

Invoice and receipt optical character recognition: review on current methods and future trends

Albana Rexhepi^{1,†}, Erijon Hasi^{1,†}, Art Haxholli^{1,†} and Eliot Bytyçi^{1,*†}

¹ University of Prishtina, Avenue Mother Teresa, No-5, 10000, Prishtinë, Republic of Kosova

Abstract

Traditional invoices and receipts remain a crucial part of financial record-keeping, but manual processing is time-consuming and error prone. Optical Character Recognition (OCR) automates text extraction, improving efficiency. This review systematically evaluates OCR solutions for invoice and receipt recognition, focusing more on open-source models. We conducted a structured search across several libraries such as IEEE Xplore, ACM Digital Library, SpringerLink, and ScienceDirect, filtering studies from 2019–2024, while using predefined inclusion criteria. Performance metrics such as Character Error Rate (CER) and Word Error Rate (WER) guided our analysis. Results highlight Tesseract as the most widely used OCR tool, with deep learning-based solutions gaining traction. Limitations include exclusion of proprietary models and older studies. Findings provide insights into current OCR advancements and their application in financial document digitization.

Keywords

optical character recognition, receipt, invoice, digitization, open source

1. Introduction

Paper-based invoices and receipts are ubiquitous in both personal and business contexts, with billions printed annually worldwide [1]. These documents are essential for tracking expenses, managing financial records, and ensuring accurate accounting. However, manually reviewing and processing these physical records is labor-intensive and time-consuming, often hindering individuals and organizations from effectively managing their finances. Moreover, this process tends to be prone to human error, which potentially accumulates over time, resulting in significant financial discrepancies and misreporting. Thus, there exists a need to automate the process of converting printed or handwritten text into digital data, through a technology called Optical Character Recognition (OCR).

Optical Character Recognition technology has been around for decades, especially in the format known today since its introduction by Kurzweil in 1974 [2]. Nevertheless, it has gained significant attention in recent years due to the development of more sophisticated machine learning algorithms and cloud-based services. Today, OCR technology is integrated into various applications, from document scanning software to mobile apps, offering users a range of tools for digitizing and processing text from images and scanned documents.

Nevertheless, despite recent advances in OCR technology, extracting accurate information from complex documents, like invoices, remains a significant challenge. This problem occurs because there exists a variety of layouts, structures, fonts, and languages making it difficult for OCR systems to accurately capture relevant data such as amounts, vendor details, and dates [3]. Furthermore, issues

¹6th International Conference Recent Trends and Applications in Computer Science and Information Technology RTA-CSIT, May 22-24, 2025, Tirana, Albania

* Corresponding author.

[†] These authors contributed equally.

✉ albana.rexhepi2@student.uni-pr.edu (A. Rexhepi); erijon.hasi@student.uni-pr.edu (E. Hasi); art.haxholli@student.uni-pr.edu (A. Haxholli); eliot.bytyci@uni-pr.edu (E. Bytyçi)

ORCID 0009-0000-9479-5358 (A. Rexhepi); 0009-0008-0979-5870 (E. Hasi); 0009-0005-9194-2654 (A. Haxholli); 0000-0001-7273-9929 (E. Bytyçi)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

such as poor image quality, background noise, and skewed scanning can also reduce OCR performance.

This research evaluated current OCR technologies, specifically in recognizing and digitizing invoices. Different kinds of metrics were used to benchmark models and some of those metrics will include character error rate and word error rate. At the end of the review, we were able to attest the best model, which we will try to use in our own application in the near future. This is not only valuable in terms of understanding the state of the art but also improving the overall digitization process.

The methodology for this literature review involved a systematic approach to identify, evaluate, and synthesize relevant research on OCR technology. Initially, a comprehensive search was conducted using four academic databases with keywords revolving around OCR. The search was focused on peer-reviewed journal articles and conference papers, but not on reviews or other technical/industry reports. After gathering the initial set of papers, inclusion and exclusion criteria were applied based on relevance, quality, and methodological rigor. Each selected paper was analyzed for key findings, methodologies used, and implications for future research.

The rest of the paper is structured as follows: section two describes methodology used for the literature review, while section three reports on the papers gathered for analysis. Section four concludes the paper and presents limitations but also future research possibilities.

2. Modifications

This literature review is aimed at evaluating current OCR solutions, with a particular focus on their application in recognizing invoices and receipts. One of the aspects of the study was to benchmark these solutions using key performance metrics, including character error rate and word error rate. Ultimately, the goal is to identify the most effective open-source OCR model that can be integrated into our own application, thereby enhancing overall digital processing efficiency and contributing to the understanding of the state-of-the-art in OCR technology.

To gather the necessary data, we turned to a systematic review of articles and conference papers published between 2019 and until our revised date in December 2024, thus covering the last five years. Our quest for knowledge was guided by the search query "OCR" AND "receipt" AND "open source", which we carefully applied across several major electronic databases, including IEEE Xplore, the ACM Digital Library, SpringerLink, and ScienceDirect. In the early stages of our investigation, we screened each document based on its title, keywords, and abstract to gather its relevance to OCR for receipt and invoice recognition.

For instance, our initial quest into IEEE Xplore included 2 documents that were promising enough to warrant further evaluation. In the ACM Digital Library, 20 documents were identified, and after a closer examination - on a second review, only 5 of them were selected for full review. Similarly, we navigated through 84 documents from SpringerLink and 35 from ScienceDirect initially, then ultimately narrowing these down to 39 and 13, respectively, based on their alignment with our research objectives.

A strict criterion was maintained throughout the process, as depicted in figure 1, to ensure our focus remained clear. Thus, only studies that concentrated on OCR algorithms applied to receipt or invoice recognition involved open-source solutions, and that reported relevant performance metrics, were included. Additionally, we restricted our review to documents published in English between 2019 and 2024 and limited our sources to articles and conference papers. Studies that did not meet these standards, whether because they were not in English or lacked a clear focus on OCR solutions were excluded from consideration. In the end, there were few studies that we were unable to fetch, even after contacting the authors directly, and thus excluded them from our review.

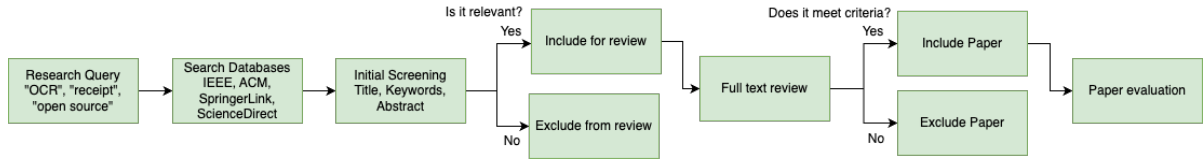


Figure 1. Paper review process

Thus, we recognize that our review is not without limitations. The focus on a specific publication window (2019–2024) means that earlier foundational work may not be fully represented, potentially omitting valuable historical context. Additionally, by concentrating solely on open-source solutions, the review might overlook advancements made in proprietary systems that could offer relevant insights. Finally, the initial screening process based solely on titles, keywords, and abstracts may have unintentionally excluded some studies that would have contributed to a more comprehensive analysis if examined in full.

At every step, we were mindful of ethical considerations, adhering to established protocols such as PRISMA to minimize bias and ensure the integrity of our review process. This commitment to ethical research practices not only enhanced the transparency of our methodology but also reinforced the credibility of our findings. In following these rigorous guidelines, we strived to create a review that was as comprehensive and objective as possible, paving the way for a meaningful contribution to the field of OCR technology.

3. Analysis of the papers

In this section of the paper, we will present the findings from selected papers included in the literature review. Initially, as presented in figure 2, we see that there is an increasing trend on the paper related to OCR, from the papers that we have selected for further analysis, thus also strengthening our motivation to deal with the research in the field.

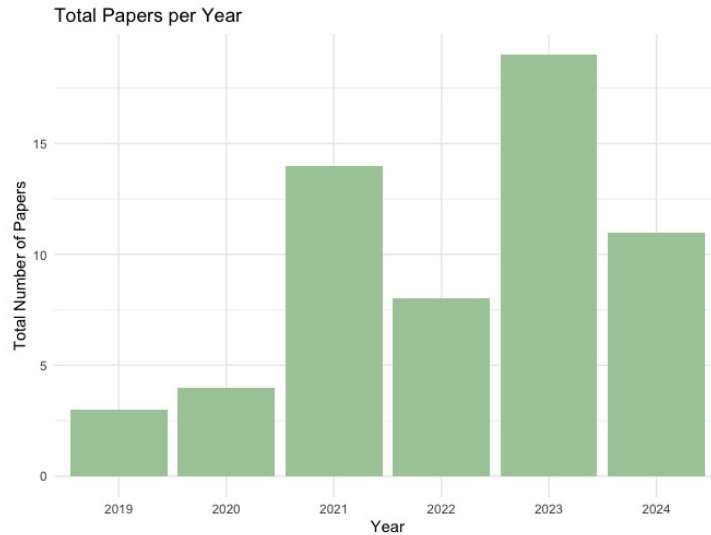


Figure 2. Yearly trend of papers selected

3.1. Open-source vs proprietary

Among the various OCR systems discussed in the literature, Tesseract emerged as the most widely used, appearing in 15 different studies [4, 6, 10, 19, 21, 25-28, 32, 33, 37, 41-43]. This contrasts with the usage of other OCR tools, such as MMOCR [5], Kraken, OCRopus, Kalamari [6], docTR and PaddleOCR [23], and EasyOCR [23, 26].

Even though they were not the focus of our study, still several proprietary OCR solutions were mentioned in the analyzed papers, including Google Vision [18, 19], Amazon Textract [19, 23-24, 37], Microsoft Computer Vision [19, 25], and Abby FineReader [6, 20]. These proprietary systems offer a ready-to-use experience, eliminating the need for extensive configuration and dependency setup, which is often required with open-source alternatives, such as Tesseract e.g..

A growing trend in the research involves integrating deep learning techniques into OCR systems, as highlighted in several studies [16-17, 28-29, 34, 39]. Furthermore, intelligent character recognition (ICR) technology, which focuses on processing handwritten documents, has also gained attention in a few papers [11, 44]. This reflects an ongoing effort to enhance OCR capabilities, particularly for more complex tasks like handwriting recognition.

In terms of distribution, the reviewed papers predominantly relied on open-source OCR systems, with 17 studies utilizing them exclusively. In contrast, 5 papers were based solely on closed-source solutions, while 4 studies incorporated both open-source and proprietary OCR tools. This demonstrates a clear preference for open-source options, likely due to their flexibility and adaptability, despite the convenience offered by proprietary systems.

We have presented some of the above-mentioned information as cumulative in table 1, for better visibility and understanding.

Table 1

Cumulative description of the OCR's

Category	Details	References
Most Frequent OCR	Tesseract (used in 15 papers)	[4, 6, 10, 19, 21, 25-28, 32, 33, 37, 41, 42, 43]
Other Open-Source OCR	MMOCR, Kraken, OCRopus, Kalamari, docTR, PaddleOCR, EasyOCR	[5, 6, 23, 26]
Proprietary OCR	Google Vision, Amazon Textract, Microsoft Computer Vision, Abby FineReader	[18-20, 23-25, 37]
Deep Learning in OCR	Integration of deep learning techniques into OCR systems	[16-17, 28-29, 34, 39]
ICR Technology	Focus on intelligent character recognition for handwritten documents	[11, 44]

3.2. Preprocessing phase

One of the important aspects to consider was also the techniques that were used in the preprocessing phase. A few papers [5, 7-8, 12, 25-26] utilized bounding boxes to identify and locate text placements within images. In contrast, a study [9] focused primarily on text-specific preprocessing techniques. Additionally, some authors proposed automated image enhancement methods to improve OCR accuracy, as seen in [10, 14, 27-28, 35, 40]. However, in certain cases, such as handling blurred images [14] or resizing images [16, 29], manual verification and intervention were also incorporated to ensure optimal results. This combination of automated and manual preprocessing highlights the diverse strategies used to address varying challenges in OCR workflows.

3.3. Metrics used to evaluate solutions

The metric used was also another important aspect of our research. Recent literature on Optical Character Recognition (OCR) demonstrates a diverse array of evaluation metrics designed to assess various dimensions of system performance. Commonly used metrics include Character Error Rate (CER) and Word Error Rate (WER), which quantify errors by counting deletions, insertions, and substitutions between OCR outputs and ground truth data [4, 26, 28, 29]. Additionally, fundamental measures such as binary accuracy, precision, recall, F1 score, and H-mean are frequently employed to evaluate performance [4, 5, 7-9, 10, 14, 16-18, 21, 27, 30-31]. Moreover, some studies focus on word/character accuracy [6] or directly compare OCR results with human-reviewed ground truth [19, 22].

For more specialized tasks like entity linking, some more specialized metrics such as mAP (mean Average Precision), mRank, and Hit@1, Hit@2, and Hit@5 are utilized [8]. Other research employs average precision, confusion matrices, and Intersection-over-Union (IoU) to evaluate performance [16, 41, 44]. Image quality, a critical factor in OCR, is often assessed using different kinds of metrics like PSNR, MS-SSIM, PieAPP, WaDIQaM, LPIPS, and DISTS [35, 43].

Furthermore, specialized criteria such as Box Error Rate, Expected Calibration Error [8], and Average Normalized Levenshtein Similarity [25] are reported in some studies. Traditional metrics like ISRI and GTmetrix [37-38], as well as k-fold cross-validation, are also used. Privacy-focused measures, including Discernibility Metric Cost and Minimal Average Group Size, are applied in certain contexts [11, 13, 39, 42]. Additional metrics like sensitivity, specificity, local distortion, and edit distance are occasionally included to provide a more comprehensive evaluation [34, 43-44]. This wide range of metrics underscores the multifaceted nature of OCR performance assessment, tailored to the specific requirements of different tasks and applications.

3.4. Specific solutions related to metrics evaluation

Among the OCR systems reviewed, Tesseract stood out as the most commonly used open-source option, showing reliable accuracy (83.36% with an 8.68% character error rate), such as license plate recognition [26, 33] and digitizing historical documents [19]. Commercial engines, such as Google Document AI and Amazon Textract, performed better on low-quality documents, likely due to refined calibration and precise text detection [23, 37].

Open-source OCR engines like MMOCR [5], which supports text detection, recognition, and downstream tasks (e.g., named entity recognition) with 14 state-of-the-art algorithms, EasyOCR [26], achieving 67.97% accuracy (CER 16.46) in evaluations, and PaddleOCR [8], integrated with frameworks like StrucTexT for entity labeling, were analyzed. Multimodal architectures like LayoutLM (79.27% F1 for form understanding) [7] and StrucTexT (state-of-the-art entity labeling via token- and segment-level representations) [8] seem to excel in layout-sensitive tasks. Preprocessing pipelines, such as binarization [10], geometric corrections (dewarping via GeoTrTemplateLarge, reducing distortion by 26.1% [43]), and super-resolution (ESRGAN improving accuracy to 85% [40]), proved critical for avoiding errors in degraded or handwritten texts [42].

Post-processing strategies, including neural network-based correction (BERT [6]) and rule-based filtering (3-step Levenshtein distance matching [18]), enhanced multilingual accuracy, particularly for scripts like Bengali [12] and Urdu [28]. OCR applications spanned domain-specific use cases, such as fraud detection in legal documents (automated classification with Apache Spark [20], using ABBYY FineReader), expense auditing via SPARQL queries [9], and historical record extraction [34]. Persistent challenges include handling mixed scripts (e.g., Latin-Arabic [28]), handwritten-mixed documents [42], and the performance disparity between open-source and commercial tools [23], underscoring the need for scalable multimodal frameworks and crowdsourced datasets for low-resource languages [12].

Several key datasets have established themselves as foundational resources in OCR research, consistently appearing across studies to benchmark and advance the field. The ICDAR series (ICDAR2015, ICDAR2017, ICDAR2019) is widely adopted, particularly for benchmarking [5, 6, 26, 28], alongside RVL-CDIP for document classification [7, 8, 13, 24]. Receipt and form-oriented datasets like SROIE, CORD, and FUNSD are frequently employed for tasks such as information extraction [7, 8, 14, 21, 23, 26, 27, 36, 38]. Specialized datasets include the IIT-CDIP/IIT-5K corpus for industrial documents [7, 24], multilingual collections like Urdu News Dataset 1M [29], Chinese Business Licenses [21], and historical archives such as Quebec Parish Registers [44]. Synthetic or augmented datasets, such as the UIC Code Recognition dataset (50 real + 1,000 generated images) [33] and Inv3D for 3D invoice unwarping [43], demonstrate efforts to address domain-specific challenges. Proprietary or private datasets are also noted, particularly for sensitive applications like medical prescriptions [18] or legal documents [34]. Evaluation frameworks like OCRBench aggregate diverse datasets (e.g., SVT, COCO-Text, DocVQA) to assess multimodal OCR performance [36].

3.5. Future opportunities

The future of OCR looks promising with better machine learning models like CNNs and transformers to improve text recognition [4, 7, 10]. Efforts will focus on fixing layout issues, handling abbreviations, and improving key-value pair extraction [8-9]. Mobile document scanning will get faster with new techniques like parameter pruning [13]. More training data, better annotation tools, and smarter AI models will help OCR work better on different types of documents, including handwritten text and legal papers [12, 18, 42]. Future research will also improve privacy filtering, make OCR easier to use on different layouts, and test it on larger datasets [29, 36, 43]. Competitions with harder datasets will help push OCR even further [31, 38-39].

It should be noted that most of the use cases of the OCR were done involving English language [5-11, 13, 18-24, 27-28, 30, 36-37, 39, 43-44]. Of course, there were studies that used other languages such as Chinese [5,8,21,31,42] but also French, Arabic, Urdu, Hindi etc., but at a lower level. Moreover, from selected studies, there was no study that involved OCR in Albanian language..

Interestingly, majority the main work on the papers was focused in:

- Preprocessing: fixing the image and preparing it to be ready for OCR [4, 19], text from image into Braille [10], pipeline for preprocessing [26]
- Training a model: deep learning model [7,12, 34, 39], own pipeline/framework [5, 17, 20-22], transformer framework [8], for different text extraction [28], printed and handwritten [45], using LSTM [47]
- Using LLM: for better text extraction [25], usage of BERT [46], GPT-4 and Gemini [36].

Recent OCR advancements focus on multimodal frameworks like LayoutLM, which integrates text, layout, and image embeddings [7] and combines token/segment representations [8]. Semantic tools detect unauthorized terms [9] and handle multilingual scripts [28], while NER extracts invoice data [30]. However, LMMs underperform on non-semantic text (e.g., codes) [36], highlighting the need for context-aware systems in such cases.

4. Conclusions

The reviewed literature highlights the critical role of preprocessing in OCR systems, as poor input quality directly compromises analysis accuracy. Much of the research focuses on model development, leveraging neural networks to create customized pipelines for printed and handwritten text. Recent advancements increasingly incorporate large language models (LLMs) for annotation tasks, improving efficiency and accuracy—exemplified by emerging commercial solutions like Mistral [48] but also Google Vision OCR². However, a deeper analysis of these advanced commercial LLMs falls outside the scope of this review, which focused exclusively on open-source models.

We believe that OCR technology holds immense potential to revolutionize data digitization and automation across industries, reducing manual errors in healthcare, finance, and legal sectors while enhancing multilingual accessibility and archival retrieval. However, challenges such as computational costs, data privacy, and dataset limitations must be addressed through sustained research.

To accelerate progress in open-source OCR, developers should enhance modular frameworks like Tesseract with AI-driven upgrades and expanded language support. Curating diverse, openly licensed training datasets will improve robustness across scripts, while community-led benchmarking can ensure scalability for low-resource applications. Embedding ethical guidelines into development practices will further mitigate biases and promote responsible innovation, ensuring OCR's benefits are widely and equitably realized.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

Acknowledgments

This research was supported by the Ministry of Education, Science, Technology and Innovation of Kosovo and HEI'25 project.

References

- [1] *Billing paper market size, demand & forecast 2025-2035. Future Market Insights.* Available at: <https://www.futuremarketinsights.com/reports/billing-paper-market> (Accessed: 22 March 2025).
- [2] *What is optical character recognition (OCR)?, IBM.* Available at: <https://www.ibm.com/think/topics/optical-character-recognition> (Accessed: 22 March 2025).
- [3] Moudgil, Aditi, Saravjeet Singh, and Vinay Gautam. "An overview of recent trends in OCR systems for manuscripts." *Cyber Intelligence and Information Retrieval: Proceedings of CIIR 2021 (2022)*: 525-533.
- [4] M. Auad, S. Alves, G. Kakizaki, J. C. S. Reis, and M. M. Silva, "A Filtering and Image Preparation Approach to Enhance OCR for Fiscal Receipts," in *2024 37th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, Manaus, Brazil: IEEE, Sep. 2024, pp. 1–6. doi: 10.1109/SIBGRAPI62404.2024.10716295.
- [5] Z. Kuang et al., "MMOCR: A Comprehensive Toolbox for Text Detection, Recognition and Understanding," in *Proceedings of the 29th ACM International Conference on Multimedia, Virtual Event China: ACM*, Oct. 2021, pp. 3791–3794. doi: 10.1145/3474085.3478328.

² <https://cloud.google.com/vision/docs/ocr>

- [6] T. T. H. Nguyen, A. Jatowt, M. Coustaty, and A. Doucet, "Survey of Post-OCR Processing Approaches," *ACM Comput. Surv.*, vol. 54, no. 6, pp. 1–37, Jul. 2022, doi: 10.1145/3453476.
- [7] Y. Xu, M. Li, L. Cui, S. Huang, F. Wei, and M. Zhou, "LayoutLM: Pre-training of Text and Layout for Document Image Understanding," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Virtual Event CA USA: ACM, Aug. 2020, pp. 1192–1200. doi: 10.1145/3394486.3403172.
- [8] Y. Li et al., "StrucTexT: Structured Text Understanding with Multi-Modal Transformers," in *Proceedings of the 29th ACM International Conference on Multimedia*, Virtual Event China: ACM, Oct. 2021, pp. 1912–1920. doi: 10.1145/3474085.3475345.
- [9] P. Jain, K. Verma, A. Gaikwad, and P. Gadde, "Understanding Financial Transaction Documents using Natural Language Processing," in *Proceedings of the 10th International Conference on Knowledge Capture*, Marina Del Rey CA USA: ACM, Sep. 2019, pp. 255–258. doi: 10.1145/3360901.3364439.
- [10] A. Ignasius, J. C. Chandra, R. Oscadinata, and D. Suhartono, "Image Pre-Processing Effect on OCR's Performance for Image Conversion to Braille Unicode," *Procedia Computer Science*, vol. 227, pp. 922–931, 2023, doi: 10.1016/j.procs.2023.10.599.
- [11] F. Martínez-Plumed, E. Gómez, and J. Hernández-Orallo, "Futures of artificial intelligence through technology readiness levels," *Telematics and Informatics*, vol. 58, p. 101525, May 2021, doi: 10.1016/j.tele.2020.101525.
- [12] Md. Y. Hossain and T. Rahman, "A crowdsourcing based framework for Bengali scene text data collection and detection," *Computers and Electrical Engineering*, vol. 112, p. 109025, Dec. 2023, doi: 10.1016/j.compeleceng.2023.109025.
- [13] L. Liu, Z. Wang, T. Qiu, Q. Chen, Y. Lu, and C. Y. Suen, "Document image classification: Progress over two decades," *Neurocomputing*, vol. 453, pp. 223–240, Sep. 2021, doi: 10.1016/j.neucom.2021.04.114.
- [14] P. A. Iglesias, C. Ochoa, and M. Revilla, "A practical guide to (successfully) collect and process images through online surveys," *Social Sciences & Humanities Open*, vol. 9, p. 100792, 2024, doi: 10.1016/j.ssaho.2023.100792.
- [15] G. Nagy, "Reflections of an ancient document processor," *Pattern Recognition Letters*, vol. 166, pp. 76–79, Feb. 2023, doi: 10.1016/j.patrec.2023.01.006.
- [16] F. Alharbi, R. Alshahrani, M. Zakariah, A. Aldweesh, and A. A. Alghamdi, "YOLO and Blockchain Technology Applied to Intelligent Transportation License Plate Character Recognition for Security," *CMC*, vol. 77, no. 3, pp. 3697–3722, 2023, doi: 10.32604/cmc.2023.040086.
- [17] A. C. Seker and S. C. Ahn, "A generalized framework for recognition of expiration dates on product packages using fully convolutional networks," *Expert Systems with Applications*, vol. 203, p. 117310, Oct. 2022, doi: 10.1016/j.eswa.2022.117310.
- [18] M. Gupta and K. Soeny, "Algorithms for rapid digitalization of prescriptions," *Visual Informatics*, vol. 5, no. 3, pp. 54–69, Sep. 2021, doi: 10.1016/j.visinf.2021.07.002.
- [19] S. Correia and S. Luck, "Digitizing historical balance sheet data: A practitioner's guide," *Explorations in Economic History*, vol. 87, p. 101475, Jan. 2023, doi: 10.1016/j.eeh.2022.101475.
- [20] W. R. Thomas et al., "Petabytes in Practice: Working with Collections as Data at Scale," *Data and Information Management*, vol. 3, no. 1, pp. 18–25, Mar. 2019, doi: 10.2478/dim-2019-0004.
- [21] P. Cao and J. Wu, "GraphRevisedIE: Multimodal information extraction with graph-revised network," *Pattern Recognition*, vol. 140, p. 109542, Aug. 2023, doi: 10.1016/j.patcog.2023.109542.
- [22] H. T. Ha and A. Horák, "Information extraction from scanned invoice images using text analysis and layout features," *Signal Processing: Image Communication*, vol. 102, p. 116601, Mar. 2022, doi: 10.1016/j.image.2021.116601.
- [23] A. Hemmer, M. Coustaty, N. Bartolo, and J.-M. Ogier, "Confidence-Aware Document OCR Error Detection," in *Document Analysis Systems*, vol. 14994, G. Sfikas and G. Retsinas, Eds., Cham: Springer Nature Switzerland, 2024, pp. 213–228. doi: 10.1007/978-3-031-70442-0_13.

- [24] A. F. Biten, R. Tito, L. Gomez, E. Valveny, and D. Karatzas, "OCR-IDL: OCR Annotations for Industry Document Library Dataset," in *Computer Vision – ECCV 2022 Workshops*, vol. 13804, L. Karlinsky, T. Michaeli, and K. Nishino, Eds., Cham: Springer Nature Switzerland, 2023, pp. 241–252. doi: 10.1007/978-3-031-25069-9_16.
- [25] M. Lamott, Y.-N. Weweler, A. Ulges, F. Shafait, D. Krechel, and D. Obradovic, "LAPDoc: Layout-Aware Prompting for Documents," in *Document Analysis and Recognition - ICDAR 2024*, vol. 14807, E. H. Barney Smith, M. Liwicki, and L. Peng, Eds., Cham: Springer Nature Switzerland, 2024, pp. 142–159. doi: 10.1007/978-3-031-70546-5_9.
- [26] A. Randika, N. Ray, X. Xiao, and A. Latimer, "Unknown-Box Approximation to Improve Optical Character Recognition Performance," in *Document Analysis and Recognition – ICDAR 2021*, J. Lladós, D. Lopresti, and S. Uchida, Eds., Cham: Springer International Publishing, 2021, pp. 481–496. doi: 10.1007/978-3-030-86549-8_31.
- [27] Q. M. Tan, Q. Cao, C. K. Seow, and P. C. Yau, "Information Extraction System for Invoices and Receipts," in *Advanced Intelligent Computing Technology and Applications*, vol. 14089, D.-S. Huang, P. Premaratne, B. Jin, B. Qu, K.-H. Jo, and A. Hussain, Eds., Singapore: Springer Nature Singapore, 2023, pp. 77–89. doi: 10.1007/978-981-99-4752-2_7.
- [28] M. Maithani, D. Meher, and S. Gupta, "Multilingual Text Recognition System," in *Advances in Signal Processing, Embedded Systems and IoT*, vol. 992, V. V. S. S. Chakravarthy, V. Bhateja, W. Flores Fuentes, J. Anguera, and K. P. Vasavi, Eds., Singapore: Springer Nature Singapore, 2023, pp. 103–114. doi: 10.1007/978-981-19-8865-3_9.
- [29] A. Maqsood, N. Riaz, A. Ul-Hasan, and F. Shafait, "A Unified Architecture for Urdu Printed and Handwritten Text Recognition," in *Document Analysis and Recognition - ICDAR 2023*, vol. 14190, G. A. Fink, R. Jain, K. Kise, and R. Zanibbi, Eds., Cham: Springer Nature Switzerland, 2023, pp. 116–130. doi: 10.1007/978-3-031-41685-9_8.
- [30] A. Hamdi, E. Carel, A. Joseph, M. Coustaty, and A. Doucet, "Information Extraction from Invoices," in *Document Analysis and Recognition – ICDAR 2021*, vol. 12822, J. Lladós, D. Lopresti, and S. Uchida, Eds., Cham: Springer International Publishing, 2021, pp. 699–714. doi: 10.1007/978-3-030-86331-9_45.
- [31] W. Yu et al., "ICDAR 2023 Competition on Reading the Seal Title," in *Document Analysis and Recognition - ICDAR 2023*, vol. 14188, G. A. Fink, R. Jain, K. Kise, and R. Zanibbi, Eds., Cham: Springer Nature Switzerland, 2023, pp. 522–535. doi: 10.1007/978-3-031-41679-8_31.
- [32] B. Akshaya and M. Rajendiran, "Automatic Inspection Verification Using Digital Certificate," in *Emerging Trends in Computing and Expert Technology*, vol. 35, D. J. Hemanth, V. D. A. Kumar, S. Malathi, O. Castillo, and B. Patrut, Eds., Cham: Springer International Publishing, 2020, pp. 826–837. doi: 10.1007/978-3-030-32150-5_83.
- [33] R. Marmo, "UIC Code Recognition Using Computer Vision and LSTM Networks," in *Dependable Computing - EDCC 2020 Workshops*, vol. 1279, S. Bernardi, V. Vittorini, F. Flammini, R. Nardone, S. Marrone, R. Adler, D. Schneider, P. Schleiß, N. Nostro, R. Løvenstein Olsen, A. Di Salle, and P. Masci, Eds., Cham: Springer International Publishing, 2020, pp. 90–98. doi: 10.1007/978-3-030-58462-7_8.
- [34] D. L. Freire et al., "Lawsuits Document Images Processing Classification," in *Progress in Artificial Intelligence*, vol. 13566, G. Marreiros, B. Martins, A. Paiva, B. Ribeiro, and A. Sardinha, Eds., Cham: Springer International Publishing, 2022, pp. 41–52. doi: 10.1007/978-3-031-16474-3_4.
- [35] K. O. M. Bogdan, G. A. S. Megeto, R. Leal, G. Souza, A. C. Valente, and L. N. Kirsten, "DDocE: Deep Document Enhancement with Multi-scale Feature Aggregation and Pixel-Wise Adjustments," in *Document Analysis and Recognition – ICDAR 2021 Workshops*, vol. 12916, E. H. Barney Smith and U. Pal, Eds., Cham: Springer International Publishing, 2021, pp. 229–244. doi: 10.1007/978-3-030-86198-8_17.
- [36] Y. Liu et al., "OCRBench: on the hidden mystery of OCR in large multimodal models," *Sci. China Inf. Sci.*, vol. 67, no. 12, p. 220102, Dec. 2024, doi: 10.1007/s11432-024-4235-6.

- [37] T. Hegghammer, "OCR with Tesseract, Amazon Textract, and Google Document AI: a benchmarking experiment," *J Comput Soc Sc*, vol. 5, no. 1, pp. 861–882, May 2022, doi: 10.1007/s42001-021-00149-1.
- [38] H. Feng et al., "DocPedia: unleashing the power of large multimodal model in the frequency domain for versatile document understanding," *Sci. China Inf. Sci.*, vol. 67, no. 12, p. 220106, Dec. 2024, doi: 10.1007/s11432-024-4250-y.
- [39] G. K. Kumar, R. R. Kumar, R. Chakka, and P. Viswanath, "A multi-pronged accurate approach to optical character recognition, using nearest neighborhood and neural-network-based principles," *Sādhana*, vol. 46, no. 4, p. 189, Dec. 2021, doi: 10.1007/s12046-021-01703-3.
- [40] A. Kabiraj, D. Pal, D. Ganguly, K. Chatterjee, and S. Roy, "Number plate recognition from enhanced super-resolution using generative adversarial network," *Multimed Tools Appl*, vol. 82, no. 9, pp. 13837–13853, Apr. 2023, doi: 10.1007/s11042-022-14018-0.
- [41] S. Kul, S. Kumcu, and A. Sayar, "Docker Container-Based Framework of Apache Kafka Node Ecosystem: Vehicle Tracking System by License Plate Recognition on Surveillance Camera Feeds," *Int. J. ITS Res.*, vol. 22, no. 2, pp. 290–297, Aug. 2024, doi: 10.1007/s13177-024-00392-6.
- [42] J. Song, H. Fu, T. Jiao, and D. Wang, "AI-enabled legacy data integration with privacy protection: a case study on regional cloud arbitration court," *J Cloud Comp*, vol. 12, no. 1, p. 145, Oct. 2023, doi: 10.1186/s13677-023-00500-z.
- [43] F. Hertlein, A. Naumann, and P. Philipp, "Inv3D: a high-resolution 3D invoice dataset for template-guided single-image document unwarping," *IJDAR*, vol. 26, no. 3, pp. 175–186, Sep. 2023, doi: 10.1007/s10032-023-00434-x.
- [44] S. Tarride et al., "Large-scale genealogical information extraction from handwritten Quebec parish records," *IJDAR*, vol. 26, no. 3, pp. 255–272, Sep. 2023, doi: 10.1007/s10032-023-00427-w.
- [45] D. S. Jabonete and M. M. De Leon, "Development of an Automatic Document to Digital Record Association Feature for a Cloud-Based Accounting Information System," in *Intelligent Computing*, vol. 283, K. Arai, Ed., Cham: Springer International Publishing, 2022, pp. 899–910. doi: 10.1007/978-3-030-80119-9_59.
- [46] R. Lima, S. Paiva, and J. Ribeiro, "Artificial Intelligence Optimization Strategies for Invoice Management: A Preliminary Study," in *Communication and Intelligent Systems*, vol. 204, H. Sharma, M. K. Gupta, G. S. Tomar, and W. Lipo, Eds., Singapore: Springer Singapore, 2021, pp. 223–234. doi: 10.1007/978-981-16-1089-9_19.
- [47] C. Thorat, A. Bhat, P. Sawant, I. Bartakke, and S. Shirsath, "A Detailed Review on Text Extraction Using Optical Character Recognition," in *ICT Analysis and Applications*, vol. 314, S. Fong, N. Dey, and A. Joshi, Eds., Singapore: Springer Nature Singapore, 2022, pp. 719–728. doi: 10.1007/978-981-16-5655-2_69.
- [48] Mistral OCR. Mistral AI. Available at: <https://mistral.ai/news/mistral-ocr> (Accessed: 22 March 2025).