

УДК 004.65

DOI: <https://doi.org/10.53920/ITS-2024-1-1>

Евгеній Олександрович ПРИЙМАК,

здобувач першого (бакалаврського) рівня вищої освіти,
Державний університет інформаційно-комунікаційних технологій

ORCID ID: [0009-0003-8341-8006](https://orcid.org/0009-0003-8341-8006)

Евген Олександрович ЗАЙЦЕВ,

доктор технічних наук, старший науковий співробітник,
Інститут електродинаміки Національної академії наук України

ORCID ID: [0000-0003-3303-471X](https://orcid.org/0000-0003-3303-471X)

Андрій Вікторович ЛЕМЕШКО,

доктор філософії з комп'ютерної інженерії, доцент,
Державний університет інформаційно-комунікаційних технологій

ORCID ID: [0000-0001-8003-3168](https://orcid.org/0000-0001-8003-3168)

Артем Васильович АНТОНЕНКО,

кандидат технічних наук, доцент,
Національний університет біоресурсів і природокористування України

ORCID ID: [0000-0001-9397-1209](https://orcid.org/0000-0001-9397-1209)

ДОСЛІДЖЕННЯ МОЖЛИВОСТЕЙ ОПТИМІЗАЦІЇ ПРОЦЕСУ ОБРОБКИ ДАНИХ В ДЕРЖАВНИХ ІНФОРМАЦІЙНИХ СИСТЕМАХ ІЗ ВИКОРИСТАННЯМ ШТУЧНОГО ІНТЕЛЕКТУ

У статті було досліджено можливості із інтеграції алгоритмів штучного інтелекту для покращення обробки інформації в державних базах даних. Показано, що алгоритм розпізнавання даних MRZ з зображення паспорта володіє значним потенціалом для використання в державних базах даних, адже він може суттєво покращити ефективність та безпеку різноманітних державних послуг. Обробка інформації в державних базах даних України є складним процесом, що охоплює збір, введення, перевірку, обробку та зберігання даних. Цей процес є важливим для ефективної роботи державних органів, але традиційні методи, засновані на людській праці, стають все менш ефективними через зростаючі обсяги і складність інформації. Вирішення цієї проблеми може бути досягнуто завдяки сучасним технологіям, таким як штучний інтелект (ШІ), який може автоматизувати рутинні процеси, підвищити точність і безпеку даних, а також зменшити витрати. У цій роботі розглянуто можливість

інтеграції алгоритмів штучного інтелекту (ШІ) для покращення обробки даних в державних базах. Було розроблено алгоритм для автоматичного вилучення інформації з машиночитабельної зони (MRZ) українського паспорта у формі ID-картки. Програмний код на Python обробляє зображення MRZ, визначає та вилучає ключові дані про власника. Функції включають попередню обробку зображення, виявлення області MRZ, аналіз розпізнаного тексту та форматування даних. Результати тестування показали, що алгоритм успішно обробляє дані з ID-картки, мінімізуючи втручання людини та скорочуючи час на обробку. Використання ШІ для автоматизації розпізнавання MRZ може стати основою для створення єдиної системи ідентифікації громадян, що покращить доступ до державних послуг, зменшить адміністративне навантаження та підвищить рівень безпеки. Таким чином, впровадження алгоритмів ШІ для обробки інформації в державних базах даних має значний потенціал для підвищення ефективності та якості державних послуг в Україні.

Ключові слова: інформаційні системи, штучний інтелект, обробка даних, алгоритм, MRZ.

Yevhenii PRYIMAK,

bachelor,

State University of Information and Communication Technologies

Ievgen ZAITSEV,

doctor of Technical Sciences, senior researcher,

Institute of Electrodynamics of the National Academy of Sciences of Ukraine

Andriy LEMESHKO,

doctor of philosophy in computer engineering, associate professor,

State University of Information and Communication Technologies

Artem ANTONENKO,

candidate of technical sciences, associate professor,

National University of Life and Environmental Sciences of Ukraine

STUDY OF THE OPPORTUNITIES OF OPTIMIZING THE DATA PROCESSING PROCESS IN STATE INFORMATION SYSTEMS USING ARTIFICIAL INTELLIGENCE

The article explored the possibilities of integrating artificial intelligence algorithms to improve information processing in state databases. It is shown that the MRZ data recognition algorithm from

the passport image has significant potential for use in government databases, as it can significantly improve the efficiency and security of various government services. The processing of information in the state databases of Ukraine is a complex process that includes the collection, input, verification, processing and storage of data. This process is essential for the efficient operation of government agencies, but traditional methods based on human labor are becoming less and less effective due to the growing volume and complexity of information. The solution to this problem can be achieved thanks to modern technologies such as artificial intelligence (AI), which can automate routine processes, improve data accuracy and security, and reduce costs. This paper considers the possibility of integrating artificial intelligence (AI) algorithms to improve data processing in state databases. An algorithm was developed for automatic extraction of information from the machine-readable zone (MRZ) of the Ukrainian passport in the form of an ID card. Python programming code processes MRZ images, identifies and extracts key owner data. Features include image preprocessing, MRZ region detection, recognized text analysis, and data formatting. The test results showed that the algorithm successfully processes data from the ID card, minimizing human intervention and reducing processing time. Using AI to automate MRZ recognition can be the basis for creating a unified citizen identification system that will improve access to public services, reduce administrative burden and increase security. Thus, the implementation of AI algorithms for processing information in state databases has significant potential for improving the efficiency and quality of public services in Ukraine.

Keywords: *information systems, artificial intelligence, data processing, algorithm, MRZ.*

Постановка проблеми. Для вирішення задач обробки даних відомо багато випадків використання ШІ для автоматизації рутинних процесів. Так компанія Amazon використовує чат-бот «Q» на базі ШІ, щоб відповідати на запитання клієнтів, надавати підтримку та створювати графіки. Бот успішно використовується самою компанією для обробки та налізу інформації, формування висновків з них та формувати звіти. Для особистого використання Amazon надає готову внутрішню базу даних, до якої ШІ звертається задля формування потрібних форматів. Популярна та поширена міжнародна платіжна система PayPal, використовує ШІ

для ведення бухгалтерського обліку, автоматичної обробки даних з рахунків, квитанцій та інших фінансових документів. Значним поширенням застосування ШІ є галузь медицини. Компанія Zebra Medical Vision використовує ШІ для аналізу рентгенівських знімків та інших медичних зображень на предмет ознак захворювань [1]. В галузі безпеки компанія Hikvision, використовує алгоритми ШІ для розпізнання рухів та обличчя в системах захисту власного виробництва [2]. Один із найвідоміших застосувань ШІ в світі є Google Translate від компанії Google, за допомогою машинного навчання штучний інтелект змогли «навчити» понад 100 мовам, а також для кожного користувача формується своя база даних [3]. Хоч штучний інтелект і не набув широкого використання в державних реєстрах, але активно використовується різними сервісами та компаніями.

В Україні використання ШІ в діяльності державних інституцій тільки набуває поширення, так наприклад для оптимізації судової системи України був створений ШІ у складі проекту «Суд на долоні» та сервісу WINCOURT [4]. Ці платформи використовують алгоритми машинного навчання для аналізу судових документів і прогнозування рішень. Для аналізу державних закупівель також був створений ШІ. Платформа DOZORRO використовує машинне навчання для моніторингу тендерів і виявлення корупційних ризиків [20]. Це дозволяє виявляти та вирішувати проблеми прозорості і корупції на ранніх стадіях, покращуючи ефективність використання державних ресурсів.

В медичній галузі, розробки Інституту кібернетики ім. В.М. Глушкова НАН України внесли значний вклад у покращення діагностики захворювань. Автоматичний аналіз медичних зображень за допомогою ШІ дозволяє виявляти патології на ранніх стадіях, значно підвищуючи шанси на успішне лікування.

В сільському господарстві ШІ забезпечує зменшення втрат урожаю та оптимізацію використання ресурсів. Компанія «Агрокедр» використовує алгоритми для аналізу зображень посівів, допомагаючи виявляти шкідників та хвороби. Це дозволяє фермерам своєчасно реагувати на загрози та застосовувати захисні заходи, збільшуючи продуктивність та знижуючи витрати.

У транспортній галузі, «Укрзалізниця» інтегрувала ШІ для оптимізації розкладів та маршрутів поїздів. Ці заходи сприяють зменшенню часу поїздок та підвищенню комфорту пасажирів,

роблячи залізничний транспорт більш привабливим для користувачів.

В освітній сфері ШІ реалізовано через платформу Prometheus, яка використовує персоналізацію для адаптації навчальних програм до індивідуальних потреб учнів. Це підвищує ефективність навчання та мотивацію студентів.

Розвиток ШІ в Україні супроводжується рядом викликів, включаючи питання захисту даних, етичні дилеми пов'язані з автоматизацією важливих рішень, та необхідність регулятивних змін для підтримки інновацій. Проте, потенціал для позитивного впливу на економіку і суспільство робить перспективу використання ШІ в Україні актуальною особливо під час вирішення завдань із імплементації положень «Національної стратегії розвитку штучного інтелекту в Україні», які описують цілі та пріоритети розвитку ШІ в Україні, а також містять рекомендації щодо використання ШІ в державному секторі [5]. У відповідності до прийнятої стратегії інтеграції ШІ у процеси обробки інформації в державних базах даних може стати ключем до їхньої модернізації та ефективного функціонування.

Метою даної роботи є вивчення можливостей інтеграції алгоритмів штучного інтелекту для покращення обробки інформації в державних базах даних.

Незважаючи на поширення використання ШІ під час вирішення задач з оптимізації, аналіз науково-технічних джерел та відкритих публікацій показав, що дані для реалізації алгоритмів штучного інтелекту, щодо покращення обробки інформації в державних базах даних, майже відсутні.

Виклад основного матеріалу. *Розробка програмного коду із використанням ШІ для обробки даних.* Одним із найбільш поширеніших документів, який використовується державними органами є паспорт громадянина України, останнім часом активно впроваджується у формі ID картки. Тому для вивчення можливостей інтеграції алгоритмів штучного інтелекту для покращення обробки інформації в державних базах даних був розроблений алгоритм скану даних з паспорта громадянина України, програма передбачає обробку зворотної частини ID картки, де знаходиться MRZ код. Програмний код представлений на рис. 1 та 2.

```

import cv2
from PIL import Image
import pytesseract
import re
import numpy as np

pytesseract.pytesseract_tesseract_cmd = 'C:\\Program Files\\Tesseract-OCR\\tesseract.exe'

def preprocess_image(image):
    gray = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)
    gray = cv2.equalizeHist(gray)
    sharpen_kernel = np.array([[ -1, -1, -1], [-1, 9, -1], [-1, -1, -1]])
    sharpen = cv2.filter2D(gray, -1, sharpen_kernel)
    thresh = cv2.adaptiveThreshold(sharpen, 255, cv2.ADAPTIVE_THRESH_GAUSSIAN_C, cv2.THRESH_BINARY, 11, 2)
    return thresh

def find_mrz_region(image):
    height, width = image.shape[:2]
    roi_height = int(height * 0.48)
    mrz_region = image[height - roi_height:height, :]
    return mrz_region

def parse_mrz(mrz_text):
    lines = mrz_text.strip().splitlines()
    if len(lines) < 3:
        return "Error"

    doc_type, country_code, doc_number = lines[0][:2], lines[0][2:5], lines[0][5:14]
    birth_date, gender, expiry_date = lines[1][:6], lines[1][7], lines[1][8:14]
    nationality, surname, given_names = lines[2][:3], lines[2][3:17].replace('<', ' '), lines[2][18:].replace('<', ' ').strip()

    return {
        'Document Type': doc_type,
        'Country Code': country_code,
        'Document Number': doc_number,
        'Birth Date': format_date(birth_date),
        'Gender': gender,
        'Expiry Date': format_date(expiry_date),
        'Nationality': nationality,
        'Surname': surname,
        'Given Names': given_names
    }

```

Рис. 1. Елементи програмного коду обробки MRZ блоку (частина 1)

```

def format_date(date_str):
    year, month, day = int(date_str[:2]), int(date_str[2:4]), int(date_str[4:6])
    return f"{day:02}/{month:02}/{'20' + str(year) if year < 38 else '19' + str(year)}"

def extract_mrz_data(image_path):
    image = cv2.imread(image_path)
    mrz_region = find_mrz_region(image)
    processed_image = preprocess_image(mrz_region)
    pil_image = Image.fromarray(processed_image)
    custom_config = r'--oem 3 --psm 6 -c tessedit_char_whitelist=ABCDEFGHIJKLMNOPQRSTUVWXYZ0123456789<'
    extracted_text = pytesseract.image_to_string(pil_image, config=custom_config)
    print(f"MRZ: {extracted_text}")
    mrz_data = parse_mrz(extracted_text)
    return mrz_data

image_path = 'C:\\passport_app\\uploaded_images\\UAID_Be.jpg'
extracted_mrz = extract_mrz_data(image_path)
print(extracted_mrz)

```

Рис. 2. Елементи програмного коду обробки MRZ блоку (частина 2)

Розроблений код Python призначений для автоматичного зчитування інформації про власника з машиночитабельної зони

(MRZ) зображення паспорта. MRZ - це стандартизований оптично зчитуваний блок, який містить ключові дані про власника паспорта у стислому вигляді. Основні функції, які використані під час розробки:

1. «preprocess_image»: Функція `preprocess_image` виконує попередню обробку зображення паспорта, щоб покращити точність розпізнавання тексту. Вона перетворює зображення в градації сірого, розмиває його та застосовує адаптивну порогову обробку для виділення тексту на тлі;
2. «find_mrz_region»: Функція `find_mrz_region` знаходить та вирізає область MRZ зображення. Вона використовує розміри вхідного зображення для визначення приблизного розташування MRZ, зазвичай займаючи нижню частину паспорта;
3. «parse_mrz»: Функція `parse_mrz` аналізує розпізнаний текст MRZ, який зазвичай складається з трьох рядків. Кожен рядок містить закодовану інформацію про власника паспорта. Функція використовує регулярні вирази або інші методи для вилучення даних, таких як тип документа, код країни, номер документа, дата народження, стать, дата закінчення терміну дії, національність, прізвище та ім'я;
4. «format_date»: Функція `format_date` перетворює дати народження та дати закінчення терміну дії, отримані з MRZ, з формату YYMMDD на більш читабельний формат, наприклад DD/MM/YYYY;
5. «extract_mrz_data»: Функція `extract_mrz_data` об'єднує інші функції для виконання повного процесу розпізнавання даних MRZ. Вона читає зображення паспорта, обробляє його за допомогою `preprocess_image`, розпізнає текст MRZ за допомогою `pytesseract`, аналізує розпізнаний текст за допомогою `parse_mrz`, формує словник з даними та повертає його.

Під час тестування використовувались данні із ID картки, яка є прикладом. Скан використаної ID картки наведено на рис. 3. Розроблене програмне забезпечення виконувало аналіз даних в декілька кроків, а саме:

Крок 1. Завантаження зображення. На цьому кроці для обробки зображення, воно було завантажено в робочу директорію та присвоєне йому ім'я, яке використано в коді.

В результаті тестування програмами на основі ШІ було отримано, що використаний алгоритм в розробленій програмі успішно обробив дані з ID картки, єдиною дією людини є завантаження зображення та запуск коду. Мінімізації дій із залученням людини дозволить скоротити час на обробку даних та зменшити вплив людського фактору на результат.

Отже, автоматизація розпізнавання MRZ на основі використання алгоритмів ШІ може стати основою для створення єдиної системи ідентифікації громадян, яка поєднуватиме дані з різних державних баз даних. Це може спростити доступ до державних послуг, зменшити адміністративний тягар та підвищити рівень безпеки. Завдяки єдиній системі ідентифікації громадяни зможуть отримувати послуги швидше та зручніше, не витрачаючи час на повторне надання своїх даних.

Висновки та пропозиції. Отже, в даній статті було досліджено можливість із інтеграції алгоритмів штучного інтелекту для покращення обробки інформації в державних базах даних. Показано, що алгоритм розпізнавання даних MRZ з зображення паспорта володіє значним потенціалом для використання в державних базах даних, адже він може суттєво покращити ефективність та безпеку різноманітних державних послуг.

© **Приймак Є.О., Лемешко А.В., Зайцев Є.О., Антоненко А.В., 2024**

ЛІТЕРАТУРА

1. Zebra medical vision unveils ai-based algorithm to detect brain bleeds. Healthcare IT News. Режим доступу: <https://hitconsultant.net/2018/03/05/zebra-medical-vision-unveils-ai-based-algorithm-detect-brain-bleeds/>.
2. Hikvision company profile. Hikvision. Режим доступу: <https://www.hikvision.com/europe/about-us/company-profile/#:~:text=Hikvision%20provides%20a%20broad%20range,safety%20management%20and%20business%20intelligence>.
3. Google translate architecture. La Vivien Post. Режим доступу: <https://www.lavivienpost.com/google-translate-and-transformer-model/>.
4. Як штучний інтелект використовується у сфері відкритих даних. Дія.Відкриті дані. Режим доступу: <https://diia.data.gov.ua/info-center/aiod>.
5. Національна стратегія розвитку штучного інтелекту в Україні 2021-2030. Режим доступу: https://wp.oecd.ai/app/uploads/2021/12/Ukraine_National_Strategy_for_Development_of_Artificial_Intelligence_in_Ukraine_2021-2030.pdf.

6. Твердохліб А.О., Коротін Д.С. Ефективність функціонування комп'ютерних систем при використанні технології блокчейн і баз даних. *Таврійський науковий вісник. Серія: Технічні науки*, 2022, (6).

7. Цвик О.С. Аналіз і особливості програмного забезпечення для контролю трафіку. *Вісник Хмельницького національного університету. Серія: Технічні науки*, 2023, (1).

8. Новіченко Є.О. Актуальні засади створення алгоритмів обробки інформації для логістичних центрів. *Таврійський науковий вісник. Серія: Технічні науки*, 2023 (1).

9. Зайцев Є.О. Smart засоби визначення аварійних станів у розподільних електричних мережах міст. *Таврійський науковий вісник. Серія: Технічні науки*, 2022, (5).

10. Scratch S. R. Object-Oriented and classical software engineering. McGraw Hill, 2007. 654 p.

REFERENCES

1. Zebra medical vision unveils ai-based algorithm to detect brain bleeds. Healthcare IT News. Режим доступу: <https://hitconsultant.net/2018/03/05/zebra-medical-vision-unveils-ai-based-algorithm-detect-brain-bleeds/>.

2. Hikvision company profile. Hikvision. Режим доступу: <https://www.hikvision.com/europe/about-us/company-profile/#:~:text=Hikvision%20provides%20a%20broad%20range,safety%20management%20and%20business%20intelligence.>

3. Google translate architecture. La Vivien Post. Режим доступу: <https://www.lavivienpost.com/google-translate-and-transformer-model/>.

4. How artificial intelligence is used in the field of open data. Action. Open data. Режим доступу: <https://diia.data.gov.ua/info-center/aiod.>

5. National strategy for the development of artificial intelligence in Ukraine 2021-2030. Режим доступу: https://wp.oecd.ai/app/uploads/2021/12/Ukraine_National_Strategy_for_Development_of_Artificial_Intelligence_in_Ukraine_2021-2030.pdf.

6. Tverdokhlіb A.O., Korotin D.S. Efektivnist funktsionuvannya kompiuternykh system pry vykorystanni tekhnolohii blokchein i baz dannykh. *Tavriiskiyi naukoviyi visnyk. Serii: Tekhnichni nauky*, 2022, (6).

7. Tsyvyk O.S. Analiz i osoblyvosti prohramnoho zabezpechennia dlia kontroliu trafiku. *Visnyk Khmelnytskoho natsionalnoho universytetu. Serii: Tekhnichni nauky*, 2023, (1).

8. Novichenko Ye.O. Aktualni zasady stvorennia alhorytmiv obrobky informatsii dlia lohistychnykh tsestriv. *Tavriiskiyi naukoviyi visnyk. Serii: Tekhnichni nauky*, 2023 (1).

9. Zaitsev Ye.O. Smart zasoby vyznachennia avariinykh staniv u rozpodilnykh elektrychnykh merezhakh mist. *Tavriiskiyi naukoviyi visnyk. Serii: Tekhnichni nauky*, 2022, (5).

10. Scratch S. R. Object-Oriented and classical software engineering. McGraw Hill, 2007, Pp. 654 [in English].

СТАТТЯ НАДІЙШЛА ДО РЕДАКЦІЇ 03.04.2024