Advancements and Challenges in Arabic Optical Character Recognition: A Comprehensive Survey

MAHMOUD SALAHELDIN KASEM, Department of Information and Communication Engineering, Chungbuk National University, Cheongju, Korea (the Republic of) and Multimedia, Assiut University, Asyut, Egypt
MOHAMED MAHMOUD, Faculty of Computer and Information, Assiut University, Assiut, Egypt and College of Electrical and Computer Engineering, Chungbuk National University, Cheongju, Korea (the Republic of)
HYUN-SOO KANG*, Department of Information and Communication Engineering, Chungbuk National University, Cheongju, Korea (the Republic of)

Optical character recognition (OCR) is a vital process that involves the extraction of handwritten or printed text from scanned or printed images, converting it into a format that can be understood and processed by machines. The automatic extraction of text through OCR plays a crucial role in digitizing documents, enhancing productivity, and preserving historical records. This paper offers an exhaustive review of contemporary applications, methodologies, and challenges associated with Arabic OCR. A thorough analysis is conducted on prevailing techniques utilized throughout the OCR process, with a dedicated effort to discern the most efficacious approaches that demonstrate enhanced outcomes. To ensure a thorough evaluation, a meticulous keyword-search methodology is adopted, encompassing a comprehensive analysis of articles relevant to Arabic OCR. In addition to presenting cutting-edge techniques and methods, this paper identifies research gaps within the realm of Arabic OCR. We shed light on potential areas for future exploration and development, thereby guiding researchers toward promising avenues in the field of Arabic OCR. The outcomes of this study provide valuable insights for researchers, practitioners, and stakeholders involved in Arabic OCR, ultimately fostering advancements in the field and facilitating the creation of more accurate and efficient OCR systems for the Arabic language.

CCS Concepts: • **Computing methodologies** → **Computer vision**.

Additional Key Words and Phrases: data processing, segmentation, classification, deep learning

1 Introduction

Optical Character Recognition (OCR) is a transformative technology designed to convert printed or handwritten text into machine-readable formats. While OCR systems for various languages have achieved remarkable progress, Arabic OCR presents unique challenges due to the script's cursive nature, contextual variations in letter shapes, diacritical marks, and rich morphological structure. These complexities demand specialized approaches to achieve high accuracy and reliability in text recognition[16, 89].

*Corresponding author

Authors' Contact Information: Mahmoud SalahEldin Kasem, Department of Information and Communication Engineering, Chungbuk National University, Cheongju, Chungcheongbuk-do, Korea (the Republic of) and Multimedia, Assiut University, Asyut, Assiut Governorate, Egypt; e-mail: mahmoud.salah@aun.edu.eg; Mohamed Mahmoud, Faculty of Computer and Information, Assiut University, Assiut, Egypt and College of Electrical and Computer Engineering, Chungbuk National University, Cheongju, Chungcheongbuk-do, Korea (the Republic of); e-mail: mohamedabokhalil@aun.edu.eg; Hyun-Soo Kang, Department of Information and Communication Engineering, Chungbuk National University, Cheongju, Chungcheongbuk-do, Korea (the Republic of); e-mail: hskang@cbnu.ac.kr.

ACM 1557-7341/2025/9-ART
https://doi.org/10.1145/3768150

Arabic OCR plays a crucial role in enabling the digitization and preservation of Arabic textual data, a task vital for cultural heritage, modern communication, and information accessibility for over 400 million Arabic speakers worldwide. The integration of Arabic OCR into applications such as document digitization, text-to-speech systems, and automated translation has the potential to bridge linguistic and technological divides. However, the cursive and highly contextual nature of Arabic script, combined with the lack of comprehensive datasets, poses substantial technical challenges.

This paper provides a comprehensive exploration of Arabic OCR, emphasizing the distinctive challenges of processing Arabic script and the latest advancements in addressing these issues. It delves into the intricacies of preprocessing, segmentation, recognition, and postprocessing techniques, analyzing state-of-the-art methodologies and their effectiveness in improving performance. By systematically reviewing current methodologies, identifying research gaps, and analyzing key datasets, this study aims to enhance the development of robust Arabic OCR systems. the study balances technical depth with clarity, ensuring that foundational concepts are explained in straightforward terms while delving into advanced topics. Illustrative examples, diagrams, and performance comparisons are incorporated to guide readers progressively through the intricacies of Arabic OCR. The insights from this research contribute to advancing applications such as document digitization, language preservation, and improved accessibility to Arabic textual data, fostering innovation and collaboration across academic and industrial domains.

1.1 Challenges in Arabic OCR

Arabic OCR presents several unique challenges that distinguish it from OCR for other languages. This section explores these challenges in detail, highlighting the complexities associated with Arabic text and the implications for OCR accuracy and performance. Understanding these challenges is crucial for developing effective OCR solutions tailored to the Arabic language.

- Complex Morphology: Arabic features a rich morphological structure, characterized by root-based word formation, extensive use of diacritical marks (i.e., vowel markers and other orthographic symbols), and various forms of ligatures and connecting letters. These complexities pose challenges for segmentation, character recognition, and word-level analysis during the OCR process.
- Contextual Variations: Arabic text exhibits contextual variations that impact character shapes, particularly due to the presence of initial, medial, final, and isolated forms of letters. Accurately recognizing and disambiguating these contextual variants is essential for maintaining OCR accuracy.
- Cursive Writing Style: Arabic is commonly written in a cursive style, where letters are connected, resulting in overlapping strokes. This cursive nature makes it challenging to separate individual characters accurately, affecting character segmentation and recognition algorithms.
- Diacritic Marks: Arabic utilizes diacritical marks to indicate vowels and other phonetic information. However, diacritical marks are often omitted in everyday writing or handwriting, making it difficult to accurately reconstruct the original text during OCR.
- Ligatures and Shaping: Arabic includes ligatures and contextual shaping, where letters change their shapes based on their position within a word. Proper recognition and interpretation of ligatures and shaping are critical for accurate OCR output.
- Limited Availability of Labeled Datasets: Building robust Arabic OCR systems requires large, high-quality, and diverse labeled datasets. However, the availability of such datasets for training and evaluation purposes is limited compared to other languages, which poses challenges for developing and benchmarking Arabic OCR algorithms.
- Arabic Dialects and Variations: Arabic is spoken across different regions, leading to dialectal variations in written form. OCR systems must account for these variations to ensure accurate recognition and understanding of Arabic text from different dialectical sources.

The unique characteristics of Arabic script, including contextual variations, cursive writing styles, diacritic marks, and ligatures, present significant challenges for OCR systems. The primary and secondary character categorizations as shown in Tables 1 and 2 emphasize how character shapes vary based on position and adjacent letters, while special characters add further complexity. Addressing these challenges is essential for developing robust OCR systems capable of processing diverse Arabic texts accurately and effectively.

In the subsequent sections of this paper, we will explore the state-of-the-art techniques and methodologies employed to tackle these challenges and improve the performance of Arabic OCR systems. By addressing these challenges head-on, researchers aim to develop more robust, accurate, and efficient OCR solutions tailored specifically to the complexities of Arabic language processing.

Arabic Letter	Start	Middle	End	Alone
Hamza				ء
Alif			ـا	ا
Baa	بـ	ـبـ	ـب	ب
Ta	تـ	ـتـ	ـت	ت
Tha	ثـ	ـثـ	ـث	ث
Jim	جـ	ـجـ	ـج	ج
HHA	حـ	ـحـ	ـح	ح
KHA	خـ	ـخـ	ـخ	خ
Dal	د	ـد	د	د
Thal	ذ	ـذ	ذ	ذ
Ra	ر	ـر	ر	ر
Zai	ز	ـز	ز	ز
Sin	سـ	ـسـ	ـس	س
Shin	شـ	ـشـ	ـش	ش
Sad	صـ	ـصـ	ـص	ص

Arabic Letter	Start	Middle	End	Alone
Dhad	ضـ	ـضـ	ـض	ض
Tua	طـ	ـطـ	ـط	ط
Zua	ظـ	ـظـ	ـظ	ظ
Ain	عـ	ـعـ	ـع	ع
Gain	غـ	ـغـ	ـغ	غ
Fa	فـ	ـفـ	ـف	ف
Qaf	قـ	ـقـ	ـق	ق
Kaf	كـ	ـكـ	ـك	ك
Lam	لـ	ـلـ	ـل	ل
Mim	مـ	ـمـ	ـم	م
Non	نـ	ـنـ	ـن	ن
Ha	هـ	ـهـ	ـه	ه
Waw	ـو	ـو	ـو	و
Ya	يـ	ـيـ	ـي	ي

Table 1. Arabic Letters and Their Forms in Different Positions

Character Name	Middle	End	Isolated
Alif-Hamzah	-	ـأ	أ
Alif-Hamzah	-	ـإ	إ
Alif-MAqsorah	-	ـى	ى
Alif-Mad	-	ـآ	آ
Ta-Marbotah	-	ـة	ة
Waw-Hamzah	-	ـؤ	ؤ
Ya-Hamzah	ـئـ	ـئ	ئ

Character Name	End	Isolated
Lam-Alif	ـلا	لا
Lam-Alif	ـلأ	لأ
Lam-Alif	ـلإ	لإ
Lam-Alif	ـلآ	لآ

Table 2. Special and Secondary Character Sets: Addressing the Challenges of Arabic OCR. The secondary category includes characters with distinct forms depending on their position (middle, end, isolated), such as Hamzah and Ya-Hamzah. The special category involves ligatures like Lam-Alif, which are particularly challenging due to their unique shapes. These complexities highlight the need for advanced OCR techniques to handle the variability of Arabic script.

1.2 Classification of OCR Systems

Various classifications of OCR systems are contingent upon the language and writing style found in the images. Documents may incorporate handwritten, printed, or scanned content and may involve one or multiple languages. Consequently, OCR systems can be categorized as unilingual or multilingual, predominantly based on language support. A unilingual OCR system is specifically designed to identify a single language, with the Arabic OCR model serving as an illustration of such a system. Conversely, certain OCR systems possess the capability to perform recognition and extraction tasks across multiple languages, earning them the designation of multilingual OCR systems.

Optical Character Recognition (OCR) systems can be broadly classified into two categories: offline OCR systems and online OCR systems. Offline OCR systems are designed to process scanned, printed, and handwritten documents [47]. These systems provide a spectrum of online services tailored to diverse applications, encompassing functionalities such as sorting mails, reading the bank cheques, verification of signatures, processing of utility bills, and applications within the insurance sector. Moreover, offline OCR systems play a crucial role in enhancing accessibility for blind or illiterate individuals by utilizing digital pens to convert text into audio format. In addition, offline recognition systems find extensive implementation in diverse domains such as number-plate recognition[62].

On the other hand, online OCR systems are specifically designed to receive and process real-time input images. These systems are capable of handling dynamic, on-the-fly recognition tasks. In contrast, offline recognition in OCR often involves the use of multiple models with different datasets and algorithms to achieve higher levels of recognition accuracy. By employing a variety of models and datasets, offline OCR systems aim to optimize performance and enhance the overall accuracy of text recognition[94].

1.3 Applications

OCR systems have become increasingly prevalent across various industries, offering expedited and precise workflows that are instrumental in digitization processes. Singh [100] provides a comprehensive survey on OCR applications and experiments with select use cases. Prominent applications include invoice imaging for efficient tracking of business records, banking services where OCR enables rapid processing of checks, and security systems like Captcha, which challenge computer programs with distorted alphanumeric images. In the legal industry, OCR streamlines document digitization, enhancing operational efficiency. Surveillance systems also utilize OCR for automatic number recognition, capturing and analyzing vehicle license plates with accuracy.

Additionally, OCR excels in handwriting recognition by learning diverse fonts and languages, boosting recognition accuracy in handwritten documents. Extracting data from scanned receipts poses challenges like layout variations and noise, as noted by Antonio [25], but OCR effectively addresses these issues for efficient data extraction. In healthcare, OCR facilitates the processing of extensive patient data, including forms and reports, demonstrating its critical role in enhancing efficiency across various sectors.

1.4 Procedural Overview of Arabic OCR Systems

Arabic OCR systems encompass a series of intricate and well-defined procedures, as depicted in Figure 1, which provides a succinct overview of the specific steps that must be followed. The initial phase involves the preprocessing of the image to enhance its quality and optimize the recognition process. The preprocessing phase involves a series of operations, including skewing correction, reduction of noise, and enhancement of the contrast. Subsequently, the area contains text in the image undergoes meticulous segmentation into individual units, comprising characters or Texts. These characters that are segmented are subjected to recognition, with the most suitable counterpart determined by utilizing a comprehensive dataset of characters known. The text that is recognized subsequently progresses through additional processing stages, focusing on error correction and overall accuracy improvement. Ultimately, the output materializes as a machine-readable text document, granting software applications the capability to seamlessly edit, search, and analyze the content.

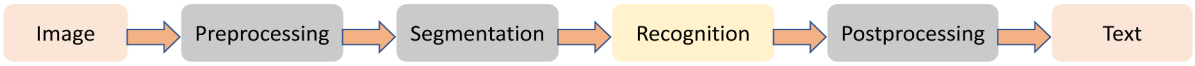


Fig. 1. Brief overview of OCR process

1.5 Goals and Outlines

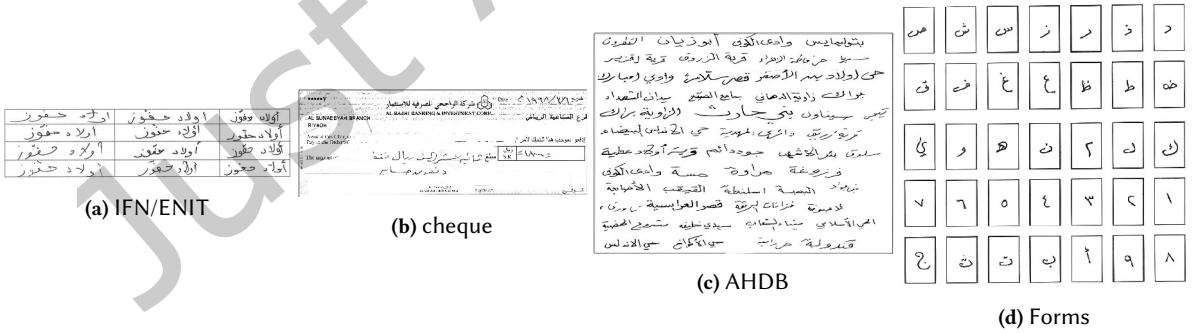
This paper provides a detailed exploration of OCR with a specific focus on Arabic language processing. Arabic, as a complex and highly contextual language, presents unique challenges that necessitate specialized OCR techniques. Understanding and accurately processing Arabic text is vital for a wide range of applications, including document digitization, text analysis, information retrieval, and language preservation.

This paper aims to provide a structured review of modern Arabic OCR, highlighting state-of-the-art methods, emerging applications, and persisting challenges. By systematically assessing existing approaches and methodologies, we emphasize the most effective techniques that contribute to improved OCR performance. Additionally, the paper addresses current research gaps, offering insight into potential future directions for the development of advanced Arabic OCR systems. The findings of this study serve as a valuable resource for researchers, practitioners, and stakeholders in Arabic OCR, guiding them towards promising avenues and fostering the advancement of more accurate and efficient OCR systems for the Arabic language.

The following sections of this paper are organized as follows: Section 2 offers a detailed analysis of the datasets available for assessing Arabic OCR systems.In Section 3, We offer a thorough examination of the current literature on every stage of Arabic OCR, highlighting significant research trends and advancements. In conclusion, we summarize the key findings from this survey, emphasizing research gaps.

2 Databases

Dataset plays a pivotal role in the validation of OCR systems, serving as a critical component to assess the accuracy of OCR results. Particularly in the case of the Arabic language, several challenges arise due to its cursive nature, the presence of diacritics, varying writing styles that alter the overall shape of each word, fluctuations in the size of the text, and related factors. Additionally, a compilation of an Arabic database is hindered by the tied availability of resources associated with the Arabic language. In prior works [5, 10], researchers have shared commonly utilized datasets, as showcased in Table 3, encompassing Arabic, Urdu, and Persian languages, including both publicly accessible and restricted datasets. Also, Figure 2,3,4,5 illustrates examples of each dataset.



(a) IFN/ENIT **(b)** cheque **(c)** AHDB **(d)** Forms

Fig. 2. Examples of images in IFN/ENIT, Cheque, AHDB, and Forms.

2.1 Handwritten Text

Arabic and Urdu exhibit numerous symmetrics in terms of their writing styles and cursive nature. Both languages adhere to the right-to-left writing direction, while Urdu comprises approximately 39 to 40 letters, with Arabic

Table 3. comprehensive list of available datasets, accompanied by their corresponding statistics, dataset types, and modes of accessibility

Dataset	Size	Content	Accessibility	Venue
IFN/ENIT [90]	115,000 words & 212,000 characters	Handwritten words	Public	CIFED 2002
AHDB[8]	30,000 words	Handwritten words & digits	Private	IEEE 2002
Cheque[9]	29,498 subwords & 15,148 digits	Handwritten subwords & digits	Private	Pattern Recognit. 2003
Forms[26]	15,800 characters & 500 writers	Handwritten characters	Private	IEC (Prague) 2005
UPTI[96]	10,000 lines	Printed text lines	Public	Document recognition and retrieval XX 2013
Numeral[27]	21,120 digits & 44 writers	Handwritten digits	Public	Signal Processing 2009
HACDB[67]	6600 characters & 50 writers	Handwritten characters	Public	EUVIP 2013
AHDBase[50]	70,000 digits & 700 writers	Handwritten digits	Public	Artificial intelligence and pattern recognition 2007
HODA[65]	102,352 digits	Handwritten digits	Public	Pattern recognition letters 2007
APTI[102]	113,284 words & 648,280 characters	Printed words	Public	DIUF 2009
KHATT[70]	9327 lines, 165,890 words & 589,924 characters	Handwritten text lines	Public	Pattern Recognit., 2014
ACTIV[109]	4824 lines & 21,520 words	Embedded text lines	Public	ICDAR 2015
ALIF[108]	1804 words & 89,819 characters	Embedded text lines	Needs request	ICDAR 2015
Digital Jawi[98]	168 words & 1524 characters	Jawi paleography images	Public	IEEE(ICIIP) 2015
SmartATID[45]	9088 pages	Printed & handwritten pages	Public	ICFHR 2016
Printed PAW[33]	415,280 unique words & 550,000 sub words	Printed subwords	Needs request	Journal of ICT, 2017
Degraded historical[103]	10 handwritten images & 10 printed images	Handwritten documents	Public	IEEE(ICEEI), 2017
ACTIV2[110]	10,415 text images	Embedded words	Public	J. Imaging, 2018
QTID[29]	309,720 words & 249,428 characters	Synthetic words	Private	Int. Journal of Adv. CS App. 2018
KAFD[68]	28,767 pages & 644,006 lines	Printed pages & lines	Public	Pattern Recognit. 2014
AHDB/FTR[93]	497 images	Handwritten Text Images	Public	Procedia Technology, 2013
ADBase & MADBase [1]	70,000 digits & 700 writers	Handwritten Digits	Public	datacenter.aucegypt.edu/shazeem
AHT2D [87][2]	-	Handwritten Text	paywall	Multim. Tools Appl. 2023
AHWD[24]	21,357 words	Handwritten words	Private	CCECE 2022

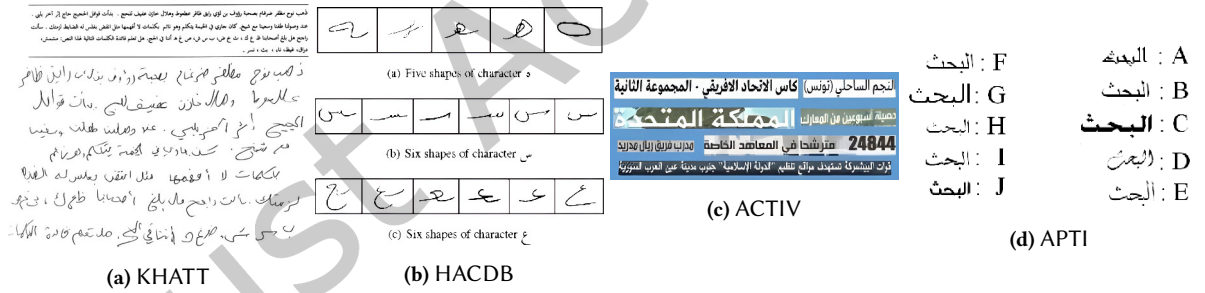


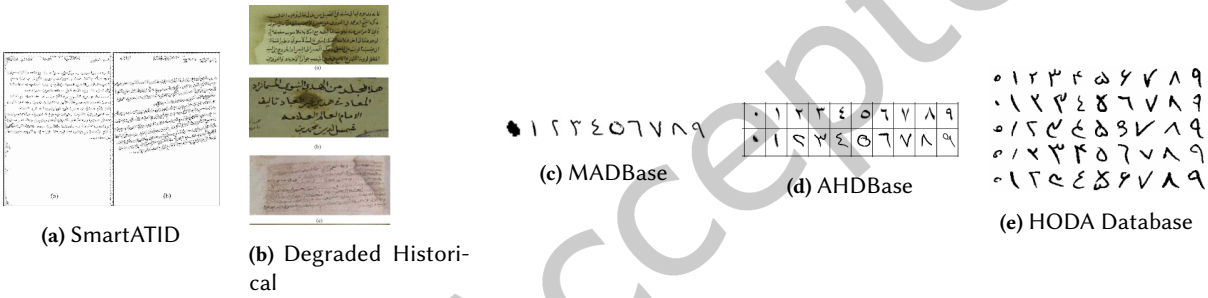
Fig. 3. Examples of images in KHATT, HACDB, ACTIV, and APTI.

possessing a slightly smaller character set. Moreover, Urdu extensively incorporates vocabulary borrowed from Arabic, accounting for nearly 30% of its lexicon. Given these commonalities, datasets and trained models developed for either language are often utilized interchangeably due to their shared characteristics. This cross-usage of resources is particularly beneficial, as it enables leveraging existing data and models, leading to synergistic advancements in both Urdu and Arabic OCR domains.

Datasets for Urdu and Arabic languages are available, showcasing the efforts of researchers in compiling valuable resources. Notably, in [36], a dataset featuring handwritten Urdu numerals was presented. Additionally, [5] introduced the Urdu Nastaliq Handwritten Dataset (UNHD), consisting of handwritten samples contributed

(a) Digital Jawi (b) AHDB/FTR (c) QTID (d) ACTIV2 (e) ALIF (f) Numeral

Fig. 4. Examples of images in Digital Jawi, AHDB/FTR, QTID, ACTIV2, ALIF, and Numeral.



(a) SmartATID (b) Degraded Historical (c) MADBase (d) AHDBase (e) HODA Database

Fig. 5. Examples of images in SmartATID, Degraded Historical, MADBase, AHDBase, and HODA Database.

by 500 writers on A4-size paper. Furthermore, Khosrobeigi et al. [66] presented a dataset for the Persian language, which was compiled from various Persian-language news websites, offering a valuable resource for research and development in the field.

Alghamdi and Teahan [13] extensively discuss the prevalent datasets commonly employed for the training and assessment of OCR systems focusing on printed form of the Arabic script. They highlight notable datasets, which incorporates datasets such as the Arabic handwritten dataset IFN/ENIT and the "Handwriting Arabic Corpus", and RIMES dataset, encompassing an extensive array of both printed and handwritten documents, this compilation represents a significant resource. The authors deliver a comprehensive overview of these accessible datasets, underscoring the paramount importance of employing high-quality datasets to augment the precision of OCR systems.

2.2 Printed Arabic

Printed Arabic OCR deals with challenges unique to the language, such as its cursive nature and diverse font styles. Several studies have tackled these challenges by introducing novel approaches to segmentation, feature extraction, and recognition pipelines. The complexity arises from the varying layouts and high interconnection of characters. H Bouressace[42] introduces a range of methodologies like bottom-up, top-down, and crossbred

[1]https://datacenter.aucegypt.edu/shazeem
[2]https://ieee-dataport.org/documents/aht2d-dataset#files/

approaches. This study extensively examines the various phases involved in the OCR pipeline, including preprocessing, segmentation, feature extraction, and classification. Also, EA El-Sherif[50] presents the introduction of a comprehensive Arabic Handwritten Digits Database (AHDBase), encompassing 60,000 digits for training and 10,000 digits for testing, contributed by 700 individuals with diverse demographic characteristics. Additionally, the authors propose a recognition system tailored for Arabic handwritten digits. The recognition system comprises two stages. In the initial stage, an Artificial Neural Network (ANN) is deployed, utilizing a concise yet potent feature vector for swift classification of non-ambiguous cases. Notably, the first stage incorporates a reject option to channel ambiguous cases to the second stage. The subsequent stage employs a Support Vector Machine (SVM), characterized by a slower yet more powerful nature. This stage utilizes a larger feature vector to effectively classify cases rejected by the first stage, particularly those with greater complexity.

2.3 Scanned Documents

Extracting information from scanned documents poses challenges due to their rough layout and low resolution compared to regular documents. Successfully preprocessing the scanned documents is crucial for accurate information extraction, as it plays a significant role in this context. An example of an initiative addressing this challenge is the competition organized by ICDAR Z Huang[61]. This competition focuses on extracting information from 1000 scanned receipts and encompasses operations like text identification, arrangement examination, and data retrieval. By participating in such competitions, researchers and practitioners aim to advance the field of information extraction from scanned documents.

2.4 Quranic Text

The Qur'an, a foundational text in Arabic, introduces additional complexity due to its scriptural conventions, linguistic variations, and diacritics. This domain has seen notable advancements in computational linguistics and NLP. MH Bashir [31] emphasized the role of computational models in tasks like morphological analysis and Quranic recitation correction. Malhas tackled the lack of datasets by creating QRCD, enabling machine reading comprehension for Quranic text. Their introduction of CL-AraBERT highlighted cross-lingual transfer learning's potential in adapting models for Classical Arabic.

3 Optical Character Recognition (OCR) Processes in Arabic

OCR, or Optical Character Recognition, is a complex process that transforms text characters from printed or handwritten sources into machine-encoded text. This involves several stages: preprocessing, segmentation, recognition, and postprocessing.

In preprocessing, the input image undergoes cleanup and enhancement to optimize recognition quality. Segmentation divides the image into characters, and feature extraction identifies characteristics like shape and size. Recognition classifies characters by comparing them to known ones, and postprocessing corrects errors for refined results. The accuracy of OCR is influenced by factors like image quality, font type, size, and language, leading to varying accuracy levels in different scenarios. As shown in Figure 6

3.1 Preprocessing

The accuracy of OCR models can be compromised by formatting issues in images. Challenges like image orientation or color correction problems can notably hinder model performance. To overcome these issues and improve OCR model accuracy during training, image preprocessing techniques are frequently utilized. These techniques are pivotal in optimizing OCR model performance by tackling formatting issues. They include procedures like resizing, grayscale conversion, skew correction, and enhancing image resolution. Implementing these preprocessing techniques improves the quality and suitability of images for OCR training, resulting in enhanced accuracy in text recognition.

AP Tafti[104] conducted detailed evaluations using popular OCR services: Google Docs OCR, Tesseract, ABBYY FineReader, and Transym. Basic image preprocessing, like grayscale conversion and brightness/contrast

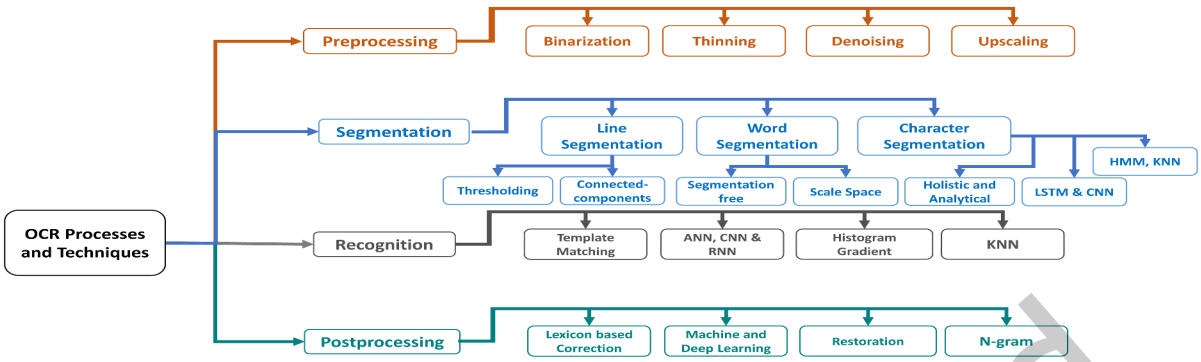


Fig. 6. OCR Processes and Techniques

adjustments, enhanced recognition accuracy by up to 9%. Illumination adjustments, emphasizing object sharpness and contour clarity, further improved results.

M Eltay[54] introduced an adaptive data augmentation method utilizing Generative Adversarial Networks (GANs) to address class imbalance in text recognition tasks. Specifically focusing on handwritten Arabic text recognition, the research presented experimental findings from two public datasets. The study demonstrated the effectiveness of GANs trained with this technique in managing class imbalances. IB Mustapha[82] introduced a method called Conditional Deep Convolutional Generative Adversarial Networks (CDCGAN) designed for generating isolated handwritten Arabic characters. Their experiments, both qualitative and quantitative, demonstrated that CDCGAN effectively produces synthetic handwritten Arabic characters.

S Majumdar[71] tackled the intricate challenge of attributing digitized handwriting in a document to multiple scribes, a task characterized by high dimensionality. The dissimilarity in unique handwriting styles arises from a combination of factors, including character size, stroke width, loops, ductus, slant angles, and cursive ligatures. The authors achieved successful hand shift detection in a historical manuscript by employing fuzzy soft clustering in conjunction with linear principal component analysis. This notable advancement highlights the effective application of unsupervised methods in the realms of writer attribution for historical documents and forensic document analysis.

3.2 Segmentation

Segmentation is a pivotal step in Optical Character Recognition (OCR), involving the intricate division of an image into its constituent parts—lines, words, and characters—to facilitate accurate recognition. In Arabic OCR, segmentation presents unique challenges due to the cursive nature of the language and the absence of clear boundaries between characters. Traditional segmentation techniques, including rule-based and heuristic approaches, have relied heavily on leveraging character features. However, recent advancements in deep learning, such as convolutional and recurrent neural networks, have shown significant promise in automatic segmentation, offering greater robustness and adaptability.

Segmentation Techniques Segmentation techniques can broadly be categorized into traditional and modern methods, each contributing to advancements in OCR and computer vision. Traditional methods such as thresholding, edge-based, region-based, and clustering-based segmentation have historically been the cornerstone of earlier applications.

- Thresholding methods segment images by applying fixed or adaptive thresholds to distinguish objects from the background.
- Edge-based approaches focus on detecting intensity discontinuities using operators like Sobel and Canny.
- Region-based methods group pixels with similar properties, such as intensity or texture, to identify coherent regions.

- Clustering-based methods, such as k-means, segment images by grouping pixels into clusters based on feature similarities.

While effective in controlled scenarios, these traditional techniques often struggle with complex environments featuring varying illumination, occlusions, or noise. Recent advancements in segmentation are largely driven by deep learning, which has revolutionized the field by enabling hierarchical feature extraction and improving adaptability. Methods like semantic segmentation using Fully Convolutional Networks (FCNs) and instance segmentation with Mask R-CNN have set new benchmarks by achieving pixel-level precision and class-specific region identification. Additionally, the emergence of Vision Transformers (ViTs) for segmentation showcases a shift towards architectures capable of modeling global context effectively. These developments highlight a clear trajectory toward improving robustness across diverse datasets while minimizing computational overhead.

The precision of segmentation is paramount for achieving accurate recognition results. The improvement of Arabic OCR segmentation techniques carries substantial implications for various applications, ranging from document digitization to text-to-speech conversion and language translation. Through the refinement of segmentation methods, the efficiency and effectiveness of Arabic OCR systems can experience notable enhancement, leading to improved performance in diverse language processing tasks.

E Mohamed[77] presents Arabic-SOS, a system for pre-processing pre-Modern Arabic text by focusing on segmentation and orthography standardization. The system utilizes the Gradient Boosting algorithm for training a morphological segmenter and a machine learner for text standardization. The results demonstrate high accuracy in segmentation and orthography standardization, outperforming other segmenters on a classical test set.

HA Abdo[3] introduced a novel approach for the analysis of Arabic text documents, a critical process in the development of Optical Character Recognition (OCR) systems. Their approach comprises four key steps: preprocessing, text line segmentation, word segmentation, and character segmentation. The horizontal projection method is employed to detect and extract text lines from preprocessed documents. In the word segmentation step, the computation of space thresholds determines spaces between connected components, facilitating the segmentation of text lines into isolated words. Finally, a thinning method is applied to identify the skeleton of segmented words and analyze geometric characteristics to detect ligatures and characters. The proposed approach was rigorously tested on a set of 115 text images, including those from the KHATT database and images produced by the authors. The experimental results are highly promising, achieving a success rate of 98.6% for line segmentation, 96% for word segmentation, and 87.1% for character segmentation.

S Naz[84] delved into the realm of Arabic script-based text recognition, an extensively researched field with applications in the educational process for students and educators seeking to comprehend educational content in Arabic script. Despite its long-standing prominence, the challenging nature of Arabic script recognition has been a focal point for researchers in both industry and academia. However, these endeavors have yet to yield satisfactory results. The authors particularly focused on the intricacies of segmenting Urdu script in the Nasta'liq writing style, a formidable task given its complexity compared to the Naskh writing style. Emphasizing the critical role of segmentation for achieving high accuracy, the study highlighted character segmentation as a pivotal phase in the Optical Character Recognition (OCR) process. The authors underscored the importance of segmentation by noting higher recognition rates for isolated characters compared to results for words or connected characters. The current study specifically investigated recent advancements in character segmentation and the challenges associated with segmenting languages based on the Arabic script.

A Qaroush[92] proposes an effective algorithm for word and character segmentation in printed Arabic documents that is independent of font variations. The algorithm incorporates profile projection for font-independent techniques, while the Inter-quartile Range is leveraged for the segmentation of words. Furthermore, two distinct methods are employed for character segmentation: the holistic approach, which involves a no-segmentation methodology, and the analytical method, characterized by an approach that is based on segmentation.

MA Al Ghamdi[7] addresses the challenges in Optical Character Recognition (OCR) specific to the Arabic language, driven by the increasing need to digitize Arabic content on the internet. Their focus lies in developing an effective printed Arabic OCR system, divided into four key stages: pre-processing, feature extraction, character segmentation, and classification. Unlike conventional Arabic OCR systems, the unique approach here involves conducting feature extraction before character segmentation. The pre-processing stage incorporates a novel thinning algorithm to generate skeletons for Arabic text images. The feature extraction stage introduces a novel chain code representation technique utilizing an agent-based model for non-dotted Arabic text images. This, in turn, informs a character segmentation technique for breaking down connected Arabic words into individual characters. For classification, the prediction by partial matching (PPM) compression-based method is employed. Experimental evaluation on a public dataset demonstrates the system's accuracy, achieving a notable 77.3% accuracy for paragraph-based text images.

Line segmentation. In the Optical Character Recognition (OCR) process, line segmentation is a crucial step that involves dividing the skew-corrected image into separate lines of text. This is essential because it allows the OCR system to process each line independently, leading to improved recognition of text within the image. Line segmentation can be achieved through various techniques, such as connected component analysis, project profile analysis, and machine learning approaches. By segmenting the text into lines, the OCR system can enhance the accuracy of character recognition.

A Qaroush[91] addressed the challenging task of extracting text lines from document images, a crucial step in optical character recognition and an ongoing issue in document analysis. They specifically tackled the complexities arising from diverse font variations, diacritics, and overlapping or touching text lines, which pose challenges for algorithms designed for machine-printed text. The authors introduced a simple yet robust two-stage algorithm for text-line extraction in printed Arabic text. The method efficiently extracts text lines, even in small font sizes, leveraging baselines, projection profiles, and a top-down divide and conquer technique. Notably, the proposed algorithm demonstrated effectiveness in handling non-uniform inter-line spacing and the overlapping problem. Through a series of experiments on a collected dataset, the authors compared the proposed method with two baseline approaches. The results showed that the proposed algorithm outperformed the baselines, achieving an average error rate of 3% for Arabic text without diacritics and 11% for Arabic text with diacritics. Additionally, the experiments highlighted the computational efficiency of the proposed algorithm, with an average running time of 0.087 seconds per document image.

Word segmentation. After completing line segmentation, the next phase involves word segmentation, which includes breaking down the line of text into individual words. Diverse methods are utilized for word segmentation. The Spacing method utilizes spaces between words to delineate and segment the text. Conversely, the dictionary-based method employs a word dictionary for comparison, identifying word boundaries by matching against the text. Character-based methods rely on recognized character patterns, like word breaks and punctuation, to execute word segmentation.

M Elkhayati[51] proposes a method to segment handwritten Arabic words into graphemes using a directed Convolutional Neural Network (CNN) and Mathematical Morphology Operations (MMO). Arabic's cursive script requires link removal for accurate segmentation. The study addresses challenges like diacritics and over-traces, introducing solutions: overcoming over-traces, robust diacritics extraction, and using a Partial Dilation (PD)-Global Erosion (GE) technique for segmentation. PD enhances vital areas, while GE removes inter-grapheme links, ensuring accurate segmentation despite information loss. The method utilizes a directed CNN for robustness.Also, MA Sabri[97] developed a method to automatically separate text from graphics in printed Arabic historical documents, a vital step for digitization. They introduced a fast and efficient technique using hybrid modeling involving graphs and structural analysis. The method employed the RLSA smoothing algorithm for segmentation and utilized the DTW (Dynamic Time Warping) algorithm for matching word features.

Character segmentation. Character segmentation is a method used to divide an image of a single word into distinct letters and characters. Its applicability depends on the specific requirements of the OCR system in use. If the text comprises separate and well-defined letters within a word, character-level segmentation may be unnecessary as the previous segmentation step can adequately handle letter and character separation using a thresholding approach. However, in cases where the text exhibits cursive handwriting or connected letter forms, character-level segmentation becomes essential.

NH Khan[64] conducted a comprehensive review and survey of the major studies in Urdu optical character recognition (OCR). The paper begins by introducing OCR technology and providing a historical overview of OCR systems, drawing comparisons between English, Arabic, and Urdu systems. Extensive background and literature are presented for the Urdu script, covering its history, OCR categories, and phases. The study reports on the latest advancements in different OCR phases, including image acquisition, pre-processing, segmentation, feature extraction, classification/recognition, and post-processing for Urdu OCR systems. Emphasis is placed on segmentation, discussing both analytical and holistic approaches for Urdu text. The feature extraction section includes a comparison between feature learning and feature engineering approaches, covering both deep learning and traditional machine learning methods. The paper also touches on Urdu numeral recognition systems. In conclusion, the research paper identifies open problems in the field and suggests future directions for Urdu OCR research. Also, M Tayyab[105] addresses the significant area of news ticker recognition, recognizing its importance for applications like information analysis, opinion mining, and language translation, particularly for media regulatory authorities. The focus of the study is on developing an automated Arabic and Urdu news ticker recognition system, encompassing ticker segmentation and text recognition to extract textual data for various online services. The research delves into character-wise explicit segmentation and syntactical models, considering Kufi and Nastaleeq fonts. Various network models are employed to learn deep representations, homogenizing classes regardless of inter-symbol correlations and linguistic taxonomy. The proposed learning model emphasizes fairness by maximizing balance among sensitive features of characters in a unified manner. The efficiency of the model is demonstrated through experiments using customized news tickers datasets with accurate character-level and component-level labeling. Additionally, the method is evaluated on the challenging Urdu Printed Text Images (UPTI) dataset, which only provides ligature-based annotations. The proposed method achieves a notable accuracy of 98.36%, surpassing the current state-of-the-art method. Ablation investigations highlight that the technique notably improves the performance of character classes with low symbol frequencies.

OA Boraik[39] proposed a hybrid approach for accurate segmentation of interconnected characters. Their method involved stages like removing extra shapes, detecting individual words using morphological operations, and employing computational analysis for touching character segmentation. Tests were conducted on various datasets, including KHATT and IFN/ENIT, showcasing the method's effectiveness.

3.3 Recognition

Recognition, also known as classification in earlier studies, is a pivotal stage in Optical Character Recognition (OCR) that involves identifying and assigning specific characters or character groups to an input image. After the preprocessing and segmentation phases, the OCR system compares the extracted features of each character or character group with a set of predefined templates or models. During the training phase, the system constructs these templates using a substantial dataset of sample images to teach the system accurate character recognition.

The classification or recognition process incorporates various algorithms, such as template matching, neural networks, and support vector machines. Template matching entails comparing the extracted features of each character with predefined templates and selecting the template that closely corresponds to the recognized character. In contrast, neural networks and support vector machines use machine learning algorithms to learn and classify patterns in the data, often achieving higher accuracy compared to template matching.

Recognition Techniques The field of recognition has undergone a significant transformation with advancements in object recognition methodologies, which have played a critical role in understanding and interpreting

visual data. Early approaches relied on handcrafted feature descriptors like Scale-Invariant Feature Transform (SIFT) and Speeded-Up Robust Features (SURF). These techniques focused on identifying and describing local features within images, enabling object detection and matching. While effective under certain conditions, these methods often struggled with robustness against challenges such as scale variations, changes in illumination, and differing viewpoints. The advent of deep learning has revolutionized object recognition, shifting from handcrafted features to data-driven feature learning. Convolutional neural networks (CNNs) have become the cornerstone of modern recognition systems, empowering applications like facial recognition, autonomous driving, and medical diagnostics. Unlike traditional methods, CNNs leverage hierarchical layers to extract complex feature representations, achieving unprecedented levels of accuracy and robustness. Furthermore, emerging architectures like MobileNets cater to resource-constrained devices, enabling efficient recognition with minimal computational overhead. Similarly, Transformer-based models such as DETR (DEtection TRansformer) have redefined the recognition landscape by eliminating the need for region proposal networks, streamlining the pipeline, and improving adaptability. This progression in object recognition reflects a clear trajectory towards achieving higher accuracy, efficiency, and versatility. The integration of advanced recognition techniques into OCR systems has greatly enhanced their ability to handle complex scripts like Arabic, where variability in font styles, ligatures, and noise presents significant challenges. By leveraging these advancements, OCR systems can achieve improved performance, broadening their applicability across diverse and challenging scenarios.

A Mostafa [80] created a custom dataset mimicking historical Arabic document noise and achieved significant progress in layout recognition, segmentation, and character recognition. The model, adept at transcribing handwritten manuscripts and detecting Arabic diacritics, demonstrated impressive results with a Character Error Rate (CER) of 0.0727, Word Error Rate (WER) of 0.0829, and Sentence Error Rate (SER) of 0.10. Additionally, M Badry[30] applied machine learning to recognize IoT sensor data, focusing on optical character recognition (OCR) for handwritten Arabic scripts. They proposed two deep learning models—sequence-to-sequence and fully convolutional. Assessing these models on the QTID dataset, the first Arabic dataset with diacritics, they surpassed benchmarks like Tesseract and ABBYY FineReader, achieving state-of-the-art results in character recognition rate, F1-score, precision, and recall. The proposed model achieved a character recognition rate of approximately 97.60% and 97.05% for text with and without diacritics, respectively, with an overall recognition rate of 99.48%. The CNN model exhibited a CRR of approximately 98.90% and 98.51% for text with and without diacritics, respectively.

MH Bashir [31] focused on the Qur'an, a sacred text in the Arabic language, read and followed by nearly two billion Muslims worldwide. As Islam rose to prominence, Arabic became a widespread language across large regions. Devout Muslims turn to the Qur'an daily for guidance. Despite the brevity of the Qur'an itself, an extensive body of supporting work, including commentaries and exegesis, spans tens of thousands of volumes. Recent advancements in computational and natural language processing (NLP) techniques have sparked a renewed interest in these religious texts, particularly among non-specialists. The paper provides a comprehensive survey of Qur'anic NLP efforts, encompassing tools, datasets, and approaches. The scope ranges from automated morphological analysis to the correction of Qur'anic recitation through speech recognition. The authors discuss multiple approaches for various tasks and outline potential future research directions in the field. Trained with diverse recitations and using mel-frequency cepstral coefficients (MFCCs), the model achieves an outstanding 99.89% recognition rate for 3-second Quranic recitations. This surpasses other techniques in the study, advancing Quranic NLP research for accurate and proficient recitation. Also, R Malhas[73] diligently tackled the scarcity of Arabic datasets for machine reading comprehension on the Holy Qur'an by creating QRCD, the pioneering Qur'anic Reading Comprehension Dataset. This dataset, meticulously crafted, contained 1,337 question-passage-answer triplets, incorporating 14% multi-answer questions to capture real-world complexities. To augment existing models, they introduced CLassical-AraBERT (CL-AraBERT), thoughtfully pre-trained on Classical Arabic data, enriching the Modern Standard Arabic (MSA) resources. Through meticulous cross-lingual transfer learning, they expertly fine-tuned CL-AraBERT using MSA-based MRC datasets and QRCD, leading to the development

of the very first MRC system tailored for the Holy Qur'an. Their innovative evaluation metric, Partial Average Precision (pAP), ingeniously accommodated partial matching in both multi-answer and single-answer MSA questions, demonstrating their keen attention to detail and commitment to precise evaluation methodologies.

T Milo[75] proposed an innovative Optical Character Recognition (OCR) strategy for Arabic, addressing current limitations. They introduced the concept of archigraphemes, considering Arabic script as an allographic rendering. A minimum unit, the letter block, formed the basis for script grammar and OCR. An algorithm reduced Unicode text to archigraphemes, enabling the synthesis of realistic images. The approach laid the groundwork for an OCR system, emphasizing controlled conditions and extended shaping. The study highlighted the need for further training, linguistic parsing of archigraphemes, and outlined steps towards static OCR technology, with future plans to make the OCR dynamic using AI software.

Character Recognition. Table 4 lists all the Character Recognition strategies. It also discusses various deep learning-based methods that have been used in these methods. L Bouchakour[41] explored Optical Character Recognition (OCR) for the Arabic language, a complex task due to the intricacies of Arabic script. Their OCR system comprises segmentation, feature extraction, and recognition stages. They proposed a novel method for recognizing printed Arabic characters, utilizing combined feature extraction techniques including densities of black pixels, Hu and Gabor features' invariant moments, and a Convolutional Neural Network (CNN) classifier.

N Altwaijry[21] introduced a novel dataset named Hijja, comprising 47,434 characters written by 591 children aged 7–12. Alongside the dataset, they proposed an automatic handwriting recognition model based on convolutional neural networks (CNN). The model was trained on both the new Hijja dataset and the Arabic Handwritten Character Dataset (AHCD), with accuracies of 97% and 88% on the AHCD dataset and Hijja dataset, respectively. Also, BH Nayef[83] proposed an optimized leaky rectified linear unit (OLReLU) to improve the handling of imbalanced positive and negative vectors, a common issue in handwritten character recognition. The proposed OLReLU was integrated into a CNN architecture with a batch normalization layer to enhance overall performance. Evaluation was conducted on four datasets, including the Arabic Handwritten Characters Dataset (AHCD), self-collected data, Modified National Institute of Standards and Technology (MNIST), and AlexU Isolated Alphabet (AIA9K). The proposed method demonstrated significant improvements in terms of accuracy, precision, and recall when compared to state-of-the-art methods, achieving remarkable results on AHCD (99%), self-collected data (95.4%), HIJJA dataset (90%), and Digit MNIST (99%). Additionally, YM Alwaqfi[22] addressed the complexity of Arabic character recognition by exploring generative adversarial networks (GANs), a compelling algorithm in neural networks. They proposed utilizing the sigmoid cross-entropy loss function for recognizing Arabic handwritten characters through multi-class GANs training algorithms. The evaluation, conducted on a dataset of 16,800 Arabic handwritten characters, demonstrated the effectiveness of the proposed approach, achieving an impressive accuracy of 99.7%. This research contributes to advancing Arabic character recognition and highlights the potential of multi-class GANs in enhancing recognition accuracy.

M Alheraki[15] introduced a CNN model designed to recognize children's handwriting. The model achieved a remarkable accuracy rate of 91% on the Hijja dataset, comprising Arabic characters written by children, and 97% on the Arabic Handwritten Character Dataset. Also, EF Bilgin Tasdemir [37] introduces a deep learning (DL)-based system for character recognition of printed Ottoman script. The study begins by creating a synthetic text image dataset sourced from a text corpus, which is then augmented using various image processing techniques. A hybrid architecture combining convolutional neural network (CNN) and bidirectional long short-term memory (BiLSTM) models is developed for the character recognition task. The recognizer is trained using both the original dataset and the augmented dataset. To further enhance the system's performance on real image data, a transfer learning procedure is applied. This procedure enables the adaptation of the DL-based character recognition system to real-world image data, resulting in improved accuracy and robustness. The presented work contributes to the advancement of character recognition technology specifically tailored for the printed Ottoman script. Also, A Bin Durayhim[38] aimed to develop models for recognizing children's Arabic handwriting. They introduced two deep

learning-based models: a Convolutional Neural Network (CNN) and a pre-trained CNN (VGG-16). These models were trained using the Hijja dataset, which comprises recent samples of Arabic children's handwriting collected in Saudi Arabia. Additionally, they conducted training and testing using the Arabic Handwritten Character Dataset (AHCD) for evaluation. Comparative analysis with similar models from existing literature demonstrated that their proposed CNN model outperformed both the pre-trained CNN (VGG-16) and other models considered. Furthermore, the authors created a prototype called Mutqin, designed to assist children in practicing Arabic handwriting. The prototype underwent evaluation by its intended users, and the study reports the results of this evaluation.

Table 4. A comparison of the benefits and drawbacks of several deep learning-based Character Recognition methods

Literature	Method	Benefits	Drawbacks	n Venue
L Bouchakour[41]	Combined Hu and Gabor features + CNN classifier	The method explores the use of combined features, resulting in improved recognition rates compared to using a single feature.	1) Computationally intensive. 2) Require large amounts of labeled data for effective training.	IEEE(ICRAMI) 2021
N Altwaijry[21]	CNN	1) Introduce a benchmark dataset Hijja. 2) Authors proposed a deep learning model to recognize handwritten letters	They didn't examine successful machine learning models like MLP, SVM, and autoencoders on this dataset.	Neural Comput. Appl. 2021
BH Nayef[83]	CNN + optimized leaky rectified linear unit (OLReLU)	The model maintains a balance between positive and negative feature maps while training the CNN.	OLReLU requires more time compared to using ReLU.	Multim. Tools Appl. 2022
YM Alwaqfi[22]	Generative Adversarial Networks (GANs) + Deep CNN	1) GANs generate diverse images for dataset augmentation, enhancing model robustness. 2) Stability during training is achieved using activation functions and dropout layers.	1) Model performance heavily relies on the quality of the generated dataset. 2) Addition of dropout layers may decrease model accuracy, especially when used with batch normalization.	Int. J. Adv. Soft Comput. Appl. 2022
M Alheraki[15]	Convolutional Neural Network (CNN)	1) The adoption of both single and multi-model approaches allows flexibility in training strategies, potentially improving recognition accuracy. 2) The LeakyReLU activation function, with a slope of 0.3, addresses the vanishing gradient problem, leading to improved results compared to the baseline.	1) The model's reliance on accurate stroke count information may make it sensitive to variations in the quality of the generated dataset. 2) The provided information lacks specific details on the number of neurons in certain layers, making it challenging to assess the complexity of the architecture comprehensively.	Hum. Centric Intell. Syst. 2023
EF Bilgin Tasdemir [37]	CNN-BLSTM-CTC(Connectionist Temporal Classification)	1) The hybrid architecture combines CNN-BLSTM-CTC has shown success in sequence labeling tasks. 2) BLSTM networks are effective in capturing temporal dependencies in sequence labeling tasks, contributing to the model's ability to recognize the structure of Ottoman text.	1) The synthetic dataset's size is constrained by the source text corpus, limiting the model's capacity for diverse learning. 2) The system's effectiveness heavily depends on the quality and quantity of the training data	Int. J. Document Anal. Recognit. (IJDAR) 2023
A Bin Durayhim[38]	CNN , VGG-16	1) Attains superior accuracy on standardized datasets. 2)Streamlines training processes, thereby optimizing model efficiency.	1) Computationally intensive. 2) Additional post-processing processes are necessary	Applied Sciences 2023

Word Recognition. Table 5 lists all the Character Recognition strategies. It also discusses various deep learning-based methods that have been used in these methods. K Hamad [58] conducted a comprehensive investigation into Optical Character Recognition (OCR), recognizing its significance in various fields for storing and retrieving information from printed or handwritten documents. They emphasized the challenges associated with OCR, including font characteristics and image quality. The paper provided a detailed overview of OCR stages, covering pre-processing, segmentation, normalization, feature extraction, classification, and post-processing. Additionally, the authors explored the historical development of OCR, its main applications, and recent advancements. This comprehensive review aims to contribute a thorough understanding of the challenges and advancements in OCR technology, addressing the complexities of extracting and processing text from diverse document formats.

M Awni[28] employs model averaging as an ensemble learning method, training three Residual Networks (ResNet18) models. The core concept involves combining diverse classifiers to compensate for individual mistakes, ultimately improving the ensemble's prediction accuracy. The ResNet18 architecture is utilized for its balance between depth and performance, with modifications to align the output nodes with the distinct words present in the IFN/ENIT database. The optimization techniques employed include gradient descent, Root Mean Square Propagation (RMSProp), and Adaptive Momentum Estimation (Adam). To efficiently determine the learning rate, the authors employ a Learning Rate Range Finder test, crucial for training convolutional neural networks. Finally, the ensemble model for offline Arabic handwritten word recognition incorporates variations in optimization techniques for each member, employing model averaging for the final prediction. The methodology is validated through experiments on the IFN/ENIT database, demonstrating the effectiveness of the proposed ensemble approach. Also, SK Jemni[63] addressed the issue of out-of-vocabulary (OOV) words in Arabic handwriting

recognition systems, which typically operate with a fixed vocabulary. They proposed a two-step approach to tackle this problem. In the first step, various sub-word units were employed to identify potential OOVs. In the recovery stage, a dynamic dictionary was created to augment the initial static word lexicon, accommodating the detected OOVs. The recovery process involved selecting the best word candidates from an external resource. The experiments conducted on the KHATT and AHTID/MW databases demonstrated that sub-word modeling contributed to better detection, and the use of a dynamic dictionary significantly enhanced recognition performance compared to one-step approaches based on a large static dictionary or the combination of different sub-word units. The proposed approach achieved state-of-the-art results on the KHATT dataset. Additionally, M Eltay[53] introduced an adaptive data augmentation algorithm assigning weights to words based on class probabilities. The RNN-LSTM architecture, configured with LSTM layers and employing the word beam search algorithm, demonstrated state-of-the-art results. They adapted the AlexNet for the task, replacing the last layer and employing transfer learning. Additionally, the authors proposed a method for out-of-vocabulary (OOV) word detection and recovery, achieving improved recognition performance compared to existing approaches.

H Butt[43] proposed a CNN-RNN model with an attention mechanism for Arabic image text recognition. The model utilizes a CNN to generate feature sequences from input images, followed by a bidirectional RNN to arrange these sequences. To enhance text segmentation, a bidirectional RNN with an attention mechanism is employed to produce the final output. This attention mechanism enables the model to selectively focus on relevant information within the feature sequences. The end-to-end training is implemented through a standard backpropagation algorithm, showcasing the effectiveness of the proposed approach in Arabic image text recognition. Also, S Bergamaschi[34] addresses the imperative need for managing multicultural heritages in a globalized context. The DigitalMaktaba project, a collaborative effort among computer scientists, historians, librarians, engineers, and linguists, aims to establish efficient procedures for cataloging archival heritage in non-Latin alphabets. The paper discusses the ongoing development of a novel workflow and tool for text sensing, focusing on automatic knowledge extraction and cataloging of documents in languages like Arabic, Persian, and Azerbaijani. The prototype utilizes advanced OCR, text processing, and information extraction techniques to ensure accurate text extraction and rich metadata content, surpassing current limitations. Additionally , N Alzrrog[24] addressed the limited progress in research on automatic Arabic handwriting word recognition using deep learning neural networks. They highlighted the absence of a general and reliable Arabic Handwritten words database, which hinders the advancement of related research. To fill this gap, the authors introduced a new Deep Convolutional Neural Network (DCNN) algorithm applied to a novel dataset called Arabic Handwritten Weekdays Dataset (AHWD). This dataset, consisting of 21,357 words distributed among seven classes, was meticulously prepared by 1,000 individuals. The proposed DCNN model was trained on this balanced dataset using various structures and enhanced with techniques like dropout, image regularization, and proper learning rates to prevent overfitting. The model demonstrated promising performance, achieving an accuracy rate of 99.39% with an error rate of 4.61% on the AHWD dataset and an accuracy rate of 99.71% with an error rate of 1.71% on the IFN/ENIT dataset. The authors conducted a blind test on the hidden test set and assessed performance using a confusion matrix and learning curves, affirming the model's effectiveness in Arabic handwriting word recognition.

GJ Salman [99] tackled the challenging task of recognizing Arabic words and texts, especially when presented in varying sizes and fonts. They developed an intelligent system that involved the creation of a comprehensive dataset comprising 1,000 words, each written in 24 different ways using various Arabic fonts. Leveraging image processing methods, the system identified and deduced words from images. The core of the system relied on a deep learning approach, specifically a Convolutional Neural Network (CNN) algorithm. This CNN algorithm was trained to extract features from truncated words and retrieve text words that closely resembled the ones that were cut. In experimental evaluations, the system demonstrated remarkable performance, achieving a 99% accuracy in word detection and a 96% accuracy in word recognition. Also, I Ouali [87] present a novel system designed to recognize and identify Arabic Handwritten Texts with Diacritics (AHTD) through the utilization of

deep learning, specifically the convolutional neural network. The system is trained, tested, and validated using our developed Arabic Handwritten Texts with a Diacritical Dataset (AHT2D). Following the recognition process, the identified text undergoes enhancement through augmented reality (AR) technology, resulting in a visually enhanced 2D image representation. Furthermore, the authors leverage AR technology to convert the recognized text into an audio output, catering to the needs of visually impaired users. By providing both voice and visual outputs, our system offers an inclusive solution to facilitate effective communication and accessibility for visually impaired individuals.

S Hamida[59] introduced a novel image processing approach that combines three image descriptors to extract features from handwritten Arabic text. Using a subset of 100 classes from the IFN/ENIT handwritten Arabic database, they applied preprocessing techniques and trained separate k-nearest neighbor (k-NN) models for each feature descriptor. These models were used to classify Arabic handwritten images, and their performance was evaluated using common metrics. Impressively, the final model achieved an exceptional recognition rate of up to 99.88%. Additionally, S Malakar[72] focused on handwritten word recognition (HWR), a persistent challenge in the field. The study chose the holistic approach for its effectiveness with limited lexicons. It noted the absence of consideration for inter-segment similarity in existing methods, which could offer valuable insights. To address this gap, the authors introduced Hausdorff and Fréchet distances to quantify similarity among different word segments. They also used shape-based descriptors and combined outputs from six classifiers using majority voting. Evaluation on standard databases (IAM and IFN/ENIT) yielded promising results compared to state-of-the-art HWR methods. Also, YM Alwaqfi[23] addressed challenges in recognizing printed Arabic words due to the language's complexity and limited available datasets. They proposed a hybrid model combining a deep convolutional neural network (DCNN) as a classifier and a generative adversarial network (GAN) for data augmentation. This hybrid model significantly improved accuracy and generalization ability, achieving a remarkable accuracy score of 99.76% on the Arabic printed text image dataset (APTI) compared to 94.81% with the DCNN alone.

Digits Recognition. Table 6 lists all the Character Recognition strategies. It also discusses various deep learning-based methods that have been used in these methods.

E Al-wajih[11] addressed the challenge of Arabic handwritten digit recognition, leveraging sliding windows for enhanced classification accuracy. Employing Random Forests (RF)and Support Vector Machine (SVM) classifiers, they introduced four feature extraction techniques—Mean-based, Gray-Level Co-occurrence Matrix (GLCM), Moment-based, and Edge Direction Histogram (EDH). Through meticulous evaluation, the study demonstrated the efficacy of sliding windows, achieving notable recognition rates, such as 98% for Mean-based and Moment-based with RF, and 98.33% and 99.13% for GLCM and EDH with linear-kernel SVM. Using a modified AHDBase dataset, the proposed model, incorporating various sliding window sizes, offers insights into optimizing feature extraction and classification. Comparative analysis against cutting-edge approaches underscores the significance of this approach in advancing Arabic handwritten digit recognition, providing a foundation for future research in optimal methodology combinations.

YS Can[44] focused on the digitization and recognition of Arabic numerals from the initial series of population registers of the Ottoman Empire in the mid-nineteenth century. To isolate numerals written in red, a color filter was applied, leveraging the distinctive structure of the registers. Initially, a convolutional neural network (CNN)-based segmentation method was employed for numeral spotting. Subsequently, the authors curated a local Arabic handwritten digit dataset from the identified numerals and tested a Deep Transfer Learning method on this dataset for digit recognition.

In the research of F Haghighi[57] a novel model for recognizing handwritten digits is introduced to address the challenge posed by the distinctive writing styles of individuals, especially in languages like Persian/Arabic. The proposed model is a stacking ensemble classifier, combining Convolutional Neural Network (CNN) and

Table 5. A comparison of the benefits and drawbacks of several deep learning-based Word Recognition methods

Literature	Method	Benefits	Drawbacks	Venue
M Awni[28]	Ensemble of Residual Networks (ResNet18)	More robust to overfitting and noise in the data	Need more post-processing processes	IEEE (ICCES) 2019
SK Jemni[63]	CNN-MDLSTM with Dynamic lexicons (DWLD)	More robust to overfitting and noise in the data	Need more post-processing processes	Pattern Recognit. 2019
M Eltay[53]	Bidirectional Long Short-Term Memory(BLSTM)-Connectionist Temporal Classification(CTC)-Word Beam Search(WBS)	Solve effectively the data imbalance problem	The method used leads to increase in the training time of the network	IEEE Access 2020
H Butt[43]	CNN-RNN with Attention for Arabic Image Text Recognition	1) Bidirectional RNNs handle sequential data, capturing contextual dependencies in Arabic text.	Need higher computational requirements.	Forecasting 2021
S Bergamaschi[34]	OCR (GoogleDocs, EasyOCR, Tesseract)	1) The tool speeds up document cataloguing, reducing manual effort significantly, especially for simpler documents 2) Automatic suggestions enhance accuracy, minimizing errors during data insertion and ensuring consistent catalogued information.	1) The system's performance relies on image quality, potentially limiting effectiveness with suboptimal images. 2) Limited development in Arabic-script text sensing poses challenges for full automation in this context. 3) Large data loads are required for machine learning features, demanding a substantial initial manual workload.	Sensors 2022
N Alzrrog[24]	Deep Convolutional Neural Network (DCNN)	The model demonstrates a high accuracy rate.	1) Making a dataset private means we cannot experience their results.	IEEE(CCECE) 2022
GJ Salman [99]	Multi-Font Arabic Word Recognition CNN	Inclusion of a dataset with 1,000 words written in 24 different ways using various Arabic fonts contributes to comprehensive training.	The need for a diverse dataset with multiple font variations necessitates extensive data collection and preprocessing efforts.	IEEE(DeSE) 2023
I Ouali [87]	AHTD based on CNN and Augmented Reality (AR)	Integrates with smart glasses through a mobile application, allowing users to capture, extract, and listen to text content, enhancing the overall user experience.	1) Limited discussion on AR engine; more details on Vuforia's suitability and alternatives would improve comprehension. 2) The system assumes user eye movement for text detection and recognition, potentially limiting applicability to all visually impaired users.	Multim. Tools Appl. 2023
S Hamida[59]	k-nearest neighbor algorithm (k-NN)	Integrates HOG, GF and LBP for precise feature extraction, enhancing Arabic handwritten text recognition.	Involves intricate preprocessing steps	Multim. Tools Appl. 2023
S Malakar[72]	Hausdorff and Fréchet distances for inter-segment similarity features in word recognition.	1) Introduces a holistic word recognition method using inter-segment similarity, offering a unique perspective. 2) Leverages diverse features like inter-segment similarity, shape-based, and contour-based for improved recognition accuracy.	1) Additional post-processing processes are necessary. 2) Relies on experimental threshold values, making it sensitive to parameter tuning	Vis. Comput. 2023
YM Alwaqfi[23]	Hybrid GAN-based Model + DCNN	1) Achieves outstanding performance. 2) The strategy based on GAN compels the network to extract similar characteristics.	1) No comparison with other methods. 2) The model relying on a generator is vulnerable when dealing with images of varying layouts.	Int. Journal of Adv. CS App. 2023

Bidirectional Long-Short Term Memory (BLSTM) techniques. An innovative aspect of the model involves using the probability vector of the images' class as input for the meta-classifier layer, leveraging BLSTM's ability to understand arrays and vectors. By doing so, the model aims to enhance the accuracy of the deep learning model, particularly in capturing the structural similarities of certain Persian/Arabic digits. Also, RS Alkhawaldeh[18] proposed an Ensemble Deep Transfer Learning (EDTL) model, combining two pre-trained transfer learning models, to effectively detect and recognize these digits. The EDTL model demonstrated exceptional performance, achieving up to 99.83% accuracy, surpassing baseline methods, including deep transfer learning models and ensemble deep transfer learning models. The proposed architecture included parallel and sequence blocks with convolution and pooling operations, incorporating ResNet-50 and MobileNetV2 models. This architecture addressed the vanishing gradient issue through skip connections and depth-wise separable convolutions. The EDTL model efficiently extracted relevant features from noisy handwritten digits. The classification phase involved merging these features into a fully-connected Artificial Neural Network (ANN) for accurate recognition, achieving state-of-the-art performance on popular datasets.

S Ali[17] addresses the challenges in recognizing Persian/Arabic handwritten digits, a task crucial for applications like office automation and document processing. They propose a modified Deep Convolutional Neural Network (DCNN) architecture, incorporating three convolutional layers with features such as batch normalization, pooling, fully connected layers, and dropout regularization to enhance generalization and prevent overfitting. The study utilizes the HODA database, applying preprocessing steps like image smoothing and resizing. Various optimization algorithms, including stochastic gradient descent and Adam, are explored to optimize the DCNN.

The research investigates the impact of different epochs on Optical Character Recognition (OCR) performance and determines optimal learning parameters.

Table 6. A comparison of the benefits and drawbacks of several deep learning-based Digits Recognition methods

Literature	Method	Benefits	Drawbacks	Venue
E Al-wajih[11]	sliding windows + random forests (RF) + support vector machine (SVM)	1) Enhace classification accuracy for Arabic digit images. 2) Comprehensive comparison with recent state-of-the-art approaches validates performance.	Experiments are constrained in image sizes and sliding window sizes.	Iran. J. Sci. Technol. Trans. Electr. Eng. 2020
YS Can[44]	Deep Transfer Learning (DTL) + CNN	The study implemented an automatic Arabic numeral spotting system on historical Ottoman Empire population registers with notable accuracy.	limited applicability of AHDBase, yielding lower accuracy (72% with CNN + MLP) than CNN alone for recognizing digits in local datasets.	Applied Sciences 2020
F Haghighi[57]	CNN + Bidirectional Long-Short Term Memory (BiLSTM)	1) provides a holistic solution to the challenges posed by diverse writing styles. 2) The model utilizes a probability vector of images' class as input for the meta-classifier layer, enhancing feature extraction capabilities, especially in capturing structural similarities.	1) To provide equivalent results, many processing stages are necessary. 2) The research acknowledges a weakness in data processing speed, primarily attributed to the size of input data	Knowl. Based Syst. 2021
RS Alkhawaldeh[18]	Ensemble Deep Transfer Learning (EDTL)	1) Reducing time and cost complexities during training. 2)Effectively extracted relevant features from noisy handwritten digits. 3) Combined the strengths of two pre-trained models, resulting in enhanced model performance	1) The Need for effective tuning of hyperparameters to achieve optimal results. 2) The combined size of ResNet-50 and MobileNetV2 may result in a larger model size, impacting deployment in resource-constrained environments.	Neural Comput. Appl. 2020
S Ali[17]	Modified Deep Convolutional Neural Network (DCNN)	1) Effective handling of sparse gradients. 2) A streamlined and effective approach for Arabic digit recognition	Larger feature sizes increase training time; the proposed approach, tested on 40x40-pixel images, may need validation for alternative sizes.	Multim. Tools Appl. 2023

Multifaceted Recognition Approaches. Table 7 lists all the Character Recognition strategies. It also discusses various deep learning-based methods that have been used in these methods.

N Alrobah [19] conducted a comprehensive survey of research projects and experiments that aim to develop machines capable of automatically recognizing handwritten characters, with a specific focus on the unique challenges posed by Arabic script. Unlike studies primarily conducted in Latin scripts, recognizing handwritten Arabic characters is inherently complex due to the nature of Arabic words. The paper particularly delves into recent advancements in the application of deep learning approaches to Arabic character recognition. The main objective is to categorize, analyze, and present a systematic survey of state-of-the-art methods, with a special emphasis on deep learning for feature extraction in offline text recognition. The analysis critically evaluates existing literature, identifies challenges and problem areas, proposes a new classification of the literature, and engages in a thorough discussion of the issues and challenges associated with recognizing Arabic scripts. And, R Ahmed[4] propose a novel context-aware model based on deep neural networks, specifically a supervised Convolutional Neural Network (CNN). This CNN model is designed to contextually extract optimal features, incorporating batch normalization and dropout regularization parameters to prevent overfitting and enhance generalization. The architecture utilizes deep stacked-convolutional layers, forming the proposed Deep CNN (DCNN). The model undergoes comprehensive evaluations across six benchmark databases, showcasing its superior classification accuracy compared to conventional OAHR approaches. Transfer learning (TL) is employed for feature extraction, demonstrating the model's superiority over state-of-the-art pre-trained VGGNet-19 and MobileNet models. Also, A Mortadi[78] proposed using the TrOCR architecture, a deep learning model based on Transformers, known for its superior performance in text recognition. The authors customized the TrOCR model's encoder and decoder specifically for Arabic script, considering its unique features. Through rigorous experimentation, their approach demonstrated outstanding accuracy.TrOCR incorporates pre-trained computer vision (ViT-style) and natural language processing (BERT-style) models for initializing the encoder and decoder. The encoder processes text line images, partitioning them into patches, which are then flattened, projected into a D-dimensional vector, and combined with positional embedding. The decoder, equipped with multi-head self-attention and feed-forward neural network (FFN) layers, features an encoder-decoder multi-head attention layer. The proposed model, based on TrOCR, modifies the encoder to employ a Vit model and integrates the BERT-Small model's self-attention and FFN layers for the decoder. When applied to recognizing text from scanned

Arabic documents at the word level, the approach achieved remarkable precision. The character error rate (CER) was 0.8, indicating precise character recognition, while the word error rate (WER) was 2.3, signifying accurate transcription of complete words. These results highlight the effectiveness of their tailored method in capturing the intricacies of Arabic script during OCR tasks.

HM Al-Barhamtoshy[6] introduces a novel framework leveraging the Fast Gradient Sign Method (FGSM) in Keras and TensorFlow for enhanced Arabic manuscript recognition in OCRing systems. The framework focuses on improving OCRing accuracy through deep learning in a multilingual context, optimizing image enhancement, alignment, and layout analysis. RoI detection, performed using a custom-trained deep learning model with bounding box regression, extends the Page Segmentation Method (PSM). The proposed framework, featuring a CNN architecture for adversarial training and FGSM implementation, achieves significant OCRing accuracy improvements, especially in language identification, document category, and diverse RoI types. The model exhibits resilience against adversarial attacks and attains a notable 99% accuracy in experimental results across various datasets. The study underscores the framework's novelty and efficacy in advancing OCRing quality through innovative approaches to RoI detection and language/document categorization. Also, M El Mamoun[74]introduced a hybrid approach that combines two powerful classification techniques. Initially, they employed a trained Convolutional Neural Network (CNN) to extract features from the character images. Subsequently, they utilized a Support Vector Machine (SVM) for classification purposes. The goal of combining CNN and SVM was to leverage the strengths of both technologies. The authors developed four hybrid models and evaluated their performance using various databases, including HACDB, HIJJA, AHCD, and MNIST. The results they achieved were promising, with test accuracy rates of 89.7%, 88.8%, 97.3%, and 99.4%

Table 7. A comparison of the benefits and drawbacks of several deep learning-based MultiFacets Recognition methods

Literature	Method	Benefits	Drawbacks	Venue
R Ahmed[4]	DCNN	The proposed approach adeptly handles high-dimensional data.	The proposed model is not suitable for real databases with an insufficient number of training samples.	Entropy 2021
A Mortadi[78]	TrOCR	1) Achieving high accuracy in word-level recognition for scanned Arabic documents. 2) Significantly reducing the need for a large dataset, distinguishing their approach from other methods.	Improvements should be made to bolster the system's robustness, encompassing enhancements in denoising algorithms, image preprocessing techniques, and the ability to manage low-quality scans	IEEE(MIUCC) 2023
HM Al-Barhamtoshy[6]	FGSM + ORCing	Authors executed noise removal procedures, identified skewing, and applied de-skewing corrections, resulting in high-quality outcomes.	Demands high computational requirements	ACM Trans. Asian Low Resour. Lang. Inf. Process. 2023

Multimodal Approaches. G Bhatia[35] introduces Qalam, a novel multimodal foundation model for Arabic Optical Character Recognition (OCR) and Handwriting Recognition (HWR), combining a SwinV2 encoder with a RoBERTa decoder to address the unique challenges of Arabic script, such as its cursive nature, diacritical marks, and context-sensitive letter forms. By leveraging a diverse dataset of over 4.5 million Arabic manuscript images and a synthetic dataset with 60,000 image-text pairs, Qalam achieves state-of-the-art performance with Word Error Rates (WER) of 0.80% for HWR and 1.18% for OCR. The model's multimodal approach, which integrates textual and visual data, enables superior processing of high-resolution inputs and diacritic recognition—critical components for accurate Arabic text recognition. Furthermore, Qalam introduces the Midad Benchmark for evaluating Arabic OCR and HWR systems, offering a comprehensive resource for the community. These advancements demonstrate the value of combining multiple modalities and datasets to improve robustness in Arabic OCR systems, setting a new standard in the field and paving the way for further research into multimodal approaches in character recognition.

End-to-End OCR Systems. M Boualam[40] developed an end-to-end offline Arabic handwriting recognition system using a hybrid Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) architecture with Connectionist Temporal Classification (CTC) for transcription. The system is trained on the IFN/ENIT database, enhanced through data augmentation techniques to expand the dataset to 487,350 images, ensuring diverse training samples. The CNN layers serve as feature extractors, while bidirectional RNN layers model

sequences, and the CTC layer provides segmentation-free transcription. The model achieved state-of-the-art performance on the IFN/ENIT dataset, with a Character Error Rate (CER) of 2.1% and a Word Error Rate (WER) of 91.79%, significantly outperforming previous approaches. The work emphasizes the effectiveness of combining CNNs and RNNs in OCR tasks, with promising future directions focused on dataset quality enhancement and model complexity improvements. Also, S Aabed[1] propose an end-to-end, segmentation-free deep learning model for Arabic handwritten recognition using the KHATT dataset. The model integrates Deep Convolutional Neural Networks (DCNN) for feature extraction with Bidirectional Long Short-Term Memory (BLSTM) layers for sequence recognition, trained with a Connectionist Temporal Classification (CTC) loss function. Extensive preprocessing techniques, such as Gaussian smoothing, adaptive thresholding, and distortion-free resizing, are applied to handle variability in handwriting styles and image quality. The model achieves an 84% character recognition rate and 71% word recognition rate without segmentation, operating efficiently at the line level. Also, A Mostafa[81] present a novel end-to-end approach for Arabic OCR by leveraging a Transformer-based architecture. They address key challenges in recognizing historical Arabic manuscripts, such as complex layouts, degraded images, and diverse fonts, by constructing the largest Arabic OCR dataset to date, comprising 30.5 million images, 270 million words, and 1.6 billion characters. Their proposed model replaces traditional CNN feature extractors with BEiT (Bidirectional Encoder representation from Image Transformers) as the encoder and a Vanilla Transformer as the decoder, significantly reducing model complexity. Additionally, they enhance OCR performance with a complete pipeline incorporating image preprocessing, segmentation using Detectron-2, and post-correction strategies such as Word Beam Search and BERT-based autocorrect. Through extensive experimentation, the authors demonstrate that their approach achieves a character error rate (CER) of 4.46%, outperforming convolutional backbones and showcasing its potential for large-scale Arabic OCR tasks.

Adversarial Attacks and Defenses in Arabic OCR. Adversarial attacks exploit vulnerabilities in OCR systems, challenging their reliability and security. Recent advancements focus on robust defenses, including adversarial perturbations, to protect OCR models while preserving accuracy and usability. Y He[60] proposed ProTegO, a novel text protection method against OCR extraction attacks, by generating adversarial underpaintings that mislead OCR models while preserving human readability through a flicker fusion effect. The framework creates universal, text-style guided adversarial perturbations using a conditional GAN and a guided network to enhance transferability across OCR models. ProTegO achieves robust protection against commercial OCR services and diverse architectures under black-box settings by incorporating robustness enhancement techniques like translation and binarization, along with a visual compensation strategy for imperceptibility. This method demonstrates high protection success rates while maintaining visual quality, offering a practical solution to safeguard text content against OCR vulnerabilities. Also C Vizcarra[106] explored adversarial attacks on License Plate Recognition (LPR) systems using methods like FGSM, C&W, and watermarking, demonstrating their impact on recognition accuracy. To counteract these vulnerabilities, they proposed image denoising and inpainting defenses, integrating them into the OCR pipeline. Testing on Saudi license plates showed improved recognition accuracy and robustness, offering practical, efficient solutions for real-world LPR systems.

Privacy concerns on Arabic OCR. The deployment of Arabic OCR technology raises significant ethical concerns, particularly around privacy and algorithmic biases. Privacy issues arise when OCR systems process sensitive documents, such as financial or personal identification records, potentially exposing data to unauthorized access or misuse. Methods like differential privacy, which masks individual data through statistical noise, and encryption techniques, which secure the storage and transmission of OCR outputs, are essential safeguards. Algorithmic biases in Arabic OCR systems often result from imbalanced training datasets that favor specific dialects, scripts, or document types. Addressing these biases requires strategies such as data augmentation, which diversifies training datasets by generating synthetic Arabic text samples with varied fonts and writing styles, and adversarial training, which improves model robustness. Explainable AI (XAI) methods also play a crucial role by offering transparency and identifying potential biases in decision-making processes. By incorporating fairness-aware

learning algorithms, such as those enforcing demographic parity, Arabic OCR systems can ensure equitable recognition while safeguarding individual rights. MT Parvez[88] highlighting their potential for tailored user authentication. These CAPTCHAs cater to 444 million Arabic speakers and are adaptable to similar scripts like Urdu and Persian. They analyzed state-of-the-art Arabic handwritten text-based CAPTCHAs, emphasizing opportunities arising from the script's complexity, including connected letters and diacritical marks, which pose unique challenges for automated systems. The study revealed limited research on Arabic CAPTCHAs, especially in multimedia formats like image, video, and audio. Key challenges include script variability, lack of standardized datasets, and linguistic nuances. The authors proposed addressing these gaps by creating robust datasets and developing advanced recognition-resistant techniques to innovate and improve Arabic CAPTCHA systems.

3.4 Postprocessing

Postprocessing serves as the final step in the OCR pipeline, refining recognized text to improve accuracy and quality. Traditional postprocessing techniques include spell-checking, contextual analysis, confidence scoring, and the integration of language models. Spell-checking compares the recognized text against comprehensive dictionaries to correct spelling errors, while contextual analysis evaluates the text within its linguistic and semantic context to resolve word-level ambiguities. Confidence scoring assigns a certainty score to each character, enabling targeted review of low-confidence outputs. Language models further enhance accuracy by aligning the recognized text with linguistic rules and context.

In computer vision, postprocessing also refines visual outputs to correct errors and enhance interpretability. Traditional approaches such as morphological operations (e.g., dilation and erosion) and edge detection techniques (e.g., Sobel and Prewitt operators) were widely used to clean segmentation masks, fill gaps, and sharpen object boundaries. Filtering methods, like Gaussian or median filters, suppressed artifacts and improved feature clarity.

Recent advancements have introduced data-driven approaches to postprocessing. Techniques such as conditional random fields (CRFs) and graph-based optimization refine segmentation outputs, particularly in applications like medical imaging and autonomous systems. Neural networks, including adversarial models for image refinement and Transformer-based methods for detail enhancement, now play a pivotal role in optimizing visual outputs. This evolution reflects a shift from static, rule-based methods to dynamic, data-driven frameworks, ensuring greater adaptability and precision across diverse tasks and datasets.

IA Doush[48] introduced a novel Optical Character Recognition (OCR) post-processing model tailored for the complexities of the Arabic language. By combining a statistical Arabic language model with innovative post-processing techniques, including the integration of error and context approaches, the system demonstrated substantial improvements. Trained on a carefully curated Arabic OCR context database, the hybrid system significantly reduced word error rates from 24.02% to 18.96%. Testing on a separate dataset further confirmed its superiority, achieving a word error rate of 14.42%. This study marks a pioneering effort in developing an end-to-end OCR post-processing model specifically designed for Arabic text, showcasing promising advancements in accuracy and efficiency.

Y Bassil[32] enhanced OCR accuracy by integrating Google's spelling suggestions based on N-gram probability. Their hybrid system combines these suggestions with a proposed postprocessing technique. The OCR-generated tokens are evaluated using a language model; unrecognized tokens trigger the error model, suggesting correct words based on Google's algorithm. Verified tokens advance to the next word, ensuring improved accuracy through advanced suggestions.

TTH Nguyen[86] emphasized the need for post-correction to enhance the quality of OCR results. The article systematically explored the post-OCR processing problem, outlining its common pipeline and reviewing contemporary approaches. It emphasized the impact of improved OCR quality on information retrieval and natural language processing applications. The work also provided valuable insights into evaluation metrics, available datasets, language resources, and practical toolkits in the context of post-OCR processing. Additionally, the study identified current trends and suggested future research directions in this evolving field.

3.5 Evaluation

Assessing the precision and caliber of identified text generated through Optical Character Recognition (OCR) for Arabic necessitates employing various crucial evaluation criteria and metrics. These parameters serve to gauge the effectiveness of OCR systems, evaluating their ability to accurately transform scanned or handwritten Arabic text into machine-readable format. Below are some prevalent evaluation criteria applied in the context of Arabic OCR.

Character Recognition Rate (C_{RR}). CRR measures the accuracy of recognizing individual characters in the recognized text compared to the ground truth.

$$C_{RR} = \frac{\text{Number of Correctly Recognized Characters}}{\text{Total Number of Ground Truth Characters}} \tag{1}$$

Word Recognition Rate (W_{RR}). WRR assesses the accuracy of recognizing entire words in the recognized text compared to the ground truth.

$$W_{RR} = \frac{\text{Number of Correctly Recognized Words}}{\text{Total Number of Ground Truth Words}} \tag{2}$$

Line Recognition Rate (L_{RR}). LRR evaluates the accuracy of recognizing entire lines of text in the recognized output compared to the ground truth.

$$L_{RR} = \frac{\text{Number of Correctly Recognized Lines}}{\text{Total Number of Ground Truth Lines}} \tag{3}$$

Character Error Rate (CER). CER quantifies the number of character-level errors in the OCR output, including substitutions, insertions, and deletions. The Levenshtein distance is determined by dividing the total number of characters in the ground truth word (N) by the combined count of character substitutions (S), insertions (I), and deletions (D) needed to transform one string into another.

$$CER = \frac{S + I + D}{N} \tag{4}$$

C Reul[95] is an open-source OCR software designed for historical print materials. It offers a user-friendly interface, advanced OCR components, and adaptability. OCR4all outperforms commercial solutions, particularly in complex layouts, delivering low Character Error Rates (CER). Its flexibility allows easy integration of new tools, making it valuable for non-technical users.

Word Error Rate (WER). WER is similar to CER but measures errors at the word level. It considers substitutions, insertions, and deletions of complete words. WER is often used to assess the overall quality of OCR output. Likewise, the *WER* is calculated by adding the count of term substitutions (Sw), insertions (Iw), and deletions (Dw) necessary to transform one string into another, and then dividing this sum by the total number of ground-truth terms (Nw). [56]

$$WER = \frac{S_w + I_w + D_w}{N_w} \tag{5}$$

MA Alghamdi[14] addressed the deficiency in performance evaluation tools for Arabic Optical Character Recognition (OCR) systems by developing an open-source automated software tool. Recognizing the limitations in existing metrics like character accuracy and word accuracy, the authors designed a tool that offers a comprehensive set of metrics tailored specifically for evaluating Arabic OCR performance. This tool is intended to contribute to the advancement of Arabic OCR research by providing researchers with a valuable resource for assessing and ranking different OCR algorithms.

Precision and Recall. Precision (P) measures the fraction of correctly recognized characters or words relative to the total recognized. Recall (R) calculates the fraction of correctly recognized characters or words relative to the total ground truth.

$$\text{Average Precision (AP)} = \frac{\text{True Positive (TP)}}{\text{(True Positive (TP) + False Positive (FP))}} = \frac{TruePositive}{TotalObservations} \tag{6}$$

$$\text{Average Recall (AR)} = \frac{\text{True Positive (TP)}}{\text{(True Positive (TP) + False Negative (FN))}} = \frac{TruePositive}{TotalGroundTruth} \tag{7}$$

F1-Score. The F1-Score combines precision and recall into a single metric, providing a balanced evaluation measure.

$$\text{F1-score} = \frac{2 * (\text{AP} * \text{AR})}{(\text{AP} + \text{AR})} \tag{8}$$

C Neudecker [85] addressed the challenges in comprehensively assessing the quality of Optical Character Recognition (OCR) results for digitized historical documents. Recognizing the limitations in existing OCR evaluation metrics and tools, especially when dealing with large-scale digitization, the authors conducted an experiment using multiple evaluation tools and diverse metrics on two distinct datasets. They emphasized the need to go beyond traditional OCR metrics and sampling methods, highlighting the importance of accurate layout analysis and detection of reading order for advanced applications such as Natural Language Processing and the Digital Humanities. The study analyzed variations in results from different evaluation tools and metrics, providing insights into areas for future improvements in OCR evaluation methodologies. Also, M Elzobi [55]conducted a comparative study to assess the performance of discriminative and generative strategies in the context of handwritten word recognition. They utilized generatively-trained hidden Markov modeling (HMM), discriminatively-trained conditional random fields (CRF), and discriminatively-trained hidden-state CRF (HCRF) on learning samples obtained from two distinct databases. Initially employing an HMM classification scheme, the researchers introduced adaptive threshold schemes to enhance the model's ability to reject incorrect and out-of-vocabulary segmentations. Additionally, the study extended its investigation by introducing CRF and HCRF classifiers for the recognition of offline Arabic handwritten words. The evaluation was comprehensive, involving two different databases, and the research presented recognition outcomes for both words and letters, shedding light on the strengths and weaknesses of each strategy. Additionally, S Singh [101] specifically addressed offline handwritten Devanagari words, leveraging statistical features. Feature vector sets were created, describing each word in the feature space through uniform zoning-, diagonal-, and centroid-based features extracted from a database of handwritten word images (comprising 50-word classes). Various classifiers, including K-nearest neighbor (KNN), decision tree, and random forest, were employed for recognition. To enhance system performance, the study proposed a combination of the aforementioned features along with the gradient-boosted decision tree algorithm. The proposed system achieved a maximum recognition accuracy of 94.53%, demonstrating competitive results compared to existing state-of-the-art methods. Additionally, the system achieved an F1-score of 94.56%, FAR of 0.11%, FRR of 5.46%, MCC of 0.945, and AUC of 97.21%.

Normalized Edit Distance (1-NED). The normalized edit distance (1-NED) metric is widely recognized in scene text recognition tasks, as it provides a robust measure of similarity between the predicted text and the ground truth. This metric calculates the edit distance (i.e., the minimum number of operations required to transform one string into another) and normalizes it by the length of the ground truth, ensuring a scale-invariant evaluation. The complement, (1-NED), provides a measure of accuracy, where a higher value indicates better recognition performance.

In the context of Arabic text recognition, the use of NED as an evaluation metric is less frequently documented in comparison to languages like English and Chinese. However, recent studies in Arabic scene text recognition and optical character recognition (OCR) have started adopting normalized edit distance metrics to evaluate models. This is particularly important for Arabic due to its unique challenges, such as cursive script, bidirectional writing, and diacritical marks, which make traditional accuracy metrics insufficient in capturing partial correctness

Recent studies in Arabic Optical Character Recognition (OCR) have increasingly adopted normalized edit distance (NED) metrics to evaluate model performance, reflecting a broader trend towards more nuanced accuracy assessments. For instance, M Alghamdi [12] discusses the importance of edit distance in determining OCR accuracy. The authors emphasize that edit distance, defined as the minimum number of operations needed to transform OCR output into the ground truth, serves as a meaningful measurement of an OCR system's performance. Similarly, MA Alghamdi [14] utilizes the Levenshtein distance, a form of edit distance, to measure the performance of Arabic OCR systems. This tool computes the minimum number of operations required to convert OCR output text into the ground truth, providing a detailed analysis of character-level accuracy.

4 Comparative Analysis of the Performance of Existing Methods

The survey by Mahdi et al. reviews deep learning methods for Arabic OCR, emphasizing CNNs, LSTMs, and hybrid models, and covers key datasets such as HACDB, AHCD, and Hijja. While it focuses on model architectures and recognition accuracy, our work extends this by analyzing scalability, cross-dataset generalization, and computational efficiency. We also provide a deeper examination of post-processing techniques critical for real-world deployment, including spell-checking, contextual analysis, and confidence scoring.

Also El Khayati et al. review CNN-based approaches for Arabic Handwriting Recognition, analyzing 62 studies across segmentation, character, digit, and word recognition. They categorize methods into simple CNNs, deep and hybrid models, transfer learning, and ensembles, highlighting the superiority of hybrid CNN-SVM and CNN-LSTM models. Key challenges such as cursive writing, diacritics, and vertical ligatures are discussed, along with segmentation techniques like FCNs, U-Net, and ARU-Net. The study also emphasizes the need for larger, high-quality datasets and suggests future improvements via augmentation and transfer learning.

ElSabagh et al. present a comprehensive survey on Arabic data augmentation (DA) techniques for NLP, addressing the challenges posed by the language's complex morphology, dialectal variations, and data scarcity. They analyze 75 primary and 9 secondary studies, categorizing DA methods into diversity-based, resampling, and secondary techniques. The survey details preprocessing strategies (e.g., diacritics and punctuation removal), feature extraction approaches (e.g., Word2Vec, AraVec, FastText), and downstream model training using both classical (SVM, NB) and deep models (CNNs, Transformers). Evaluation strategies, similarity metrics (e.g., BLEU, cosine), and the impact of DA on Arabic NLP tasks are also discussed. And Minoofam et al. proposes DB-QM, an analytical framework for evaluating Persian OCR databases. It introduces a structured classification of datasets and defines both structural (e.g., samples, fonts, resolution) and qualitative (e.g., integrity, diversity, legibility) evaluation criteria. DB-QM enables fair comparison and informed selection of OCR datasets, highlighting strengths, weaknesses, and challenges in Persian and Arabic database development.

4.1 Character Recognition Results

In our analysis, we navigate through diverse character recognition methods, scrutinizing their performance across benchmark datasets such as AHCD, Hijja, APTI, and Pat-A01. This examination serves to unravel the intricacies of recent studies' character recognition methodologies, emphasizing their distinctive approaches, dataset choices, and the performance metrics they achieve. The insights gleaned from this exploration are encapsulated in Table 8, providing a consolidated view of the comparative analysis of these character recognition methods.

In the realm of character recognition for Arabic OCR, several methodologies have been explored, each presenting distinct approaches and achieving varying degrees of success. Bouchakour et al. (2021) employed a combination of Hu and Gabor features along with a CNN classifier on the PAT-A01 dataset, achieving an accuracy of 97.23%.

Altwaijry and Al-Turaiki (2021) delved into the use of CNN on datasets like Hijja and AHCD, showcasing accuracies of 88% and 97%, respectively. Nayef et al. (2022) introduced a CNN architecture with OLReLU activation on the Hijja and AHCD datasets, demonstrating a character error rate (CER) of 52.9% for Hijja and 8.4% for AHCD, both coupled with 90% accuracy. Alwaqfi et al. (2022) adopted a unique strategy, integrating GANs with Deep CNN, achieving an impressive accuracy of 99.78% on the AHCD dataset. Alheraki et al. (2023) employed CNNs on both Hijja and AHCD datasets, achieving accuracies of 91% and 97%, respectively. Bilgin Tasdemir (2023) introduced a CNN-BLSTM-CTC architecture on the APTI dataset, resulting in a CER of 16%. Bin Durayhim et al. (2023) utilized CNN and VGG-16 architectures on Hijja and AHCD datasets, demonstrating accuracies ranging from 83% to 99%. These diverse methods highlight the ongoing exploration and innovation in character recognition, offering a spectrum of approaches to cater to the nuanced challenges of Arabic OCR.

AlShehri proposed DeepAHR, a CNN-based model for Arabic Handwritten Character Recognition (AHCR) with five convolutional layers and two fully connected layers, using LeakyReLU, batch normalization, and dropout to enhance performance and prevent overfitting. The model was trained on the AHCD and Hijaa datasets, achieving 98.66% and 88.24% accuracy, respectively. By optimizing hyperparameters and using the Nadam optimizer, DeepAHR outperformed state-of-the-art models.

Table 8. Character Recognition

Literature	Method	Dataset	CER	Accuracy	Precision	Recall	F1-score	Year
L Bouchakour[41]	Combined Hu and Gabor features + CNN classifier	PAT-A01	-	0.9723	-	-	-	2021
N Altwaijry[21]	CNN	Hijja	-	0.88	0.8788	0.8781	0.878	2021
N Altwaijry[21]	CNN	AHCD	-	0.97	0.9678	0.9673	0.9673	2021
BH Nayef[83]	CNN + OLReLU	Hijja	0.529	0.90	0.90	0.90	0.90	2022
BH Nayef[83]	CNN + OLReLU	AHCD	0.084	0.99	0.99	0.99	0.99	2022
YM Alwaqfi[22]	GANs + Deep CNN	AHCD	-	0.9978	-	-	-	2022
M Alheraki[15]	CNN	Hijja	-	0.91	0.91	0.91	0.91	2023
M Alheraki[15]	CNN	AHCD	-	0.97	0.97	0.97	0.97	2023
EF Bilgin Tasdemir [37]	CNN-BLSTM-CTC	APTI	0.16	-	-	-	-	2023
A Bin Durayhim[38]	CNN	Hijja	-	0.99	0.99	0.99	0.99	2023
A Bin Durayhim[38]	VGG-16	Hijja	-	0.83	0.85	0.83	0.83	2023
A Bin Durayhim[38]	CNN	AHCD	-	0.98	0.99	0.99	0.99	2023
A Bin Durayhim[38]	VGG-16	AHCD	-	0.94	0.95	0.94	0.94	2023
H AlShehri[20]	DeepAHR	AHCD	-	0.9866	0.9868	0.9866	0.9866	2024
H AlShehri[20]	DeepAHR	Hijja	-	0.8824	0.914	0.914	0.915	2024

Each method contributes uniquely, whether through the integration of advanced features, novel architectures, or sophisticated activation functions, providing valuable insights and paving the way for enhanced character recognition in Arabic text processing.

4.2 Words Recognition Results

In the domain of word recognition for Arabic OCR, we embark on an insightful journey, investigating a myriad of methodologies applied to various benchmark datasets, including IFN/ENIT, KHATT, AHDB, Alif, AcTiV, APTI, ACTIV, AHT2D, and IAM, detailed in Table 9. Word recognition stands as a pivotal component in the optical character recognition process, involving the interpretation and understanding of entire words within a document image. Our analysis peels back the layers of recent studies, shedding light on the unique approaches, dataset selections, and performance metrics achieved by different word recognition methods.

Awni et al. (2019) harnessed the power of an Ensemble of Residual Networks (ResNet18) on the IFN/ENIT dataset, resulting in a Word Error Rate (WER) of 6.63% and an accuracy of 93.37%. Jemni et al. (2019) explored a CNN-MDLSTM with Dynamic Lexicons (DWLD) on the KHATT dataset, achieving a WER of 20.83%. Eltay et al. (2020) introduced the BLSTM-CTC-WBS architecture on both IFN/ENIT and AHDB datasets, demonstrating

a WER of 1% and 1.9%, respectively, coupled with high accuracies. Butt et al. (2021) employed a CNN-RNN + Attention model on Alif and AcTiV datasets, showcasing recognition rates with precision (CRR), word (WRR), and line (LRR) of 97.09%, 79.91%, and 85.98%, respectively. Alzrrog et al. (2022) implemented DCNN on IFN/ENIT and AHWD datasets, achieving high accuracies of 99.76% and 99.39%, respectively. Salman and Altaei (2023) proposed a Multi-Font Arabic Word Recognition CNN, attaining a noteworthy accuracy of 96.77%. Ouali et al. (2023) introduced an Augmented Horizontal Text Detection (AHTD) based on CNN and Augmented Reality (AR) for datasets IFN/ENIT, ACTIV, and AHT2D, achieving varying precision scores. Hamida et al. (2023) explored k-NN on IFN/ENIT, exhibiting a high accuracy of 99.88%. Malakar et al. (2023) employed novel methods utilizing the Hausdorff and Fréchet distances for inter-segment similarity features in word recognition, showcasing high accuracies on IFN/ENIT and IAM datasets. Alwaqfi et al. (2023) introduced both a Hybrid GAN-based Model and DCNN on the APTI dataset, achieving remarkable accuracies of 99.76% and 94.81%, respectively. And Waly et al. propose an Arabic OCR system using CNNs and Transformers, achieving high accuracy on printed and handwritten text. Their pipeline combines DBNet++ detection, Transformer-based recognition, and real-world augmentations for robust document understanding. Also Chan et al. propose HATFORMER, a Transformer-based model for historical Arabic handwritten text recognition. It uses a custom image processor and tokenizer, with two-stage training on synthetic and real data. The model achieves state-of-the-art CER

Table 9. Word Recognition

Literature	Method	Dataset	C_{RR}	W_{RR}	L_{RR}	CER	WER	Accuracy	Precision	Recall	F1-score	Year
M Awni[28]	Ensemble of Residual Networks (ResNet18)	IFN/ENIT	-	-	-		0.0663	0.9337	-	-	-	2019
SK Jemni[63]	CNN-MDLSTM with Dynamic lexicons (DWLD)	KHATT	-	-	-	-	0.2083	-	-	-	-	2019
M Eltay[53]	BLSTM-CTC-WBS	IFN/ENIT	-	-	-	-	0.01	0.9899	-	-	-	2020
M Eltay[53]	BLSTM-CTC-WBS	AHDB	-	-	-	-	0.019	0.9810	-	-	-	2020
H Butt[43]	CNN-RNN + Attention	Alif	0.9709	0.7991	0.8598	-	-	-	-	-	-	2021
H Butt[43]	CNN-RNN + Attention	AcTiV	0.9071	0.6107	0.7564	-	-	-	-	-	-	2021
N Alzrrog[24]	DCNN	IFN/ENIT	-	-	-	-	-	0.9976	-	-	-	2022
N Alzrrog[24]	DCNN	AHWD	-	-	-	-	-	0.9939	-	-	-	2022
GJ Salman [99]	Multi-Font Arabic Word Recognition CNN	their own dataset	-	-	-	-	-	0.9677	-	-	-	2023
I Ouali [87]	AHTD based on CNN and Augmented Reality (AR)	IFN/ENIT	-	-	-	-	-	-	0.72	0.88	0.79	2023
I Ouali [87]	AHTD based on CNN and Augmented Reality (AR)	ACTIV	-	-	-	-	-	-	0.79	0.82	0.80	2023
I Ouali [87]	AHTD based on CNN and Augmented Reality (AR)	AHT2D	-	-	-	-	-	-	0.92	0.98	0.95	2023
S Hamida[59]	k-NN	IFN/ENIT	-	-	-	-	-	0.9988	0.9099	0.9790	-	2023
S Malakar[72]	Hausdorff and Fréchet distances for inter-segment similarity features in word recognition.	IFN/ENIT	-	-	-	-	-	0.9736	-	-	-	2023
S Malakar[72]	Hausdorff and Fréchet distances for inter-segment similarity features in word recognition.	IAM	-	-	-	-	-	0.9202	-	-	-	2023
YM Alwaqfi[23]	Hybrid GAN-based Model	APTI	-	-	-	-	-	0.9976	-	-	-	2023
YM Alwaqfi[23]	DCNN	APTI	-	-	-	-	-	0.9481	-	-	-	2023
A Waly [107]	Invizo (CNN + Transformer)	Online-KHATT	-	-	-	7.91	31.41	-	-	-	-	2023
A Waly [107]	Invizo (CNN + Transformer)	Printed Text	-	-	-	0.59	1.72	-	-	-	-	2023
A Chan [46]	HATFORMER	Muharaf	-	-	-	8.6%	-	-	-	-	-	2023
A Chan [46]	HATFORMER	KHATT	-	-	-	15.4%	-	-	-	-	-	2023

This comprehensive analysis unveils the dynamic landscape of word recognition in Arabic OCR, highlighting the richness and diversity of methods applied to tackle the intricacies posed by different datasets. Each method brings its unique strengths, and this exploration sets the stage for further advancements in the field.

4.3 Digits Recognition Results

In the domain of digits recognition for Arabic OCR, our analysis extends across various methodologies applied to distinct benchmark datasets. The datasets under scrutiny include AHDBase, HODA, ADBase, MADBase, and

their respective performance metrics are compiled in Table 10. our analysis navigates through recent studies, unraveling the distinct approaches, dataset choices, and performance metrics achieved by various digit recognition methods.

Al-wajih and Ghazali (2020) introduced a method combining sliding windows, random forests (RF), and support vector machine (SVM) on the AHDBase dataset, achieving an impressive accuracy of 98%. Al-wajih and Ghazali (2020) employed DTL + CNN on HODA and ADBase datasets, demonstrating high accuracies of 99.47% and 99.34%, respectively. Haghighi and Omranpour (2021) delved into the use of CNN and Bidirectional Long-Short Term Memory (BiLSTM) on the HODA dataset, resulting in a remarkable accuracy of 99.98%. Alkhawaldeh et al. (2021) proposed Ensemble Deep Transfer Learning (EDTL) on ADBase and MADBase datasets, achieving high accuracies of 99.83% and 99.78%, respectively. Ali et al. (2023) presented a Modified DCNN approach on the HODA dataset, showcasing an accuracy of 99.5% and demonstrating precision, recall, and F1-score values of 99.5%, 99.5%, and 99.45%, respectively.

Table 10. Digits Recognition

Literature	Method	Dataset	Accuracy	Precision	Recall	F1-score	Year
E Al-wajih[11]	sliding windows + random forests (RF) + support vector machine (SVM)	AHDBase	0.98	-	-	-	2020
YS Can[44]	DTL + CNN	HODA	0.9947	-	-	-	2020
YS Can[44]	DTL + CNN	ADBase	0.9934	-	-	-	2020
F Haghighi[57]	CNN and Bidirectional Long-Short Term Memory (BiLSTM)	HODA	0.9998	0.994	0.9938	-	2021
RS Alkhawaldeh[18]	Ensemble Deep Transfer Learning (EDTL)	ADBase	0.9983	-	-	-	2021
RS Alkhawaldeh[18]	Ensemble Deep Transfer Learning (EDTL)	MADBase	0.9978	-	-	-	2021
S Ali[17]	Modified DCNN	HODA	0.995	0.995	0.995	0.9945	2023

This in-depth exploration provides a comprehensive overview of the advancements in digit recognition for Arabic OCR, highlighting the efficacy of diverse methods across different datasets. These methodologies contribute significantly to the robustness and accuracy of digit recognition systems, paving the way for enhanced performance in practical applications.

4.4 Multifaceted Recognition Results

In the realm of multifaceted recognition within OCR, our exploration encompasses a spectrum of methodologies applied to diverse datasets, including HACDB, MADBase, SUST-ALT, KAFD, ADBase, AHCD, HIJJA and their distinctive characteristics are outlined in Table 11.

The multifaceted recognition paradigm involves handling different types of data, including characters, digits, words, and more. In the study by Ahmed et al. (2021), a Deep Convolutional Neural Network (DCNN) was employed for character recognition on the HACDB dataset, achieving a remarkable accuracy of 99.91%. The same methodology was applied to MADBase for digit recognition, yielding an accuracy of 99.906%. Additionally, on the SUST-ALT dataset focusing on word recognition, an accuracy of 99.952% was achieved. These findings showcase the adaptability and robustness of the applied DCNN approach across various facets of recognition. A unique approach was presented by Mortadi et al. (2023) using TrOCR on the KAFD dataset for word recognition. Despite a character error rate (CER) of 0.82 and a word error rate (WER) of 2.39, TrOCR demonstrated its potential for handling diverse textual content. Al-Barhamtoshy et al. (2023) introduced an innovative method involving the Fast Gradient Sign Method (FGSM) coupled with Optical Recognition Characterization (ORCing) for text recognition on ADBase. While the CER is not specified, the approach achieved a low WER of 0.0310, emphasizing

its efficacy in recognizing textual content. Mamoun (2023) presented a Combined CNN + SVM approach applied to HACDB for character recognition, achieving an accuracy of 89.7%. This methodology was further extended to AHCD, focusing on character recognition, and to HIJJA for word recognition, achieving accuracies of 97.3% and 88.8%, respectively.

Also, Mosbah et al. proposed ADOCRNet, a novel system combining CNNs for feature extraction, BLSTMs for sequence modeling, and CTC for output decoding. It achieved notable recognition rates on the P-KHATT, APTI, and IFN/ENIT datasets, with Character Error Rates (CERs) of 0.01% and 0.03%, and a Word Error Rate (WER) of 1.09%. The system surpasses existing models in handling mixed-font Arabic scripts, thanks to its hybrid architecture and the use of advanced data augmentation techniques like distortion, stretching, and perspective transformations.

Table 11. MultiFacets Recognition

Literature	Method	Dataset	DatabaseType	CER	WER	Accuracy	Precision	Recall	F1-score	Year
R Ahmed[4]	DCNN	HACDB	Characters	-	-	0.9991	0.96967	0.96967	0.96967	2021
R Ahmed[4]	DCNN	MADBase	Digits	-	-	0.99906	0.9953	0.9953	0.9953	2021
R Ahmed[4]	DCNN	SUST-ALT	Words	-	-	0.99952	0.99038	0.99038	0.99038	2021
A Mortadi[78]	TrOCR	KAFD	Words	0.82	2.39	-	-	-	-	2023
HM Al-Barhamtoshy[6]	FGSM + ORCing	ADBase	Text	-	0.0310	0.99	-	-	-	2023
M El Mamoun[74]	Combined CNN + SVM	HACDB	Characters	-	-	0.897	-	-	-	2023
M El Mamoun[74]	Combined CNN + SVM	AHCD	Characters	-	-	0.973	-	-	-	2023
M El Mamoun[74]	Combined CNN + SVM	HIJJA	Words	-	-	0.888	-	-	-	2023
L Mosbah[79]	ADOCRNet	P-KHATT	Characters	0.01	-	-	-	-	-	2024
L Mosbah[79]	ADOCRNet	APTI	Characters	0.03	-	-	-	-	-	2024
L Mosbah[79]	ADOCRNet	IFN/ENIT	Words	-	1.09	-	-	-	-	2024

The results demonstrate the adaptability of the approach across different facets of recognition, emphasizing its potential for multifaceted OCR applications.

4.5 Case Study Analysis: Arabic OCR on Receipt Line Segmentation Using the CORU Dataset

This case study examines Arabic OCR applied to receipt line segmentation using the CORU dataset[2], designed for multilingual receipt processing (Arabic and English). The analysis focuses on OCR accuracy for Arabic script extracted from segmented lines, addressing challenges such as cursive writing, diverse layouts, and multilingual content. The CORU dataset includes over 20,000 annotated receipts. The methodology involved:

- Preprocessing: Noise reduction and contrast enhancement to improve OCR performance.
- Line Segmentation: Custom tools were used to extract individual receipt lines, enabling precise OCR processing.
- OCR Application: A fine-tuned Arabic OCR model (CNN + BiLSTM) was applied to segmented lines for text extraction.

Challenges Addressed Cursive Writing: The model tackled the contextual variations of Arabic script effectively. Diverse Layouts: Adaptations ensured robust segmentation for different receipt structures. Multilingual Content: Mixed Arabic and English content required advanced OCR capabilities.

Results, Analysis, and Lessons Learned. **Line Segmentation Accuracy:** Segmentation successfully extracted specific lines: As shown in Figure 7

- Line 1: ليالي الشام تعتز بثقتكم (Item description)
- Line 2: التاريخ : 04:54:24 19/12/2022 م (Item description)

- Line 3: البيان الكمية السعر الاجمالي (Item description)
- Line 4: 25 25 1 سندوتش فلافل صاج (Item description)

OCR Results: The Character Error Rate (CER) is 7.83% and the Word Error Rate (WER) is 27.24%. lines were accurately recognized, with minor diacritical errors (البيان in Line 3) that did not impact overall interpretation.

Fig. 7. Examples of CORU dataset after Line Segmentation

The study illustrates the importance of segmentation and OCR adaptation for real-world applications. It confirms the potential of CORU to enhance automation in receipt digitization. This analysis demonstrates the value of OCR in automating Arabic receipt processing and validates the efficacy of the CORU dataset for multilingual applications.

5 Discussion and Conclusions

This comprehensive survey delves into the multifaceted landscape of Arabic Optical Character Recognition (OCR), shedding light on the diverse methodologies and datasets employed by researchers to enhance recognition rates. The key stages of the OCR process, such as preprocessing, segmentation (including text-area detection, line, word, and character segmentation), recognition, and postprocessing, are thoroughly examined in the literature review. The survey meticulously discusses the strengths and limitations associated with each technique, offering nuanced insights into their effectiveness.

An important revelation from the survey is the superior performance of segmentation-based approaches compared to segmentation-free methods, with techniques such as vertical/horizontal projection proving particularly effective in word and character segmentation. However, the survey acknowledges that the quality of OCR outcomes is intrinsically linked to the availability of suitable datasets. Notably, the accessibility of Arabic OCR datasets is limited, prompting a call for increased focus on postprocessing techniques. The incorporation and refinement of algorithms akin to Google's spelling checker are highlighted as particularly promising avenues, given their potential to substantially enhance the overall performance of OCR systems.

In conclusion, the survey provides a panoramic view of the current state-of-the-art in Arabic OCR, highlighting significant advancements in recent years. Despite these strides, the survey recognizes persistent challenges, including addressing the inherent variability and complexity of the Arabic script, accommodating diverse dialects, and handling language variations. The potential advantages of effective Arabic OCR systems are evident, propelling researchers and developers to dedicate ongoing efforts to further enhance and refine this technology. The survey optimistically anticipates the emergence of even more accurate and reliable Arabic OCR systems in the future, driven by continued research and development endeavors.

6 funding

This work was partially supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education under Grant 2020R1I1A3A04037680, and partly by the Innovative Human Resource Development for Local Intellectualization program through the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea Government [Ministry of Science and ICT (MSIT)] (IITP-2025-RS-2020-II201462, 50%).

References

[1] Sondos Aabed and Ahmad Khairaldin. 2024. An End-to-End, Segmentation-Free, Arabic Handwritten Recognition Model on KHATT. *arXiv preprint arXiv:2406.15329* (2024).

[2] Abdelrahman Abdallah, Mahmoud Abdalla, Mahmoud SalahEldin Kasem, Mohamed Mahmoud, Ibrahim Abdelhalim, Mohamed Elkasaby, Yasser ElBendary, and Adam Jatowt. 2024. CORU: Comprehensive Post-OCR Parsing and Receipt Understanding Dataset. *arXiv preprint arXiv:2406.04493* (2024).

[3] Hakim A Abdo, Ahmed Abdu, Ramesh R Manza, and Shobha Bawiskar. 2022. An approach to analysis of Arabic text documents into text lines, words, and characters. *Indonesian Journal of Electrical Engineering and Computer Science* 26, 2 (2022), 754–763.

[4] Rami Ahmed, Mandar Gogate, Ahsen Tahir, Kia Dashtipour, Bassam Al-Tamimi, Ahmad Hawalah, Mohammed A El-Affendi, and Amir Hussain. 2021. Novel deep convolutional neural network-based contextual recognition of Arabic handwritten scripts. *Entropy* 23, 3 (2021), 340.
[5] Saad Bin Ahmed, Saeeda Naz, Salahuddin Swati, and Muhammad Imran Razzak. 2019. Handwritten Urdu character recognition using one-dimensional BLSTM classifier. *Neural Computing and Applications* 31 (2019), 1143–1151.
[6] Hassanin M Al-Barhamtoshy, Kamal M Jambi, Mohsen A Rashwan, and Sherif M Abdou. 2023. An Arabic Manuscript Regions Detection, Recognition and Its Applications for OCRing. *Transactions on Asian and Low-Resource Language Information Processing* 22, 1 (2023), 1–28.
[7] Mansoor A Al Ghamdi. 2022. A novel approach to printed Arabic optical character recognition. *Arabian Journal for Science and Engineering* 47, 2 (2022), 2219–2237.
[8] Somaya Al-Ma'adeed, Dave Elliman, and Colin A Higgins. 2002. A data base for Arabic handwritten text recognition research. In *Proceedings eighth international workshop on frontiers in handwriting recognition*. IEEE, 485–489.
[9] Yousef Al-Ohali, Mohamed Cheriet, and Ching Suen. 2003. Databases for recognition of handwritten Arabic cheques. *Pattern Recognition* 36, 1 (2003), 111–121.
[10] I Saleh Al-Sheikh, MASNIZAH Mohd, and L Warlina. 2020. A review of Arabic text recognition dataset. *Asia-Pacific J. Inf. Technol. Multimedia* 9, 1 (2020), 69–81.
[11] Ebrahim Al-wajih and Rozaida Ghazali. 2020. Improving the accuracy for offline Arabic digit recognition using sliding window approach. *Iranian Journal of Science and Technology, Transactions of Electrical Engineering* 44 (2020), 1633–1644.
[12] Mansoor Alghamdi and William Teahan. 2017. Experimental evaluation of Arabic OCR systems. *PSU Research Review* 1, 3 (2017), 229–241.
[13] Mansoor Alghamdi and William Teahan. 2018. Printed Arabic script recognition: A survey. *International Journal of Advanced Computer Science and Applications* 9, 9 (2018).
[14] Mansoor A Alghamdi, Ibrahim S Alkhazi, and William J Teahan. 2016. Arabic OCR evaluation tool. In *2016 7th international conference on computer science and information technology (CSIT)*. IEEE, 1–6.
[15] Mais Alheraki, Rawan Al-Matham, and Hend Al-Khalifa. 2023. Handwritten Arabic Character Recognition for Children Writing Using Convolutional Neural Network and Stroke Identification. *Human-Centric Intelligent Systems* (2023), 1–13.
[16] Lutfieh S Alhomed and Kamal M Jambi. 2018. A survey on the existing arabic optical character recognition and future trends. *International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)* 7, 3 (2018), 78–88.
[17] Saqib Ali, Sana Sahiba, Muhammad Azeem, Zeeshan Shaukat, Tariq Mahmood, Zareen Sakhawat, and Muhammad Saqlain Aslam. 2023. A recognition model for handwritten Persian/Arabic numbers based on optimized deep convolutional neural network. *Multimedia Tools and Applications* 82, 10 (2023), 14557–14580.
[18] Rami S Alkhawaldeh, Moatsum Alawida, Nawaf Farhan Funkur Alshdaifat, Wafa'Za'al Alma'aitah, and Ammar Almasri. 2022. Ensemble deep transfer learning model for Arabic (Indian) handwritten digit recognition. *Neural Computing and Applications* 34, 1 (2022), 705–719.
[19] Naseem Alrobah and Saleh Albahli. 2022. Arabic handwritten recognition using deep learning: A Survey. *Arabian Journal for Science and Engineering* 47, 8 (2022), 9943–9963.
[20] Helala AlShehri. 2024. DeepAHR: a deep neural network approach for recognizing Arabic handwritten recognition. *Neural Computing and Applications* (2024), 1–13.
[21] Najwa Altwaijry and Isra Al-Turaiki. 2021. Arabic handwriting recognition system using convolutional neural network. *Neural Computing and Applications* 33, 7 (2021), 2249–2261.
[22] Yazan M Alwaqfi, Mumtazimah Mohamad, and Ahmad T Al-Taani. 2022. Generative Adversarial Network for an Improved Arabic Handwritten Characters Recognition. *International Journal of Advances in Soft Computing & Its Applications* 14, 1 (2022).
[23] Yazan M Alwaqfi, Mumtazimah Mohamad, Ahmad T Al-Taani, and Nazirah Abd Hamid. 2023. A Novel Hybrid DL Model for Printed Arabic Word Recognition based on GAN. *International Journal of Advanced Computer Science and Applications* 14, 1 (2023).
[24] Nori Alzrrog, Jean-François Bousquet, and Idris El-Feghi. 2022. Deep Learning Application for Handwritten Arabic Word Recognition. In *2022 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*. IEEE, 95–100.
[25] Jason Antonio, Aditya Rachman Putra, Harits Abdurrohman, and Moch Shandy Tsalasa. [n. d.]. A Survey on Scanned Receipts OCR and Information Extraction. ([n. d.]).
[26] Adil M Asiri and Mohammad S Khorsheed. 2005. Automatic Processing of Handwritten Arabic Forms using Neural Networks.. In *IEC (Prague)*. 313–317.
[27] Sameh M Awaidah and Sabri A Mahmoud. 2009. A multiple feature/resolution scheme to Arabic (Indian) numerals recognition using hidden Markov models. *Signal Processing* 89, 6 (2009), 1176–1184.
[28] Mohamed Awni, Mahmoud I Khalil, and Hazem M Abbas. 2019. Deep-learning ensemble for offline Arabic handwritten words recognition. In *2019 14th International Conference on Computer Engineering and Systems (ICCES)*. IEEE, 40–45.

[29] Mahmoud Badry, Hesham Hassan, Hanaa Bayomi, and Hussien Oakasha. 2018. QTID: Quran Text Image Dataset. *International Journal Of Advanced Computer Science And Applications* 9, 3 (2018).
[30] Mahmoud Badry, Mohammed Hassanin, Asghar Chandio, and Nour Moustafa. 2021. Quranic script optical text recognition using deep learning in IoT systems. *CMC-Comput. Mater. Contin* 68 (2021), 1847–1858.
[31] Muhammad Huzaifa Bashir, Aqil M Azmi, Haq Nawaz, Wajdi Zaghouani, Mona Diab, Ala Al-Fuqaha, and Junaid Qadir. 2023. Arabic natural language processing for Qur'anic research: A systematic review. *Artificial Intelligence Review* 56, 7 (2023), 6801–6854.
[32] Youssef Bassil and Mohammad Alwani. 2012. Ocr post-processing error correction algorithm using google online spelling suggestion. *arXiv preprint arXiv:1204.0191* (2012).
[33] Bilal Bataineh. 2017. A Printed PAW Image Database of Arabic Language for Document Analysis and Recognition. *Journal of ICT Research & Applications* 11, 2 (2017).
[34] Sonia Bergamaschi, Stefania De Nardis, Riccardo Martoglia, Federico Ruozzi, Luca Sala, Matteo Vanzini, and Riccardo Amerigo Vigliermo. 2022. Novel perspectives for the management of multilingual and multialphabetic heritages through automatic knowledge extraction: The digitalmaktaba approach. *Sensors* 22, 11 (2022), 3995.
[35] Gagan Bhatia, El Moatez Billah Nagoudi, Fakhraddin Alwajih, and Muhammad Abdul-Mageed. 2024. Qalam: A Multimodal LLM for Arabic Optical Character and Handwriting Recognition. *arXiv preprint arXiv:2407.13559* (2024).
[36] Aamna Bhatti, Ameera Arif, Waqar Khalid, Baber Khan, Ahmad Ali, Shehzad Khalid, and Atiq ur Rehman. 2023. Recognition and classification of handwritten urdu numerals using deep learning techniques. *Applied Sciences* 13, 3 (2023), 1624.
[37] Esma F Bilgin Tasdemir. 2023. Printed Ottoman text recognition using synthetic data and data augmentation. *International Journal on Document Analysis and Recognition (IJDAR)* (2023), 1–15.
[38] Anfal Bin Durayhim, Amani Al-Ajlan, Isra Al-Turaiki, and Najwa Altwaijry. 2023. Towards Accurate Children's Arabic Handwriting Recognition via Deep Learning. *Applied Sciences* 13, 3 (2023), 1692.
[39] Omar Ali Boraik, M Ravikumar, and Mufeed Ahmed Naji Saif. 2022. Characters Segmentation from Arabic Handwritten Document Images: Hybrid Approach. *International Journal of Advanced Computer Science and Applications* 13, 4 (2022).
[40] Manal Boualam, Youssef Elfakir, Ghizlane Khaissidi, and Mostafa Mrabti. 2022. Arabic handwriting word recognition based on convolutional recurrent neural network. In *WITS 2020: Proceedings of the 6th International Conference on Wireless Technologies, Embedded, and Intelligent Systems*. Springer, 877–885.
[41] Lallouani Bouchakour, Fariza Meziani, Houda Latrache, Khadija Ghribi, and Mustapha Yahiaoui. 2021. Printed Arabic Characters Recognition Using Combined Features and CNN classifier. In *2021 International Conference on Recent Advances in Mathematics and Informatics (ICRAMI)*. IEEE, 1–5.
[42] Hassina Bouressace. 2022. A Review of Arabic Document Analysis Methods. In *2022 4th International Conference on Pattern Analysis and Intelligent Systems (PAIS)*. IEEE, 1–7.
[43] Hanan Butt, Muhammad Raheel Raza, Muhammad Javed Ramzan, Muhammad Junaid Ali, and Muhammad Haris. 2021. Attention-based CNN-RNN Arabic text recognition from natural scene images. *Forecasting* 3, 3 (2021), 520–540.
[44] Yekta Said Can and M Erdem Kabadayı. 2020. Automatic cnn-based Arabic numeral spotting and handwritten digit recognition by using deep transfer learning in Ottoman population registers. *Applied Sciences* 10, 16 (2020), 5430.
[45] Fatma Chabchoub, Yousri Kessentini, Slim Kanoun, Veronique Eglin, and Frank Lebourgeois. 2016. SmartATID: A mobile captured Arabic Text Images Dataset for multi-purpose recognition tasks. In *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*. IEEE, 120–125.
[46] Adrian Chan, Anupam Mijar, Mehreen Saeed, Chau-Wai Wong, and Akram Khater. 2024. HATFormer: Historic Handwritten Arabic Text Recognition with Transformers. *arXiv preprint arXiv:2410.02179* (2024).
[47] Soumia Djaghbellou, Abderraouf Bouziane, Abdelouahab Attia, and Zahid Akhtar. 2021. A Survey on Arabic Handwritten Script Recognition Systems. *International Journal of Artificial Intelligence and Machine Learning (IJAIML)* 11, 2 (2021), 1–17.
[48] Iyad Abu Doush, Faisal Alkhateeb, and Anwaar Hamdi Gharaibeh. 2018. A novel Arabic OCR post-processing using rule-based and word context techniques. *International Journal on Document Analysis and Recognition (IJDAR)* 21 (2018), 77–89.
[49] Mohsine El Khayati, Ismail Kich, and Youssef Taouil. 2024. CNN-based Methods for Offline Arabic Handwriting Recognition: A Review. *Neural Processing Letters* 56, 2 (2024), 115.
[50] Ezzat Ali El-Sherif and Sherif Abdelazeem. 2007. A Two-Stage System for Arabic Handwritten Digit Recognition Tested on a New Large Database.. In *Artificial intelligence and pattern recognition*. 237–242.
[51] Mohsine Elkhayati, Youssfi Elkettani, and Mohammed Mourchid. 2022. Segmentation of handwritten arabic graphemes using a directed convolutional neural network and mathematical morphology operations. *Pattern Recognition* 122 (2022), 108288.
[52] Ahmed Adel ElSabagh, Shahira Shaaban Azab, and Hesham Ahmed Hefny. 2025. A comprehensive survey on Arabic text augmentation: approaches, challenges, and applications. *Neural Computing and Applications* (2025), 1–34.
[53] Mohamed Eltay, Abdelmalek Zidouri, and Irfan Ahmad. 2020. Exploring deep learning approaches to recognize handwritten arabic texts. *IEEE Access* 8 (2020), 89882–89898.

[54] Mohamed Eltay, Abdelmalek Zidouri, Irfan Ahmad, and Yousef Elarian. 2022. Generative adversarial network based adaptive data augmentation for handwritten Arabic text recognition. *PeerJ Computer Science* 8 (2022), e861.
[55] Moftah Elzobi and Ayoub Al-Hamadi. 2018. Generative vs. Discriminative Recognition Models for Off-Line Arabic Handwriting. *Sensors* 18, 9 (2018), 2786.
[56] Volkmar Frinken and Horst Bunke. 2014. Continuous Handwritten Script Recognition. In *Handbook of Document Image Processing and Recognition*, David Doermann and Karl Tombre (Eds.). Springer London, London, 391–425. doi:10.1007/978-0-85729-859-1_12
[57] Fatemeh Haghighi and Hesam Omranpour. 2021. Stacking ensemble model of deep learning and its application to Persian/Arabic handwritten digits recognition. *Knowledge-Based Systems* 220 (2021), 106940.
[58] Karez Hamad and Kaya Mehmet. 2016. A detailed analysis of optical character recognition technology. *International Journal of Applied Mathematics Electronics and Computers* Special Issue-1 (2016), 244–249.
[59] Soufiane Hamida, Bouchaib Cherradi, Oussama El Gannour, Abdelhadi Raihani, and Hassan Ouajji. 2023. Cursive Arabic handwritten word recognition system using majority voting and k-NN for feature descriptor selection. *Multimedia Tools and Applications* (2023), 1–25.
[60] Yanru He, Kejiang Chen, Guoqiang Chen, Zehua Ma, Kui Zhang, Jie Zhang, Huanyu Bian, Han Fang, Weiming Zhang, and Nenghai Yu. 2023. ProTegO: Protect Text Content against OCR Extraction Attack. In *Proceedings of the 31st ACM International Conference on Multimedia*. 7424–7434.
[61] Zheng Huang, Kai Chen, Jianhua He, Xiang Bai, Dimosthenis Karatzas, Shijian Lu, and CV Jawahar. 2019. Icdar2019 competition on scanned receipt ocr and information extraction. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 1516–1520.
[62] Noman Islam, Zeeshan Islam, and Nazia Noor. 2017. A survey on optical character recognition system. *arXiv preprint arXiv:1710.05703* (2017).
[63] Sana Khamekhem Jemni, Yousri Kessentini, and Slim Kanoun. 2019. Out of vocabulary word detection and recovery in Arabic handwritten text recognition. *Pattern Recognition* 93 (2019), 507–520.
[64] Naila Habib Khan and Awais Adnan. 2018. Urdu optical character recognition systems: Present contributions and future directions. *IEEE Access* 6 (2018), 46019–46046.
[65] Hossein Khosravi and Ehsanollah Kabir. 2007. Introducing a very large dataset of handwritten Farsi digits and a study on their varieties. *Pattern recognition letters* 28, 10 (2007), 1133–1141.
[66] Zohreh Khosrobeigi, Hadi Veisi, Ehsan Hoseinzade, and Hanieh Shabanian. 2022. Persian Optical Character Recognition Using Deep Bidirectional Long Short-Term Memory. *Applied Sciences* 12, 22 (2022), 11760.
[67] Ahmed Lawgali, Maia Angelova, and Ahmed Bouridane. 2013. HACDB: Handwritten Arabic characters database for automatic character recognition. In *European workshop on visual information processing (EUVIP)*. IEEE, 255–259.
[68] Hamzah Luqman, Sabri A Mahmoud, and Sameh Awaida. 2014. KAFD Arabic font database. *Pattern Recognition* 47, 6 (2014), 2231–2240.
[69] Mohamed G Mahdi, Ahmed Sleem, and Ibrahim Elhenawy. 2024. Deep Learning Algorithms for Arabic Optical Character Recognition: A Survey. *Multicriteria Algorithms with Applications* 2 (2024), 65–79.
[70] Sabri A Mahmoud, Irfan Ahmad, Wasfi G Al-Khatib, Mohammad Alshayeb, Mohammad Tanvir Parvez, Volker Märgner, and Gernot A Fink. 2014. KHATT: An open Arabic offline handwritten text database. *Pattern Recognition* 47, 3 (2014), 1096–1112.
[71] Sriparna Majumdar and Aaron Brick. 2022. Recognizing Handwriting Styles in a Historical Scanned Document Using Scikit-Fuzzy c-means Clustering. *arXiv preprint arXiv:2210.16780* (2022).
[72] Samir Malakar, Samanway Sahoo, Anuran Chakraborty, Ram Sarkar, and Mita Nasipuri. 2023. Handwritten Arabic and Roman word recognition using holistic approach. *The Visual Computer* 39, 7 (2023), 2909–2932.
[73] Rana Malhas and Tamer Elsayed. 2022. Arabic machine reading comprehension on the Holy Qur'an using CL-AraBERT. *Information Processing & Management* 59, 6 (2022), 103068.
[74] Mamouni El Mamoun. 2023. An Effective Combination of Convolutional Neural Network and Support Vector Machine Classifier for Arabic Handwritten Recognition. *Autom. Control Comput. Sci.* 57, 3 (jun 2023), 267–275. doi:10.3103/S0146411623030069
[75] Thomas Milo and Alicia Gonzalez Martinez. 2019. A new strategy for Arabic OCR: archigraphemes, letter blocks, script grammar, and shape synthesis. In *Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage*. 93–96.
[76] Seyyed Amir Hadi Minoofam, Azam Bastanfard, and Mohammad Reza Keyvanpour. 2025. DB-QM: A Comparative Quality Measurement and Its Prospective on Persian/Arabic Databases for OCR. *ACM Transactions on Asian and Low-Resource Language Information Processing* (2025).
[77] Emad Mohamed and Zeeshan Ali Sayyed. 2019. Arabic-SOS: segmentation, stemming, and orthography standardization for classical and pre-modern standard Arabic. In *Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage*. 27–32.
[78] Ahmad Mortadi, Ahmed Mohamed, Ahmed Talima, Ahmed Alkhattip, Ahmed Ibrahim, Ahmed Osman, and Yasser Hifny. 2023. ALNASIKH: An Arabic OCR System Based on Transformers. In *2023 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*. IEEE, 74–81.

[79] Lamia Mosbah, Ikram Moalla, Tarek M Hamdani, Bilel Neji, Taha Beyrouthy, and Adel M Alimi. 2024. ADOCRNet: A Deep Learning OCR for Arabic Documents Recognition. *IEEE Access* (2024).

[80] Aly Mostafa, Omar Mohamed, Ali Ashraf, Ahmed Elbehery, Salma Jamal, Ghada Khoriba, and Amr S Ghoneim. 2021. Ocformer: A transformer-based model for arabic handwritten text recognition. In *2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*. IEEE, 182–186.

[81] Aly Mostafa, Omar Mohamed, Ali Ashraf, Ahmed Elbehery, Salma Jamal, Anas Salah, and Amr S Ghoneim. 2022. An end-to-end ocr framework for robust arabic-handwriting recognition using a novel transformers-based model and an innovative 270 million-words multi-font corpus of classical arabic with diacritics. *arXiv preprint arXiv:2208.11484* (2022).

[82] Ismail B Mustapha, Shafaatunnur Hasan, Hatem Nabus, and Siti Mariyam Shamsuddin. 2022. Conditional deep convolutional generative adversarial networks for isolated handwritten arabic character generation. *Arabian Journal for Science and Engineering* 47, 2 (2022), 1309–1320.

[83] Bahera H Nayef, Siti Norul Huda Sheikh Abdullah, Rossilawati Sulaiman, and Zaid Abdi Alkareem Alyasseri. 2022. Optimized leaky ReLU for handwritten Arabic character recognition using convolution neural networks. *Multimedia Tools and Applications* (2022), 1–30.

[84] Saeeda Naz, Arif I Umar, Syed H Shirazi, Saad B Ahmed, Muhammad I Razzak, and Imran Siddiqi. 2016. Segmentation techniques for recognition of Arabic-like scripts: A comprehensive survey. *Education and Information Technologies* 21 (2016), 1225–1241.

[85] Clemens Neudecker, Konstantin Baierer, Mike Gerber, Christian Clausner, Apostolos Antonacopoulos, and Stefan Pletschacher. 2021. A survey of OCR evaluation tools and metrics. In *The 6th International Workshop on Historical Document Imaging and Processing*. 13–18.

[86] Thi Tuyet Hai Nguyen, Adam Jatowt, Mickael Coustaty, and Antoine Doucet. 2021. Survey of post-OCR processing approaches. *ACM Computing Surveys (CSUR)* 54, 6 (2021), 1–37.

[87] Imene Ouali, Mohamed Ben Halima, and Ali Wali. 2023. An augmented reality for an arabic text reading and visualization assistant for the visually impaired. *Multimedia Tools and Applications* (2023), 1–29.

[88] Mohammad T Parvez and Suliman A Alsuhibany. 2024. Challenges and opportunities for Arabic CAPTCHAs. *Multimedia Tools and Applications* 83, 5 (2024), 14047–14062.

[89] Mohammad Tanvir Parvez and Sabri A Mahmoud. 2013. Offline Arabic handwritten text recognition: a survey. *ACM Computing Surveys (CSUR)* 45, 2 (2013), 1–35.

[90] Mario Pechwitz, S Snoussi Maddouri, Volker Märgner, Noureddine Ellouze, Hamid Amiri, et al. 2002. IFN/ENIT-database of handwritten Arabic words. In *Proc. of CIFED*, Vol. 2. Citeseer, 127–136.

[91] Aziz Qaroush, Abdalkarim Awad, Abualsoud Hanani, Khader Mohammad, Basam Jaber, and Ala Hasheesh. 2022. Learning-free, divide and conquer text-line extraction algorithm for printed Arabic text with diacritics. *Journal of King Saud University-Computer and Information Sciences* 34, 9 (2022), 7699–7709.

[92] Aziz Qaroush, Bassam Jaber, Khader Mohammad, Mahdi Washaha, Eman Maali, and Nibal Nayef. 2022. An efficient, font independent word and character segmentation algorithm for printed Arabic text. *Journal of King Saud University-Computer and Information Sciences* 34, 1 (2022), 1330–1344.

[93] Jabril Ramdan, Khairuddin Omar, Mohammad Faidzul, and Ali Mady. 2013. Arabic handwriting data base for text recognition. *Procedia Technology* 11 (2013), 580–584.

[94] Dhuha Rashid and Naveen Kumar Gondhi. 2022. Scrutinization of Urdu Handwritten Text Recognition with Machine Learning Approach. In *Emerging Technologies in Computer Engineering: Cognitive Computing and Intelligent IoT: 5th International Conference, ICETCE 2022, Jaipur, India, February 4–5, 2022, Revised Selected Papers*. Springer, 383–394.

[95] Christian Reul, Dennis Christ, Alexander Hartelt, Nico Balbach, Maximilian Wehner, Uwe Springmann, Christoph Wick, Christine Grundig, Andreas Büttner, and Frank Puppe. 2019. OCR4all—An open-source tool providing a (semi-) automatic OCR workflow for historical printings. *Applied Sciences* 9, 22 (2019), 4853.

[96] Nazly Sabbour and Faisal Shafait. 2013. A segmentation-free approach to Arabic and Urdu OCR. In *Document recognition and retrieval XX*, Vol. 8658. SPIE, 215–226.

[97] My Abdelouahed Sabri, Assia Ennouni, and Abdellah Aarab. 2023. A Robust Approach for Arabic Document Images Segmentation and Indexation. In *International Conference on Digital Technologies and Applications*. Springer, 540–549.

[98] Khairun Saddami, Khairul Munadi, and Fitri Arnia. 2015. A database of printed Jawi character image. In *2015 Third International Conference on Image Information Processing (ICIIP)*. IEEE, 56–59.

[99] Ghufran Jafar Salman and Mohammed Sahib Mahdi Altaei. 2023. Proposed Deep Learning System for Arabic Text Detection and Recognition. In *2023 15th International Conference on Developments in eSystems Engineering (DeSE)*. IEEE, 39–44.

[100] Amarjot Singh, Ketan Bacchuwar, and Akshay Bhasin. 2012. A survey of OCR applications. *International Journal of Machine Learning and Computing* 2, 3 (2012), 314.

[101] Sukhjinder Singh, Naresh Kumar Garg, and Munish Kumar. 2023. On the performance analysis of various features and classifiers for handwritten devanagari word recognition. *Neural Computing and Applications* 35, 10 (2023), 7509–7527.

[102] Fouad Slimane, Rolf Ingold, Slim Kanoun, Adel M Alimi, and Jean Hennebert. 2009. Database and evaluation protocols for arabic printed text recognition. *DIUF-University of Fribourg-Switzerland* 1 (2009).
[103] Alaa Sulaiman, Khairuddin Omar, and Mohammad F Nasrudin. 2017. A database for degraded Arabic historical manuscripts. In *2017 6th International Conference on Electrical Engineering and Informatics (ICEEI)*. IEEE, 1–6.
[104] Ahmad P Tafti, Ahmadreza Baghaie, Mehdi Assefi, Hamid R Arabnia, Zeyun Yu, and Peggy Peissig. 2016. OCR as a service: an experimental evaluation of Google Docs OCR, Tesseract, ABBYY FineReader, and Transym. In *Advances in Visual Computing: 12th International Symposium, ISVC 2016, Las Vegas, NV, USA, December 12-14, 2016, Proceedings, Part I 12*. Springer, 735–746.
[105] Moeen Tayyab, Ayyaz Hussain, Mohammed Ali Alshara, Shakir Khan, Reemiah Muneer Alotaibi, and Abdul Rauf Baig. 2022. Recognition of Visual Arabic Scripting News Ticker From Broadcast Stream. *IEEE Access* 10 (2022), 59189–59204.
[106] Conrado Vizcarra, Shadan Alhamed, Abdulelah Algosaibi, Mohammed Alnaeem, Adel Aldalbahi, Nura Aljaafari, Ahmad Sawalmeh, Mahmoud Nazzal, Abdallah Khreishah, Abdulaziz Alhumam, et al. 2024. Deep learning adversarial attacks and defenses on license plate recognition system. *Cluster Computing* (2024), 1–18.
[107] Alhossien Waly, Bassant Tarek, Ali Feteha, Rewan Yehia, Gasser Amr, Walid Gomaa, and Ahmed Fares. 2025. Invizo: Arabic Handwritten Document Optical Character Recognition Solution. *arXiv preprint arXiv:2502.05277* (2025).
[108] Sonia Yousfi, Sid-Ahmed Berrani, and Christophe Garcia. 2015. ALIF: A dataset for Arabic embedded text recognition in TV broadcast. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 1221–1225.
[109] Oussama Zayene, Jean Hennebert, Sameh Masmoudi Touj, Rolf Ingold, and Najoua Essoukri Ben Amara. 2015. A dataset for Arabic text detection, tracking and recognition in news videos-AcTiV. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 996–1000.
[110] Oussama Zayene, Sameh Masmoudi Touj, Jean Hennebert, Rolf Ingold, and Najoua Essoukri Ben Amara. 2018. Open datasets and tools for arabic text detection and recognition in news video frames. *Journal of Imaging* 4, 2 (2018), 32.

Received 20 December 2023; revised 18 May 2025; accepted 12 August 2025