**УДК 004.032.2**

**В. І. Яркун**

Національний університет “Львівська політехніка”,

кафедра електронних обчислювальних машин

**ПРОГРАМНИЙ СЕРВІС РОЗПІЗНАВАННЯ РУКОПИСНОГО ТЕКСТУ ДЛЯ МОБІЛЬНОГО ОБЧИСЛЮВАЛЬНОГО ПРИСТРОЮ**

*© Яркун В. І., 2019*

**Розроблено метод розпізнавання тексту, рукописного чи друкованого, на зображенні. Метод базується на емпіричному опрацюванні зображень статистичних моделях машинного навчання та функціонування. Він забезпечує ефективне розв’язання задач двох класів: виявлення тексту на зображенні та розпізнавання тексту. Розроблені алгоритмічні підходи що об’єднують ці два класи задач.**

**Ключові слова: розпізнавання символів, зображення, згорткова нейронна мережа, машинне навчання.**

**V. Yarkun**

Lviv Polytechnic National University,

Computer Engineering Department

**SOFTWARE RECOGNITION SERVICE FOR MOBILE COMPUTING DEVICE**

*© Yarkun V., 2019*

**In this article, a system of handwritten or printed text recognition in the image has been developed. Empirical methods of image processing and statistical models of machine learning and simulation are being developed in two directions: the detection of text on the image and the recognition of the text.**

**Key words: character recognition, image, convolutional neural network, machine learning.**

**Вступ**

Дослідники, які займаються виявленням та розпізнаванням тексту на зображенні та відео намагаються розвинути комп’ютерну систему з можливістю автоматично зчитувати текстовий вміст із зображень чи відео з складним фоном. Загальна комп’ютерна система розпізнавання тексту повинна давати відповіді на два питання: «де знаходиться текст на зображенні?» і «що цей текст означає?». Інакше кажучи, використовуючи таку систему, текст повинен бути автоматично виявленим і кожен символ може бути розпізнаний.

Безупинне вивчення та вдосконалення виявлення і розпізнавання тексту мотивується сучасним станом та широкою областю застосування цифрового мультимедіа. Сьогодні все більше візуальної та аудіо інформації збирається, зберігається, доставляється та керується у цифрових формах. Використання пристроїв мультимедіа, а саме смартфонів, які є на даний момент широко розповсюдженні у повсякденному людському застосуванні, призводить до появи нових викликів та до керування великими мультимедійними базами даних.

**Окреслення проблеми**

Розробка методів машинного розпізнавання дозволяє розширити коло виконуваних комп'ютерами завдань і зробити машинну переробку інформації більш інтелектуальною. Прикладами сфер застосування розпізнавання можуть служити системи розпізнавання тексту, машинний зір, розпізнавання мови, відбитків пальців та інше. Незважаючи на те, що деякі з цих завдань вирішуються людиною на підсвідомому рівні з великою швидкістю, до теперішнього часу ще не створено комп'ютерних програм, які вирішують їх у настільки ж загальному вигляді. Існуючі системи призначені для роботи лише в спеціальних випадках зі строго обмеженою сферою застосування. Тому створення програмних засобів для розпізнавання рукописного тексту та дослідження даної області є актуальним.

**Аналіз останніх досліджень та публікацій**

Через те, що вартість обробки зображень чи відео досить висока та вимагає значних затрат зі сторони процесора, більшість мобільних додатків залежать від даних інтерпретованих у цифровому вигляді. Дані повинні бути проструктуровані щоб вони могли бути легко опрацьовані комп’ютером. Також важливо щоб дані були релевантні та доступні для пошуку, наприклад в текстовому форматі. Щоб автоматично чи інтерактивно отримувати такі дані, необхідно виконувати пошук по мультимедійних базах даних, використовуючи описові функції (особливості), які відносяться до об’єкту дослідження – фотографій, відео та т. п. [3, 4].

Особливості, такі як колір, текстура, форма, рух, макет, можуть бути отримані при обробці на низькому рівні, використовуючи технології комп’ютерного бачення. Різновиди особливостей які відносяться до кольору використовуються, як атомні блоки для пошуку зображень та відео, наприклад, домінуючий колір, регіональна чи глобальна гістограма, вектори усередненого кольору та власне зображення. Особливості текстури, такі як гістограми орієнтації краю, параметри випадкових полів Маркова, коефіцієнти перетворення вейвлет та Фур’є, локальні бінарні шаблони також широко використовуються для індексування природних образів.

Особливості форми в основному використовуються в задачах пошуку та моделювання об’єктів. Типовими особливостями форми є область накладання, параметри перетворення Хафа, елементарні дескриптори, вимірювання зміни кривизни простору, дескриптори Фур’є та кути країв. Особливості руху, при обробці на низькому рівні, включають величину руху та похідні. Особливостями макету можуть бути особливості просторових вимірювань в зображеннях або в кадрах відео [4 - 6].

Існує два види тексту на зображеннях, це: текст сцени та накладений текст. Текст сцени – це ті текстові рядки, які написані на деяких об’єктах у зображеннях. Вони, як правило, мають великі розміри, але можуть мати оклюзії, різні вирівнювання та рухи. Накладений текст немає оклюзій, як правило, є малого розміру та релевантний до відповідного зображення.

**Мета статті**

Метою цієї роботи є розробити та дослідити алгоритмічно-програмні засоби розпізнавання тексту, використовуючи згорткову нейронну мережу. Текст для розпізнавання може бути, як писаний, так і друкований. Розробити програмне забезпечення для реалізації та демонстрації даних засобів розпізнавання тексту.

**Основний матеріал дослідження**

Для побудови алгоритмічно-програмних засобів розпізнавання рукописного тексту, слід розділити роботу системи на чотири етапи.

Спочатку система повинна виконати задачу виявлення тексту на зображенні, де результатом можуть бути певні координати тексту чи символів на зображенні. Наступним етапом буде обробка зображення. На цій стадії роботи системи можна виконувати нарізку символів із зображення для подачі на розпізнавання. Останніми етапами є виділення ключових особливостей, які будуть використовуватись на момент розпізнавання символів, та власне розпізнавання символів [1-4].

На рис. 1 зображено основні засоби для побудови системи розпізнавання рукописного тексту, де на вході система буде отримувати певне зображення, а на виході демонструвати результат оцифрованого документу.

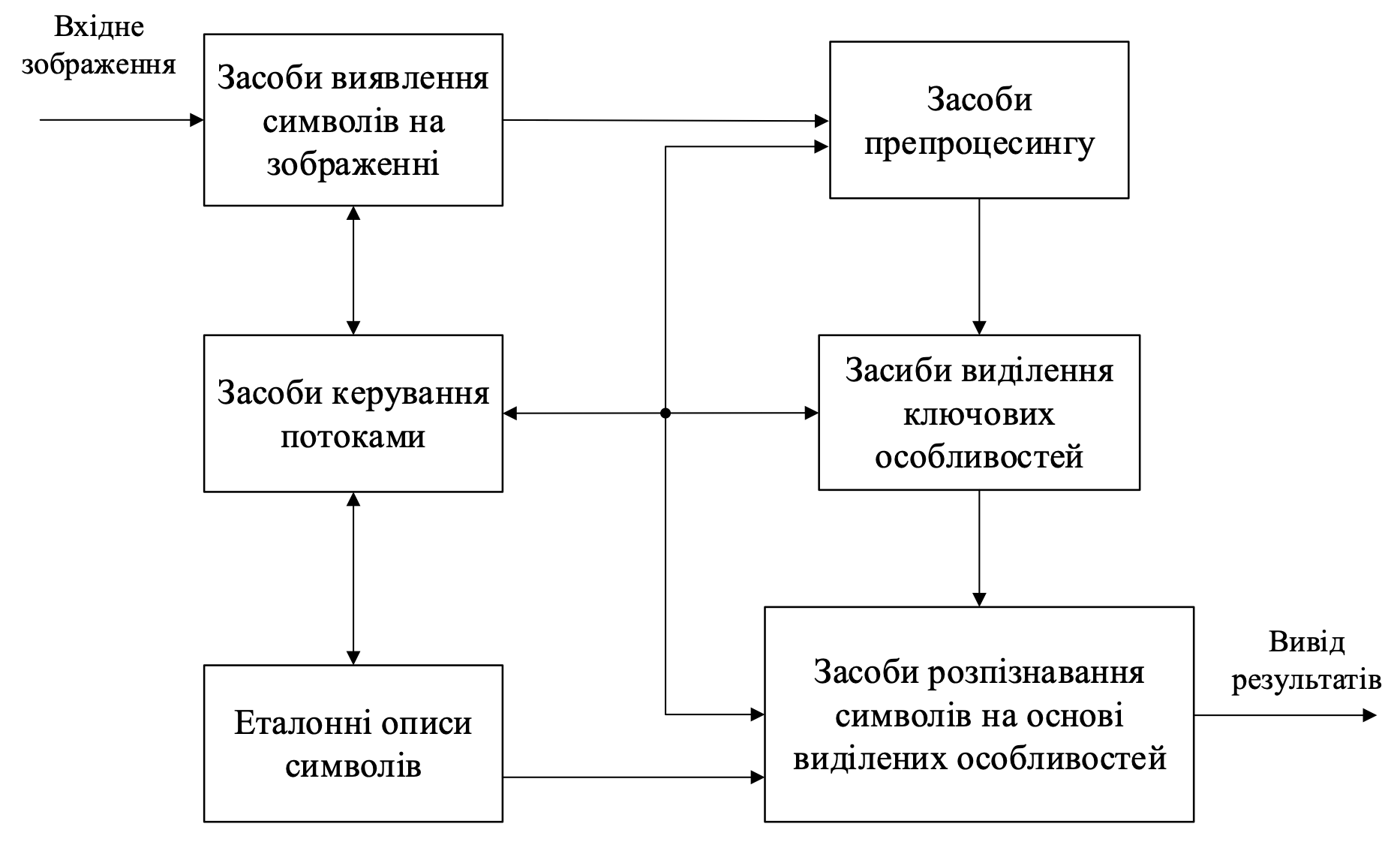


Рис. 1. Структура побудови алгоритмічно-програмних засобів розпізнавання рукописних текстів

Перш ніж почати розробку системи слід скласти план роботи та побудови системи, що в свою чергу допоможе зрозуміти суть поставленої задачі. Якщо в етапах розробки будуть присутні певні незалежні завдання, то їх можна легко побачити та розділити роботу, що в свою чергу пришвидшить реалізацію системи.

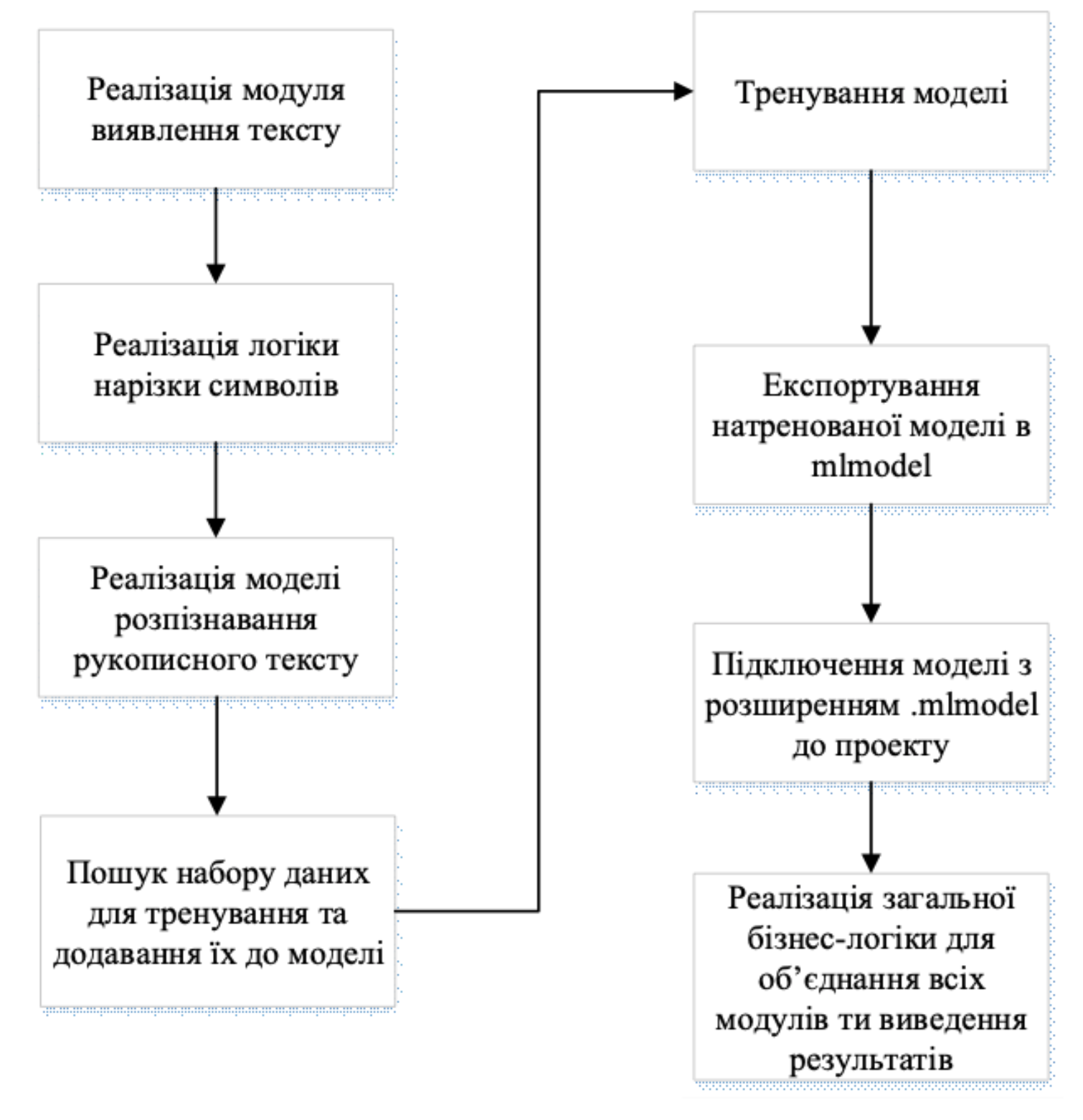


Рис. 2. Структурна схема етапів реалізації системи

На рис. 2 зображено загальну структурну схеми етапів розробки системи розпізнавання рукописного тексту. Спочатку слід реалізувати модуль виявлення тексту, використовуючи фреймворк Vision. Після цього відбувається реалізація логіки нарізки символів, використовуючи вихідні дані з Vision. Далі реалізація моделі розпізнавання рукописного тексту, яка розроблялась на мові Python використовуючи бібліотеку tensorflow. Також важливим етапом є пошук даних для тренування, бо у випадку їх відсутності система не буде працювати, неможливо натренувати нейронну мережу без прикладів. Далі відбувається безпосереднє тренування мережі та експортування натренованої моделі у формат .mlmodel. Після цього можна підключати дану модель до проекту та використовувати для розпізнавання нових вхідних символів. Останній, але не менш важливий етап – це реалізація загальної логіки програми, яка буде об’єднувати всі програмні модулі та виводити результат розпізнавання користувачу.

Для опису роботи системи створено UML діаграми, які демонструють можливі випадки та роботу головних процесів реалізованого програмного забезпечення. Діаграма прецедентів зображена на рис. 3.

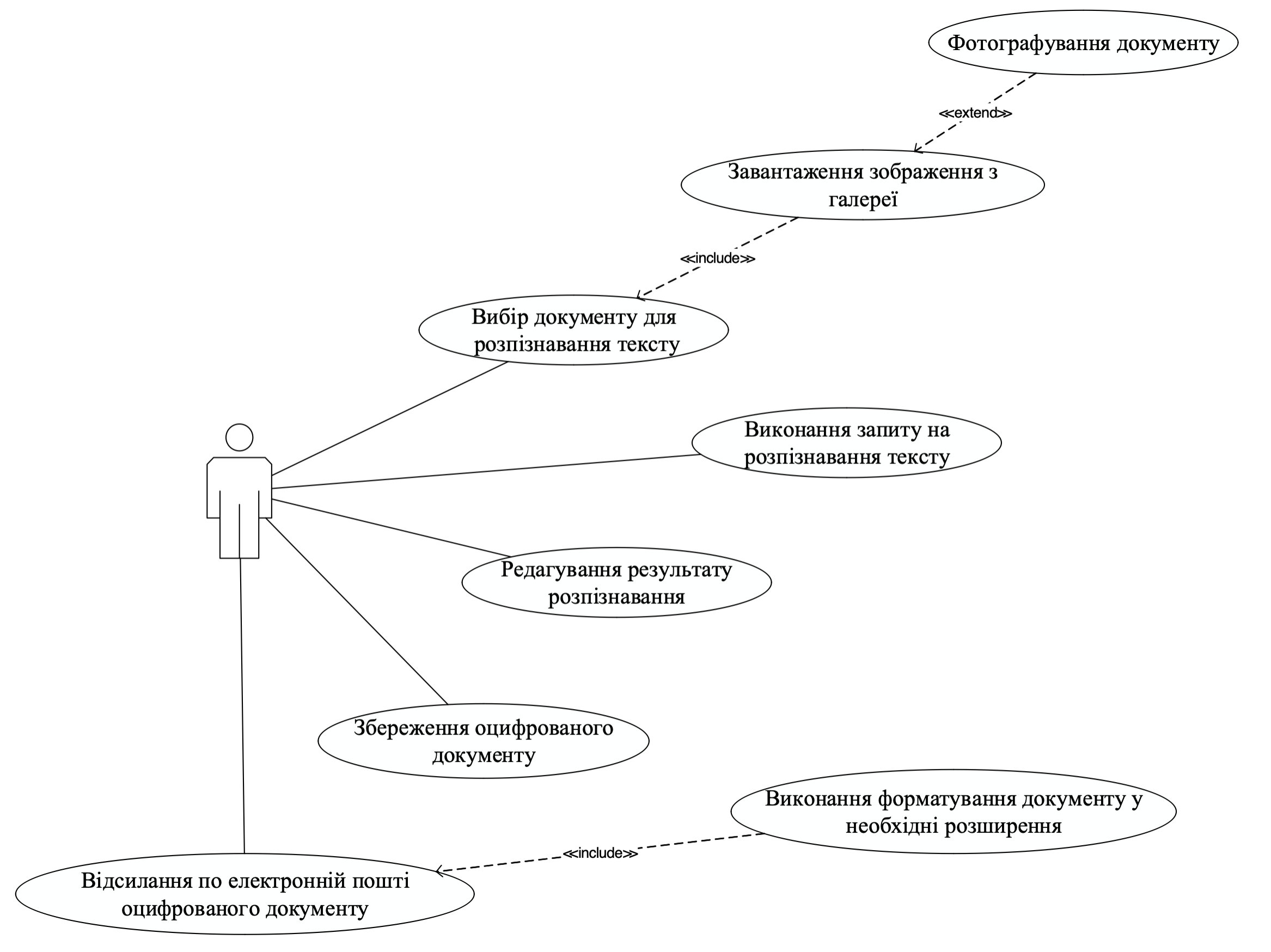


Рис. 3. Діаграма прецедентів

Користувач спочатку повинен вибрати зображення документу, якого потрібно розпізнати. Він це може зробити, завантаживши існуюче зображення з галереї або сфотографувати і тоді завантажити його.

Після цього користувач виконує запит на розпізнавання, натиснувши на кнопку початку обробки, таким чином розпочинаючи процес розпізнавання. Коли процес розпізнавання завершився, користувач буде бачити на екрані смартфону оцифрований текст документу. Цей текст відображатиметься в інтерактивному вікні, дозволяючи користувачу його редагувати при необхідності.

Після завершення процесів розпізнавання та редагування, користувач може зберегти оцифрований документ.

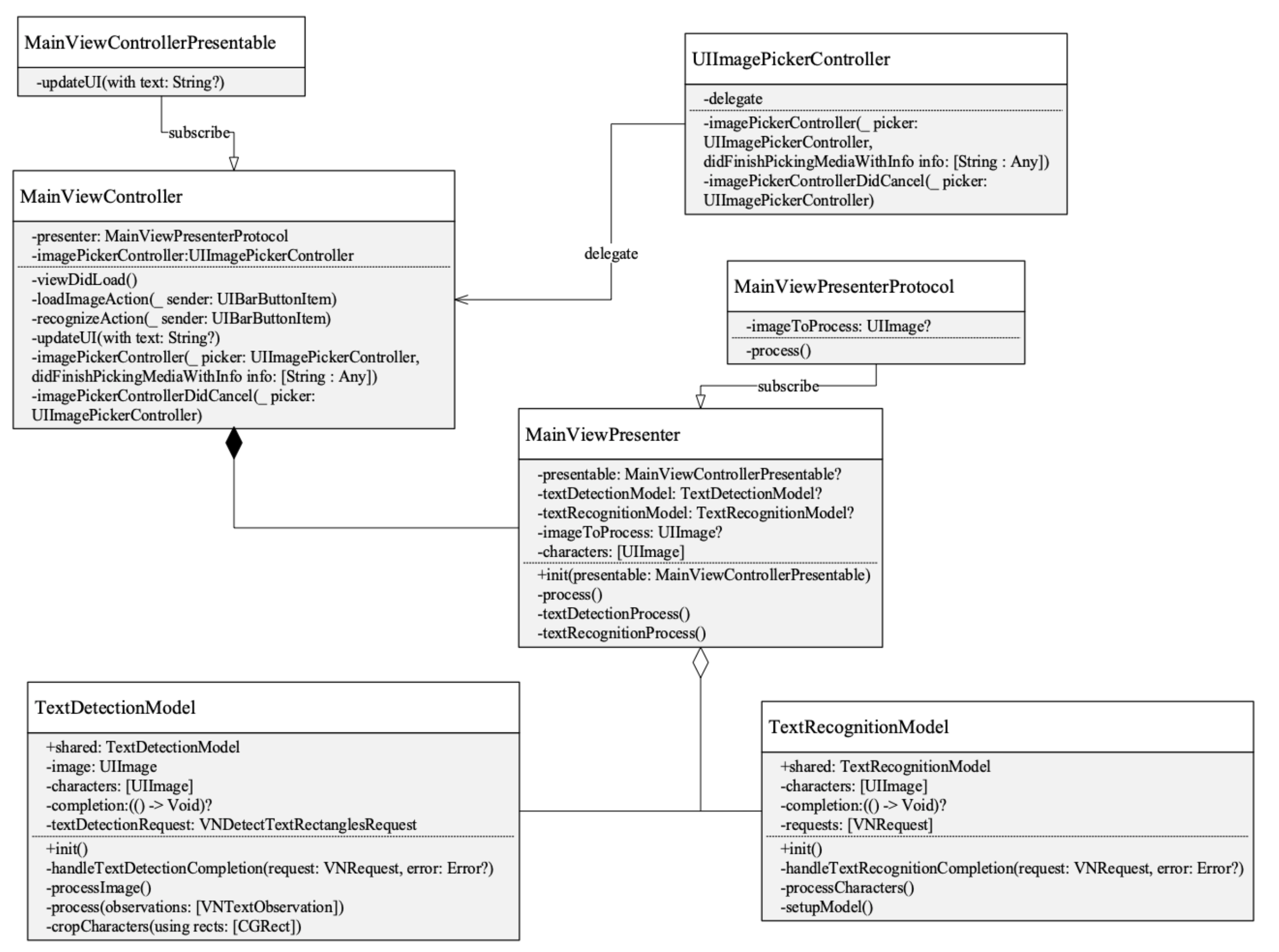


Рис. 4. Діаграми класів

На рис. 4 зображено діаграму класів, де можна побачити архітектуру програмного додатку та зв’язки між різними компонентами. Робота системи починається з класу MainViewController, який підписується на протокол MainViewControllerPresentable, через цей протокол MainViewPresenter зможе отримати доступ до певного функціоналу MainViewController. Також MainViewController є делегатом класу UIImagePickerController, який використовується для доступу до галереї та вибору зображення на розпізнавання. MainViewControllerє композитом для MainViewPresenter, якщо він знищиться і не буде існувати то так само і знищиться MainViewPresenter. MainViewPresenter підписується на протокол MainViewPresenterProtocol для того щоб MainViewController мав доступ до певного функціоналу MainViewPresenter. MainViewPresenter агрегує функціонал на виявлення та розпізнавання тексту у два модулі: TextDetectionModel та TextRecognitionModel. TextDetectionModel відповідає за виявлення тексту на зображенні, використовуючи фреймворк Vision, та нарізку символів. TextRecognitionModel виконує розпізнавання символів, використовуючи натреновану модель та повертає результати класифікації символів.

Архітектура зображена на рис. 5 називається MVP, де можна виділити клас Model, який реалізовує певну логіку роботи програми, View, яка відображає користувачу результати роботи класу Model, Presenter, який використовується як комунікатор між попередніми двома компонентами.

**Висновки**

Проведено аналіз існуючих проблем при створенні таких систем. Розроблено метод розпізнавання тексту, рукописного чи друкованого, на зображенні. Розроблено алгоритмічні підходи що об’єднують ці два класи задач. Для опису роботи системи створено UML діаграми, які демонструють можливі випадки та роботу головних процесів реалізованого програмного забезпечення.

Дані засоби розпізнавання використовуються локально на мобільному пристрої, що при цьому їх виділяє серед аналогів, оскільки цей підхід не вимагає підключення пристрою до мережі Інтернет та виконання задачі розпізнавання, використовуючи мережу. Це також збільшить швидкодію, тому що затримки, які виникають при виконанні запитів через мережу, будуть відсутні.

**Використані джерела**

*1. Yarkun V., Paramud Y. Algorithmic and software means of handwritten symbol recognition // Bulletin of the National University “Lviv Polytechnic” “Computer Science and Information Technologies” No 881, Lviv, Ukraine. 2017. - P. 98-107. 2. Yarkun V., Paramud Y. Algorithmic and software synchronization of information exchange // Bulletin of the National University “Lviv Polytechnic” “Computer Science and Information Technologies” No 857, Lviv, Ukraine. 2016. - P.111-118. 3. Datong Chen, Jean-Marc Odobez, Herve Bourlard. Text detection and recognition in images and video frames // Pattern Recognition journal, 2003. – P. 595-608. 4. Convolutional Neural Network [Electronic resource]/ Stanford. – Access mode: http://ufldl.stanford.edu/tutorial /supervised/ConvolutionalNeuralNetwork/. 5. Convolutional networks [Electronic resource]/ Stanford. – Access mode: http://cs231n.github.io/convolutional-networks/#overview. 6. Сoreml [Electronic resource]/ Apple. – Access mode: https://developer.apple.com/documentation/coreml. 7. Bachelor’s degree dissertation [Electronic resource]/ Github. – Access mode: https://github.com/VitaliyYarkun/Bachelor-s-degree-dissertation-*