозволяють виявляти зразки різного розміру та орієнтації в одному шарі мережі	Класифікація виявлення об'єктів
MobileNet	Легкий архітектурний підхід, що використовує "depthwise separable convolution" для зменшення кількості параметрів та обчислень	Мобільні додатки вбудовані системи, малопотужні пристрої
DenseNet	Використовує густі з'єднання між шарами, що сприяє покращенню передачі градієнту та використанню обмежених ресурсів	Класифікація виявлення об'єктів, сегментація

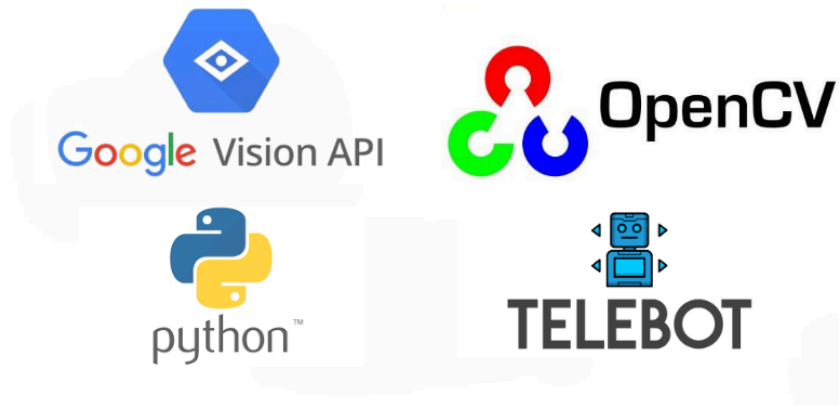
АНАЛІЗ ЗАПРОПОНОВАНОГО СПОСОБУ



АНАЛІЗ ЗАПРОПОНОВАНОГО СПОСОБУ



ВИКОРИСТАНІ ТЕХНОЛОГІЇ



ДЕМОНСТРАЦІЯ РОЗРОБЛЕНОГО ТЕЛЕГРАМ БОТУ

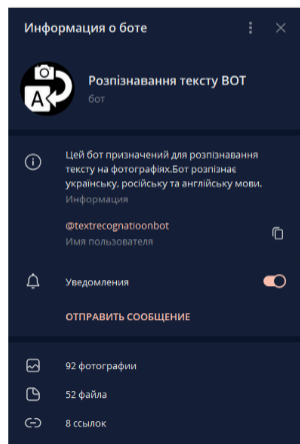


рис 1.4 - Сторінка бота

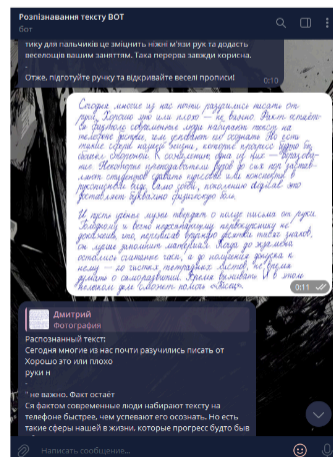


рис 1.5 - Працюючий бот

Висновки

У цій класифікаційній роботі ми розробили телеграм-бота, а також ознайомилися з бібліотекою Google vision api. Також ми дізналися про глибокі нейронні мережі.

Швидкість роботи розробленого бота становить близько 1,7 секунди на збереження, обробку фотографії та відправку фото. Системи розпізнавання зображень відіграють важливу роль у подальшому розвитку нейронних мереж. Крім того дана нейронна мережа постійно покращує функціональність, якість обробки, та підвищення ефективності обробки великих обсягів даних.

Дякую за увагу!