



Поточні тенденції у сфері розпізнавання тексту із застосуванням штучного інтелекту для криміналістичних досліджень

Тетяна Варава

провідний судовий експерт, Український науково-дослідний інститут спеціальної техніки
та судових експертиз Служби безпеки України, м. Полтава, Україна, e-mail: grafio4tan@gmail.com

Досліджено сучасні технології розпізнавання тексту на основі штучного інтелекту та їх застосування у криміналістичних дослідженнях. Проаналізовано методи розпізнавання рукописного тексту, алгоритми машинного навчання та можливості оптичних систем обробки документів. Визначено перспективи використання інтелектуальних систем для вирішення ідентифікаційних завдань.

Ключові слова: штучний інтелект; OCR; криміналістика; почерк; ідентифікація.

Current Trends in Text Recognition Using Artificial Intelligence

Tetiana Varava

The article examines modern text recognition technologies based on artificial intelligence and their application in forensic research. Methods for handwritten text recognition, machine learning algorithms, and optical document processing systems are analyzed. Prospects for the use of intelligent systems for identification tasks are defined.

Keywords: artificial intelligence; OCR; forensics; handwriting; identification.

Стрімкий розвиток інформаційних технологій зумовлює необхідність переходу до автоматизованих методів обробки документальної інформації. Особливо актуальним це питання є для криміналістичної діяльності, де значні обсяги текстових матеріалів потребують швидкого аналізу, систематизації та подальшого використання в експертній практиці. Одним з основних напрямів виконання зазначених завдань є застосування технологій розпізнавання тексту на основі штучного інтелекту.

Сучасні системи інтелектуального аналізу зображень дають змогу здійснювати цифровізацію документів різного типу. Це значно розширює можливості криміналістичних досліджень, оскільки забезпечує автоматизоване виокремлення інформативних ознак, їх порівняльний аналіз і формування обґрунтованих експертних висновків.

Особливу складність становить аналіз рукописного тексту, який характеризується значною варіативністю графічних ознак, індивідуальними особливостями почерку та впливом зовнішніх чинників під час письма. Традиційні методи почеркознавчої експертизи передбачають тривале дослідження зразків і залежать від досвіду експерта. Застосування алгоритмів машинного навчання дає змогу автоматизувати визначення стійких оз-

нак почерку, підвищити об'єктивність результатів та скоротити час проведення експертиз.

Окрім того, інтеграція технологій комп'ютерного зору й нейронних мереж створює передумови для побудови комплексних криміналістичних систем, здатних одночасно виконувати розпізнавання тексту, аналіз структури документів та ідентифікацію автора рукопису. Це відкриває нові можливості для автоматизації експертної діяльності й підвищення ефективності розслідування правопорушень.

Метою статті є дослідження сучасних технологій розпізнавання тексту із застосуванням штучного інтелекту та визначення їх потенціалу для виконання криміналістичних ідентифікаційних завдань.

Аналіз технологій і програмних засобів для розпізнавання тексту. Системи розпізнавання тексту базуються на технологіях оптичного розпізнавання символів (OCR), інтелектуального розпізнавання символів (ICR) і розпізнавання рукописного тексту (HTR). Вони послуговуються методами комп'ютерного зору, глибокого навчання та статистичного аналізу. Основними компонентами таких систем є модулі попередньої обробки, сегментації, виокремлення ознак і класифікації [1].

Сучасні програмні засоби застосовують згорткові нейронні мережі, рекурентні нейронні мережі та трансформерні моделі. Такі підходи дають змогу аналізувати послідовності символів і враховувати контекст тексту. Комбінація *CNN* і *LSTM* забезпечує ефективне розпізнавання складних рукописних текстів [2].

Аналіз і розпізнавання рукописного тексту. Розпізнавання рукописного тексту містить такі етапи:

- попередня обробка зображення (фільтрація шумів, нормалізація);
- бінаризація;
- сегментація рядків і символів;
- виокремлення ознак;
- класифікація;
- постобробка результатів.

Для виокремлення ознак користуються гістограмами орієнтованих градієнтів, контурними дескрипторами, скелетизацією символів та аналізом текстур. Сучасні підходи передбачають застосування глибоких нейронних мереж, що дають змогу автоматично визначати релевантні ознаки без ручного програмування [3].

Дослідження стилів почерку. Аналіз почерку є важливим напрямом криміналістичних досліджень. Основні характеристики почерку містять нахил символів, розмір і пропорції літер, міжсимвольні інтервали, швидкість письма, силу натиску й особливості з'єднання елементів. Методи машинного навчання дають змогу створювати моделі ідентифікації автора за почерком. Для цього застосовують метод опорних векторів, нейронні мережі, алгоритми кластеризації та статистичні моделі [4].

Основні принципи розпізнавання тексту. До основних алгоритмів розпізнавання тексту належать:

- метод *k*-ближчих сусідів;
- метод опорних векторів;
- згорткові нейронні мережі;
- рекурентні нейронні мережі;
- трансформерні моделі.

Процес розпізнавання містить формування векторів ознак і їх подальшу класифікацію. Користування трансформерними моделями дає змогу враховувати контекст символів і підвищує точність розпізнавання [5].

Можливості оптичних систем обробки рукописного тексту. Оптичні системи обробки рукописного тексту забезпечують:

- автоматичну цифровізацію документів;
- розпізнавання рукописних записів;
- пошук у рукописних архівах;
- визначення автора тексту;
- аналіз структури документа.

Інтеграція таких систем з базами даних дає змогу створювати автоматизовані криміналістичні комплекси для дослідження документів [6].

Провідні підприємства й установи у сфері розпізнавання тексту. Сучасний ринок систем розпізнавання тексту представлений міжнародними компаніями й відкритими дослідницькими проектами. Вони створюють програмні рішення для автоматичної цифровізації документів, аналізу рукописів та інтелектуальної обробки текстових даних. Поміж них можна назвати корпоративні платформи, хмарні сервіси та відкриті бібліотеки для досліджень.

Моделі, методи й інформаційні технології для виконання ідентифікаційних криміналістичних завдань.

Основні методи містять:

- нейромережевий аналіз почерку;
- статистичний аналіз параметрів письма;
- порівняльний аналіз зразків;
- біометричні методи ідентифікації;
- методи кластеризації.

Користування глибоким навчанням дає змогу визначати автора тексту з високою точністю навіть за обмеженої кількості зразків [7].

Тенденції розвитку технологій розпізнавання тексту у криміналістиці. Сучасні системи створюють із застосуванням бібліотек комп'ютерного зору, фреймворків машинного навчання, технологій обробки зображень, хмарних платформ і систем керування базами даних. Користування хмарними технологіями дає змогу масштабувати системи й забезпечити доступ до значних обчислювальних ресурсів [8].

Аналіз сучасних досліджень у сфері розпізнавання тексту дає змогу виокремити



кілька основних тенденцій розвитку технологій, що мають безпосереднє значення для криміналістичної практики.

По-перше, спостерігається перехід від класичних алгоритмів розпізнавання символів до комплексних моделей глибокого навчання, які одночасно здійснюють сегментацію, виокремлення ознак і класифікацію тексту. Такі системи дають змогу мінімізувати кількість попередніх операцій та підвищити точність розпізнавання рукописних документів.

По-друге, активно впроваджуються трансформерні архітектури, що забезпечують урахування контексту під час розпізнавання тексту. Це особливо важливо для криміналістичних досліджень, де необхідно аналізувати короткі записи, підписи та фрагменти документів з обмеженим обсягом інформації.

По-третє, значного розвитку набувають технології мультимодального аналізу, які поєднують дослідження текстових, графічних і біометричних ознак. Користування такими підходами дає змогу одночасно оцінювати структуру документа, особливості почерку й параметри письма.

По-четверте, спостерігається інтеграція систем розпізнавання тексту з криміналістичними інформаційними базами даних. Це дає змогу формувати електронні профілі почерку та проводити автоматизований порівняльний аналіз зразків.

По-п'яте, перспективним напрямом є користування хмарними технологіями для обробки великих масивів документів. Такий підхід дає змогу масштабувати обчислювальні ресурси й забезпечити доступ до інтелектуальних систем для експертних установ.

Висновки. Системи розпізнавання тексту на основі штучного інтелекту є перспективним напрямом розвитку криміналістичних досліджень. Користування сучасними алгоритмами машинного навчання дає змогу автоматизувати аналіз рукописного тексту, підвищити точність ідентифікації автора та зменшити суб'єктивність експертних оцінок [9—11]. Подальші дослідження слід спрямувати на вдосконалення алгоритмів глибокого навчання, розширення баз даних почерків та інтеграцію інтелектуальних систем у криміналістичну практику.

В Україні аналогічні дослідження проводять у Львівському національному університеті імені Івана Франка, Державному університеті інформаційно-комунікаційних технологій і Чорноморському національному університеті імені Петра Могили. Це створює передумови для розроблення сучасних аналітичних інструментів у напрямках почеркознавства, технічної експертизи документів тощо.

Перелік джерел посилання

1. Smith R. An overview of the Tesseract OCR engine. *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)*. 2007. P. 629—633.
2. Graves A., Schmidhuber J. Offline handwriting recognition with multidimensional recurrent neural networks. *Advances in Neural Information Processing Systems*. 2009. P. 545—552.
3. Goodfellow I., Bengio Y., Courville A. Deep learning. Cambridge : MIT Press, 2016. 800 p.
4. Srihari S. Handwriting identification: facts and fundamentals. *Journal of Forensic Sciences*. 2002. Vol. 47. № 4.
5. Vaswani A. Attention is all you need. *Advances in Neural Information Processing Systems*. 2017.
6. Trier O., Jain A., Taxt T. Feature extraction methods for character recognition. *Pattern Recognition*. 1996. Vol. 29.
7. Impedovo D., Pirlo G. Automatic signature verification: the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics*. 2008.
8. Bishop C. Pattern recognition and machine learning. New York : Springer, 2006.
9. Свобода Є., Григорович Т., Кравець В. Застосування математичного аналізу при визначенні ідентифікаційних ознак почерку (практичний аспект). *Актуальні проблеми права*. 2025. № 2. С. 118—125.
10. Kerniakevych-Tanasiichuk Yu. V., Sezonov V. S., Nychytailo I. M., et al. Problems of forensic identification of handwriting in forensic examination. *Journal of the National Academy of Legal Sciences of Ukraine*. 2021. Vol. 28 (1). P. 195—204. DOI: 10.37635/jnalsu.28(1).2021.195-204 (дата звернення: 12.04.2026).
11. Krasniuk I. P. Problems of forensic handwriting examination in the analysis of signatures and short notes. *Scientific Journal of the National Academy of Internal Affairs*. 2022. Vol. 27 (1). P. 73—78. DOI: 10.33270/0122271.73 (дата звернення: 12.04.2026).