# A Systematic Literature Review of Deep Learning Methods for Handwritten Text Recognition in Historical Arabic Manuscripts

**Bilal Abdulrahman Tuama**
Department of Emergent Computing, Faculty of Computing, University of Technology Malaysia, Malaysia | Department of Pathological Analysis, College of Applied Science, University of Samarra, Iraq
tuamabilal@gmail.com (corresponding author)

**Farhan Mohamed**
Department of Emergent Computing, Faculty of Computing, University of Technology Malaysia, Malaysia
farhan@utm.my

*Received: 13 May 2025 | Revised: 2 June 2025 | Accepted: 16 June 2025*

 *DOI: https://doi.org/10.48084/etasr.12123*

## ABSTRACT

**Arabic Handwritten Text Recognition (AHTR) in historical manuscripts poses significant challenges due to the script's cursive nature, variability in calligraphy styles, and document degradation over time. This paper presents a Systematic Literature Review (SLR) of recent Deep Learning (DL)-based approaches applied to AHTR, focusing on methods developed between 2020 and 2025. It analyzes key DL architectures and provide a comparative analysis of the most commonly used datasets and segmentation strategies. Additionally, this review highlights essential preprocessing and postprocessing techniques that enhance the recognition performance and discusses the common evaluation metrics for AHTR. Finally, it identifies the current challenges and proposes future research directions to improve the recognition accuracy and model generalization. This review aims to guide researchers in building more robust and effective systems for the preservation and digitization of Arabic cultural heritage.**

***Keywords-Arabic handwritten text recognition; historical manuscripts; deep learning; segmentation***

## I. INTRODUCTION

Handwriting recognition has emerged as a major area of focus within the field of artificial intelligence. In today's digital age, it has attracted a substantial research interest for more than thirty years. Although the recognition technologies have advanced significantly, the handwriting analysis continues to be a fundamental classification challenge in AI. For example, identifying text often relies on geometric characteristics, such as shape, placement, and symmetry, whether horizontal, vertical, circular, or elliptical [1]. Optical Character Recognition (OCR) is a crucial component of computer vision. OCR is a field that empowers the translation of various document types and images into data that can be analyzed, edited, and searched [2].

Since 1973, Arabic has held the status of an official language of the United Nations. With more than 400 million native speakers globally [3], it is the primary language in 26 countries across the Middle East and North Africa, underlining its global significance. Beyond its native usage, Arabic plays a vital role in religious contexts, as over a billion Muslims use it in their daily worship. The Arabic script also serves as the writing system for several other languages, including Farsi, Kurdish, Urdu, and Jawi. Whether in print or handwritten form, Arabic text is widely encountered, highlighting its extensive reach and cultural importance [4]. Arabic is written and read from right to left, and unlike some other languages, it does not distinguish between uppercase and lowercase letters. However, the shape of each letter changes depending on its position within a word whether it appears at the beginning, middle, or end. Modern punctuation marks were not originally a part of the traditional Arabic writing. They were formally introduced in the early 20th century by Ahmed Zaki Pasha, who adapted them from European languages to suit Arabic writing. Prior to this, Arabic relied on linguistic tools, such as word order, conjunctions, and context to convey meaning, rather than using visual punctuation [5]. Short vowel marks, placed above or below letters, assist with proper pronunciation. These diacritics are typically found in the Qur'an to ensure accurate recitation, and in educational texts aimed at beginners. The Arabic alphabet consists of 18 basic letter shapes that represent 28 distinct phonetic sounds through the use of diacritical dots. For instance, the same base shape produces the sound "b" when one dot is placed below (ب), "t" when two dots are placed above (ت), and "th" with three dots above (ث). The form of an Arabic

letter varies based on its placement within a word; each letter can appear in as many as four distinct forms: isolated, at the last (joined on the right), at the first (joined on the left), or in the middle (joined on both sides). More examples can be seen in Table I. A key feature that distinguishes the Arabic script is its rich tradition of calligraphy. Arabic calligraphy is not merely a form of writing, but an esteemed artistic discipline that emphasizes the beauty, harmony, and flow in letterforms. Styles, such as Naskh, Thuluth, Diwani, and Ruqah, each have distinct visual characteristics, contributing to the cultural and historical depth of the Arabic handwriting. Some of Arabic calligraphy styles are illustrated in Figure 1. These calligraphic styles often feature highly cursive, interconnected letters and variable stroke thickness, which add aesthetic value but also pose significant difficulties for the automated recognition systems [6].

TABLE I. THE ARABIC ALPHABET

| Character | Isolated | First | Middle | Last |
|---|---|---|---|---|
| ALEF | أ | - | - | ـا |
| BAA | ب | بـ | ـبـ | ـب |
| TAA | ت | تـ | ـتـ | ـت |
| THAA | ث | ثـ | ـثـ | ـث |
| JEM | ج | جـ | ـجـ | ـج |
| HAA | ح | حـ | ـحـ | ـح |
| KHAA | خ | خـ | ـخـ | ـخ |
| DAL | د | - | - | ـد |
| ZAL | ذ | - | - | ـذ |
| RAA | ر | - | - | ـر |
| ZAI | ز | - | - | ـز |
| SEN | س | سـ | ـسـ | ـس |
| SHIN | ش | شـ | ـشـ | ـش |
| SAD | ص | صـ | ـصـ | ـص |
| DAD | ض | ضـ | ـضـ | ـض |
| TAH | ط | طـ | ـطـ | ـط |
| THAH | ظ | ظـ | ـظـ | ـظ |
| AIN | ع | عـ | ـعـ | ـع |
| GAIN | غ | غـ | ـغـ | ـغ |
| FAA | ف | فـ | ـفـ | ـف |
| QAF | ق | قـ | ـقـ | ـق |
| KAF | ك | كـ | ـكـ | ـك |
| LAM | ل | لـ | ـلـ | ـل |
| MEM | م | مـ | ـمـ | ـم |
| NON | ن | نـ | ـنـ | ـن |
| HAA | ه | هـ | ـهـ | ـه |
| WAW | و | - | - | ـو |
| YAA | ي | يـ | ـيـ | ـي |

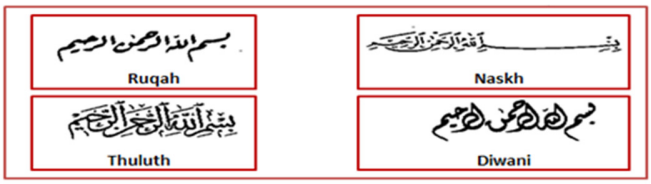


Fig. 1. Examples of Arabic calligraphy styles.

Recognizing handwritten text in historical Arabic manuscripts is challenging. The main challenges include the variability in handwriting styles, cursive script characteristics, contextual dependencies of letters, diacritical marks, and noise introduced by aging and environmental factors. Figure 2 presents an example of the historical Arabic manuscript.

Fig. 2. Sample image of historical Arabic manuscripts.

Traditional OCR methods often fail to deliver satisfactory results when applied to such manuscripts, as they are primarily designed for modern, clean, and printed text [7]. The need for an accurate and robust approach to transcribing these manuscripts is critical for preserving, analyzing, and making accessible a vast corpus of cultural, historical, and linguistic knowledge. These challenges are especially important in the fields of research and historic preservation. The extensive collection of Arabic manuscripts contains valuable cultural, scientific, and historical information. But manual transcription is labor intensive and time consuming. As a result, the access to these documents is limited and prevents a broader analysis [8].

This systematic review examines the DL methods applied to AHTR in historical Arabic manuscripts, aiming to evaluate the current progress, identify the stages of AHTR architecture, highlight challenges, and propose future research directions. The key contributions of the review include:

- Overview of DL architectures (Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), LSTMs, Transformers) used in AHTR especially in historical manuscripts.
- Analysis of preprocessing, segmentation, and postprocessing techniques in AHTR.
- Evaluating the available datasets by examining their size and data types to better understand their characteristics and suitability for the study.
- Highlighting the challenges and proposing future directions for the AHTR model development.

## II. METHODOLOGY

This section outlines the systematic approach adopted to identify, select, and analyze the relevant literature on DL methods for AHTR in historical Arabic.

### *A. Review Protocol*

This SLR follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 [9] guidelines, ensuring transparency, reproducibility, and rigor throughout the review process. PRISMA provides a structured approach to identifying, selecting, and critically evaluating relevant research, as well as systematically collecting and analyzing data from the involved studies. The review protocol was designed in four main stages:

### *B. Identification*

The current work conducted a comprehensive search across major academic databases including IEEE Xplore, SpringerLink, PubMed, and Google Scholar. The Boolean operators and specific keywords were used to maximize the retrieval of relevant studies. The search was limited to studies published between 2020 and 2025. The keywords for the search phrases used in the online databases that were considered and the number of the resulting papers are shown in Table II.

TABLE II. KEYWORDS FOR THE SEARCH STRINGS

| Search Keywords | Databases | Number of papers |
|---|---|---|
| ("Handwritten Arabic Text Recognition" OR "Arabic text recognition of manuscripts " OR "DL for Arabic manuscripts OCR") | Google Scholar | 173 |
| | PubMed | 13 |
| | Springer | 11 |
| | IEEE Xplore | 93 |

The search results vary across databases due to their different scopes. Google Scholar returned the most papers because of its broad coverage, while IEEE Xplore had a significant number focused on technical research. PubMed and Springer yielded fewer results, reflecting their more specialized areas. This highlights the importance of searching multiple databases for comprehensive coverage.

### *C. Screening*

Since browsing sites often provide the ability for direct filtering, in this work, only articles published between 2020 and 2025 were taken in consideration, resulting in 290 items overall. The titles and abstracts were then screened for relevance based on the inclusion criteria as follows:

- Studies focused on handwritten text recognition in historical Arabic manuscripts
- Published between 2020 and 2025
- Articles written in English
- Availability of full-text access

### *D. Eligibility*

Full-text articles meeting the inclusion criteria were assessed for methodological quality and relevance. Only those that demonstrated a clear application of DL approaches in the context of historical AHTR were retained for the final review.

### *E. Inclusion*

Studies that met all eligibility criteria were included in the final synthesis. The included papers were analyzed for information such as:

- Datasets and preprocessing methods
- DL architectures (e.g., CNN, RNN, Transformer)
- Post-processing techniques and evaluation metrics
- Identified challenges and research gaps

A PRISMA flow diagram depicted in Figure 3 was created to illustrate the study selection process, showing the number of records identified, screened, and excluded. A total of 31 research papers were, finally, included in the review.

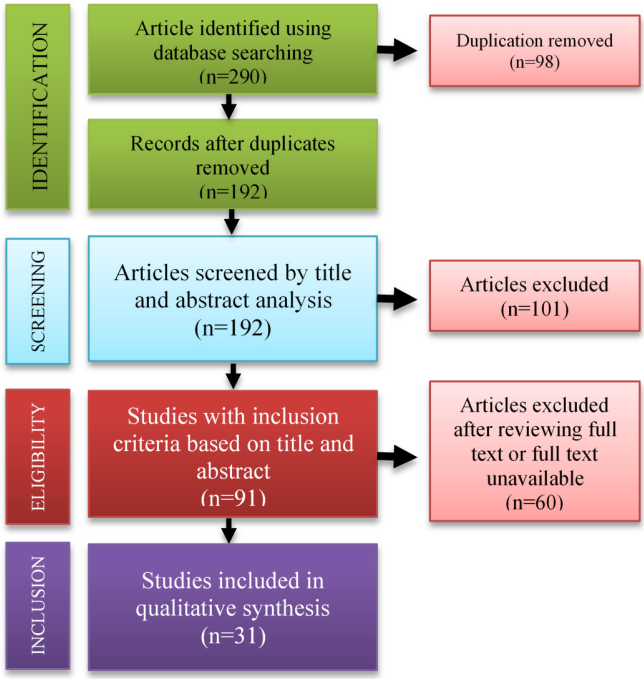


Fig. 3. PRISMA flow diagram.

### *F. Research Questions*

In this paper, a review of handwritten text recognition in old Arabic manuscripts was addressed. Thus, the following sentences provide a summary of the main research questions:

- Q1: What DL architectures have been applied to Arabic handwritten text recognition, particularly in historical manuscripts?
- Q2: What datasets are currently available for training and evaluating AHTR systems?
- Q3: What preprocessing techniques are commonly used to enhance the quality of historical Arabic texts before recognition?
- Q4: How is the performance of these methods evaluated, and which metrics are the most prevalent?
- Q5: What trends and future directions are emerging in this field?

## III. REVIEW FINDINGS

AHTR systems can be broadly classified based on two principles of data acquisition (online vs offline). This classification is essential for understanding the design, capabilities, and constraints of different recognition systems, particularly given the cursive and context-sensitive nature of the Arabic script. The taxonomy in Figure 4 serves as a

foundation for selecting or designing an appropriate AHTR solution, depending on the specific application domain, available resources, and data characteristics [10].

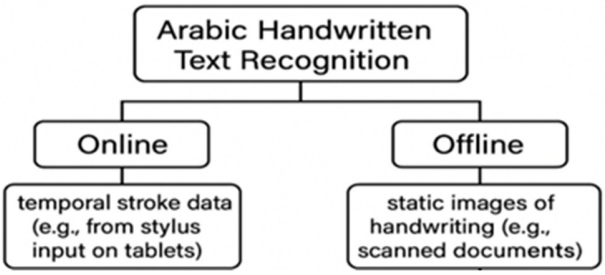


Fig. 4. Taxonomy of AHTR.

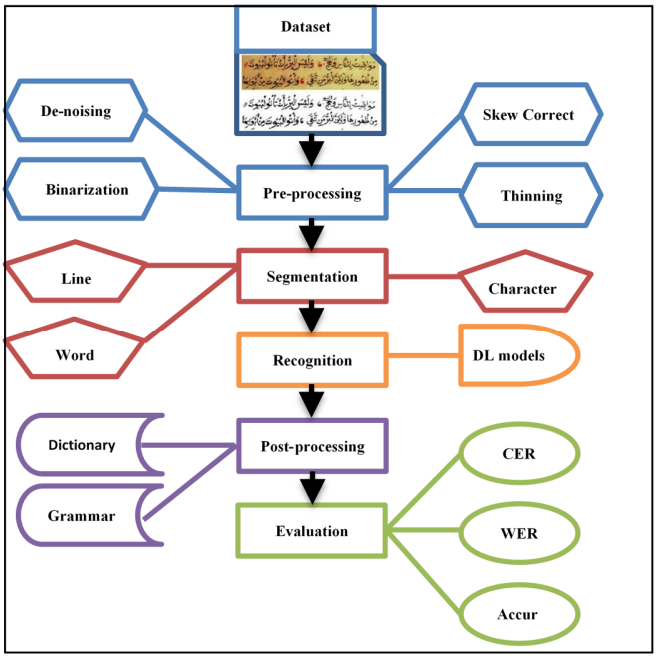


Fig. 5. General stages of AHTR.

Online AHTR systems operate by capturing handwriting in real-time using digitizing devices, such as styluses and tablets. These systems record dynamic stroke-level data including spatial coordinates, pressure, and timing information. The availability of temporal data significantly aids the disambiguation of complex Arabic character shapes and writing styles. Consequently, online systems often achieve higher accuracy, especially for cursive scripts. However, their practical deployment is limited by the need for specialized hardware and their inapplicability to the existing handwritten documents [10]. In contrast, offline AHTR systems function on static images of handwriting acquired from scanners or cameras. These systems rely solely on the visual representation of the text, without access to temporal or stroke-order information. While they are more broadly applicable for document archiving, historical manuscript analysis, and form, processing them is difficult due to the lack of temporal cues and the frequent occurrence of character overlap and ligatures in Arabic handwriting [11]. Although OCR generally focuses on identifying printed or handwritten text, AHTR is a specialized branch that tackles the unique difficulties of recognizing handwritten Arabic script. AHTR systems share a similar architectural structure with the traditional OCR systems, typically involving five key stages: pre-processing, segmentation, recognition, post-processing, and evaluation. Figure 5 provides a visual representation of the typical pipeline used in AHTR systems employing DL techniques [12].

### *A. Datasets*

Prior to outlining the stages of recognizing AHTR, it is essential to examine the existing datasets researchers rely on, especially those containing historical manuscripts, as summarized in Table III.

- Ibn Sina : The dataset is named after the renowned Persian scholar Ibn Sina, as it originates from one of his philosophical manuscripts. It comprises 60 pages and contains approximately 25,000 sub-words. The dataset includes around 1,200 distinct classes, with the sub-words distributed across these various categories [13].
- VML-HD: The dataset is derived from five historical books authored by different writers between the years 1088 and 1451. A total of 680 pages were selected from these manuscripts and meticulously annotated at the sub-word level. The dataset comprises 121,636 sub-word instances, representing 244,553 characters drawn from a vocabulary of 1,731 unique sub-word forms [14].
- The RASM: The dataset comprises 120 original pages from historical manuscripts written in Arabic, sourced from both Arab countries and regions with Arab or Muslim communities. The collection features diverse and intricate layouts, including margins, diagrams, tables, illustrations, and text lines, all of which exhibit rich stylistic and script variations [15].
- KALIMA: This dataset has been introduced to address the issue of word detection; it is constructed from 26 historical pages selected from the RASM2019 dataset and this collection includes images sourced from two distinct books [16].
- Muharaf: The dataset comprises 1,644 images of historical handwritten pages, which are meticulously transcribed by experts in archival Arabic. Each document is annotated with spatial polygon coordinates for text lines, along with key page elements. Muharaf features a diverse range of handwriting styles and document types, including personal letters, diaries, notes, poems, church records, and legal documents [17].
- KHATT: The KHATT database represents a comprehensive offline Arabic handwritten text resource compiled from 1,000 handwritten forms produced by writers from various countries. The dataset includes scanned images at multiple resolutions (200 dpi, 300 dpi, and 600 dpi), along with a manually validated ground truth containing metadata at the page, paragraph, and line levels. A structured verification process ensures alignment

between the handwritten content and its corresponding annotations [18].

- MADCAT: It is a large-scale resource developed to support the Arabic handwriting recognition and document analysis tasks. It consists of approximately 740,000 handwritten Arabic line images collected under controlled conditions, with specifications regarding the writing speed, implementation guidelines, and paper type. All images feature black ink text on clean, white backgrounds to ensure high visual clarity. Each handwritten sample is paired with detailed ground truth annotations, including transcriptions and layout information, making the dataset suitable for line-level recognition, machine translation, and document structure analysis [19].
- Naqsh: This dataset was generated using the Abstract Window Toolkit (AWT) graphics package in Java, with its textual content sourced from Arabic-language articles on Wikipedia. It is a line-synthesized Arabic OCR dataset that features 21 distinct calligraphy styles across various fonts, some of which include multiple weight variations. The dataset comprises line-level text images along with corresponding font labels. Each sample image is paired with its respective font name, making the dataset valuable for font recognition, OCR benchmarking, and script-style classification tasks [20].
- MOJ-DB: It is a database of Arabic historical manuscripts used for sub-word extraction. The database comprises 560,000 sub-words categorized into 5,600 distinct classes. It was constructed using 64 pages sourced from 10 books dating back to the 16th and 17th centuries [21].
- HICMA: The Handwriting Identification of Manuscripts and Calligraphy in Arabic (HICMA) dataset is a publicly accessible resource to provide authentic and diverse samples of Arabic handwritten text, encompassing both manuscript and calligraphic styles. It contains over 5,000 image-text pairs annotated with style labels, spanning five distinct writing styles. The dataset is designed to support handwriting identification and recognition tasks, and it includes benchmark results from state-of-the-art Arabic OCR models, providing a reference baseline for future research [22].

TABLE III. SUMMARY OF REVIEWED DATASETS

| Dataset | Size | Type |
|---|---|---|
| Ibn Sina [13] | 25000 | Sub-words |
| VML-HD [14] | 680 | Pages |
| RASM [15] | 120 | Pages |
| KALIMA [16] | 26 | Pages |
| Muharaf [17] | 1644 | Pages |
| KHATT [18] | 1000 | Handwritten forms |
| MADCAT [19] | 740000 | Line |
| Naqsh [20] | 132000 | Line |
| MOJ-DB [21] | 560000 | Sub-words |
| HICMA [22] | 5,000 | Line |

After outlining the primary datasets used in AHTR research, the technical pipeline is examined, beginning with preprocessing techniques aimed at improving the input quality.

## B. *Preprocessing*

Preprocessing is a critical step in AHTR pipelines, aimed at enhancing the quality of the input images and simplifying the recognition task. Especially in offline AHTR systems, where the input is a scanned or photographed image of handwritten text, preprocessing helps normalize the visual variations, reduce the noise, and standardize the input format. This section discusses four essential preprocessing techniques shown in Figure 6: skew correction, binarization, denoising, and thinning. In combination, skew correction, binarization, denoising, and thinning transform raw, degraded manuscript images into clean, standardized representations that are suitable for accurate recognition. Each step must be carefully designed and adapted to the unique challenges of Arabic handwriting and the specific historical conditions of the documents [4, 23].


Fig. 6. Preprocessing operation.

- De-skewing (Skew correction): It addresses the problem of tilted text lines, a common issue in scanned manuscripts due to imperfect document alignment or the natural slant of handwriting. In Arabic manuscripts, where the script flows continuously from right to left, even a small skew can disrupt the perception of word and character boundaries. Skew correction involves detecting the degree of tilt and rotating the image to restore horizontal alignment [24].
- Binarization: It is performed to simplify the image by converting it from grayscale or color into a binary format, where the pixels are either black (foreground) or white (background). Binarization is crucial because it isolates the handwritten text from the noisy background of old, stained, or damaged paper. Standard thresholding techniques work well in clean images, but historical Arabic manuscripts often require more sophisticated approaches, such as adaptive or local thresholding. These methods adjust the threshold dynamically based on the local properties of the

image, preserving the fine structures of thin strokes and faded characters that are common in ancient handwriting [23, 24].

- De-noising: Once the manuscript has been binarized, denoising becomes necessary to remove any unwanted artifacts that can mislead the recognition system. Historical documents frequently contain noise, such as ink bleed-through from the reverse side of the paper, spots, smudges, or degradation marks. Effective denoising aims to clean the image while preserving the fine details of the handwriting, which is especially delicate in Arabic due to the many subtle curves, dots, and diacritics [4, 23].
- Thinning: It is applied to the cleaned binary image. Thinning reduces the width of the handwritten strokes to a single pixel wherever possible, producing a skeletonized version of the text. In Arabic handwriting, where many characters are distinguished by subtle differences in curves and loops, thinning is particularly useful for simplifying the shapes while retaining their essential structure. This skeletonization not only facilitates feature extraction, but also improves the performance of subsequent recognition stages, such as character segmentation and classification [24].

### *C. Segmentation*

AHTR systems are divided segmentation-based and segmentation-free approaches [25]. Segmentation-based approaches decompose the input text into smaller units prior to recognition. This segmentation can occur at various levels, as shown in Figure 7, including lines [26], words [27, 30], or characters [31]. Line-level segmentation typically involves detecting text baselines and separating individual lines. Word-level segmentation divides the lines into discrete words, while character-level segmentation seeks to isolate the individual characters. Although segmentation can simplify the recognition task, it is particularly difficult in Arabic script due to its highly connected and variable character forms. Errors introduced during segmentation can severely degrade the recognition performance. On the other hand, segmentation-free (holistic) approaches avoid explicit segmentation altogether. These systems typically leverage DL models, CNNs for visual feature extraction, RNNs for sequence modeling, and Connectionist Temporal Classification (CTC) for alignment-free prediction. Holistic models are better suited for Arabic handwriting as they circumvent the ambiguity and error-prone nature of segmentation. However, they demand large annotated datasets and significant computational resources [32].

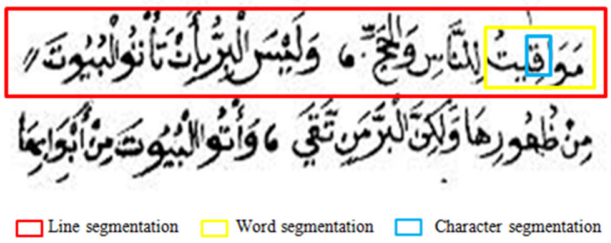


Fig. 7. Segmentation levels.

### *D. AHTR Techniques*

The literature on processing ancient and historical manuscripts is relatively limited compared to that on modern documents. Nevertheless, the techniques developed for modern document processing can still be applied to historical materials. However, processing these older manuscripts is often more challenging due to the typically poor quality of the available images. Authors in [26] proposed a method for recognizing the text from manuscript image using CNNs. The main stages of the proposed algorithm are: image enhancement through preprocessing, segmentation of the manuscript into lines and characters, creation of an Arabic character dataset, and text recognition using classification. In the classification phase, a CNN is applied to three datasets of images of the Arabic manuscripts from the Islamic period of $7^{th}$-$8^{th}$ Hijri centuries, achieving accuracy rates between 74.29% and 88.20%. Authors in [27] introduced a system which incorporates a novel feature extraction technique designed to derive robust features from Arabic words. Before feature extraction, each input image undergoes pre-processing and segmentation to isolate individual words. The features of each word or sub-word are then extracted using Multiscale Convexity Concavity (MCC) analysis of the word's contour shape. To reduce the computational overhead, the system replaces the commonly used Dynamic Time Warping (DTW) method with a circular shift technique for feature matching. The experimental results demonstrated that the proposed system achieves a strong balance between efficiency and accuracy, attaining a recognition rate of 96.6%.

Authors in [28] presented an end-to-end system to recognize Arabic handwritten sub-words in historical documents. They introduced a hybrid Convolutional Neural Network and Gated Recurrent Unit (CNN-GRU) model, where the shallow CNN learns robust feature representations while the GRU layers carry out the sequence modeling and generate the transcription of the text. The proposed system was evaluated on two different datasets, IBN SINA and VML-HD reporting recognition rates of 96.10% and 98.60% respectively.

Authors in [16] proposed an end-to-end approach for the offline recognition of ancient manuscripts. Initially, a critical preprocessing stage is carried out, where text lines and words are extracted by applying transfer learning to the You Only Look Once (YOLO) architecture, thereby addressing the segmentation task as a detection problem. For the recognition of historical handwritten words, ensemble learning techniques are employed, combining RNNs with a CTC layer and CNNs enhanced by Squeeze-and-Excitation blocks. High accuracy in detecting text lines and words is achieved, even in cases of overlapping or touching components, and the words with multiple connected elements are correctly identified. The method is evaluated using a collection of 20 pages for text line detection. Additionally, a new, consistent, and reliable dataset for word detection and recognition is introduced. Promising results are obtained, with F1-scores of 98.1% for the text line detection and 94.38% for word detection, along with a Character Error Rate (CER) of 8.27% in recognition.

A comprehensive framework was proposed in [29] for classifying ancient Arabic manuscripts into known and

unknown categories by processing their associated catalog cards. Due to the deterioration of these cards, traditional OCR methods are found to be insufficient. The framework addresses this limitation by employing DL techniques, including Faster R-CNN and other classification methods, to locate, extract, and classify textual information. The key processing steps involve region-of-interest detection, rotation correction, alignment, feature extraction, and classification. A dedicated dataset is compiled from various libraries and annotated with bounding boxes for relevant objects. The experimental results demonstrate that the proposed approach achieves a classification accuracy of 92.5%, outperforming classical image classification (83.5%) and OCR-based methods (81.5%), with a 10% improvement in accuracy over traditional techniques.

Authors in [30] tackle the challenge of handwritten Arabic word recognition in historical manuscripts, a field hindered by the scarcity of annotated training data. To bridge this gap, the researchers introduce a novel, word-level annotated dataset specifically curated for the detection and recognition of Arabic script in historical documents. The proposed methodology follows a two-stage pipeline: initially, a YOLO-based model is used to accurately detect the text lines and segment words, even in cases of overlap or contact. For the recognition phase, they explore and compare two DL architectures CNN-BLSTM with skip connections and a Transformer-based model, both integrated with an Arabic language model to enhance accuracy. These models demonstrate a strong performance validated on a 20-page manuscript dataset and the established IFN/ENIT benchmark. Fine-tuning the pre-trained models on IFN/ENIT data enabled high accuracy, achieving a CER of 4.8% with data augmentation, outperforming existing state-of-the-art methods despite the limited size of the proposed dataset.

Authors in [31] introduce HATFORMER, a Transformer-based encoder–decoder architecture adapted from a state-of-the-art English HTR model and tailored specifically for Arabic script. HATFORMER leverages the Transformer's attention mechanism to capture spatial and contextual information critical for distinguishing visually similar cursive characters and accurately identifying diacritics. The key components of the system include a Vision Transformer (ViT)-based image processor for efficient visual encoding, a custom Arabic text tokenizer for compact sequence representation, and a training pipeline optimized for low-resource historical data scenarios. Extensive experiments show that HATFORMER achieves a CER of 8.6% on the largest publicly available historical Arabic handwriting dataset—representing a 51% improvement over existing baselines. Moreover, it achieves a CER of 4.2% on the largest private modern Arabic dataset, demonstrating its adaptability across domains. These results validate the feasibility of adapting English HTR techniques to Arabic, and underscore HATFORMER's potential to advance document digitization, information retrieval, and the preservation of Arabic cultural heritage.

A comprehensive end-to-end OCR solution capable of recognizing Arabic handwritten text has been presented in [20]. For the recognition stage, the model integrates state-of-the-art CNN-based feature extraction with Transformer-based sequence modeling to effectively handle the variability in handwriting styles, stroke widths, alignments, and noisy inputs. The experimental results demonstrate the model's strong performance across both printed and handwritten Arabic texts. They achieved a precision of 81.66%, a recall of 78.82%, and an F-measure of 79.07% in the text detection task that underpins the proposed solution. On handwritten text, it records a CER of 7.91% and a WER of 31.41%, indicating solid robustness even in challenging conditions. The proposed system is designed for general-purpose applications and proves highly practical for real-world OCR tasks involving Arabic documents and receipts.

A novel OCR framework has been introduced in [33] that leverages transfer learning to enhance the recognition of ancient Arabic handwritten texts, addressing the critical need to preserve and digitize historically valuable manuscripts and rare books. This method was implemented and evaluated using the Tesseract OCR engine enhanced through transfer learning. The experimental results showed a significant boost in performance over the base model, achieving an average CER of 14.02% and Word Error Rate (WER) of 41.39%, with an overall accuracy of 87.89%. This work highlights the potential of adapting existing models to low-resource languages, paving the way for more advanced and scalable Arabic OCR solutions.

Authors in [34] investigated the use of advanced DL techniques for the restoration and recognition of ancient Arabic manuscript documents of profound cultural and historical value. At the core of the proposed approach lies the Modified Attention-Based Bidirectional Long Short-Term Memory (M-AB-LSTM) model, which demonstrated exceptional performance by achieving an accuracy rate of 99.50%. This model marks a significant advancement over traditional image enhancement methods, contributing not only to the digital accessibility of historical manuscripts, but also to their long-term preservation. Moreover, this work emphasizes the extensive effort required to construct high-quality datasets. A manually classified collection of 3,745 ancient Arabic manuscripts is presented, underscoring the complexity and dedication involved in preparing data suitable for DL applications. By combining technical innovation with cultural preservation, this research offers a meaningful contribution to the fields of heritage conservation, historical document analysis, and the development of specialized OCR solutions for Arabic script.

Finally, authors in [35] focus on developing a robust OCR system tailored for detecting and recognizing text in historical Arabic manuscripts. By leveraging Machine Learning (ML) models, particularly a CNN trained on a carefully curated dataset, the system aims to generate highly accurate, digitized text that supports an efficient analysis by historians, archaeologists, and researchers. The project contributes directly to the digitization of ancient Arabic scrolls and manuscripts through a comprehensive pipeline that includes preprocessing steps, such as binarization, dilation, erosion, and word and character segmentation. These steps prepare the manuscripts for effective text line extraction and recognition using a CNN-based OCR model, specifically designed for historical Arabic script. The implemented CNN model achieved an impressive

accuracy of 99% in recognizing historical Arabic characters, demonstrating the system's effectiveness and reliability. The aforementioned AHTR systems are summarized in Table IV.

TABLE IV. SUMMARY OF REVIEWED AHTR SYSTEMS

| Reference | Method | Dataset | Performance |
|---|---|---|---|
| [26] | CNN | (~5000) Historical images | Accuracy: 74.29%-88.20% |
| [27] | MCC | Ibn Sina | Recognition rate: 96:6% |
| [28] | Hybrid CNN-GRU | IBN SINA and VML-HD | Accuracy: 96.10% and 98.60% respectively |
| [16] | RNN with CTC | RASM 2019 and KALIMA | F1-score of 98.1% and 94.38, respectively |
| [29] | Region-based CNN | Total of 8478 images | Accuracy of 92.5% |
| [30] | Transformers and CNN-BLSTM | RASM2019, KALIMA, and IFN/ENIT | Best CER of 4.8% |
| [31] | HATFORMER | Muharaf, KHATT, and MADCAT | CER of 8.6%, 15.4%, and 4.2%, respectively |
| [20] | CNN-Transformers | KHATT and Naqsh Arabic | CER of 7.91% and WER of 31.41% |
| [33] | TL using Tesseract | Publicly available 25000 images | Accuracy of 87.89% |
| [34] | M-AB-LSTM | Collected ancient Arabic manuscripts (3745) | Accuracy of 99.50% |
| [35] | CNN | Images from MAKHTOTA library | Accuracy of 99 % |

### *E. Postprocessing Methods*

Postprocessing represents a phase in the AHTR workflow, aimed at enhancing the precision and overall quality of the extracted text. Postprocessing plays a vital role in refining the raw outputs of the recognition systems by correcting errors and enhancing the linguistic accuracy of the text. Due to the complexity of the Arabic script, such as the presence of similar-shaped characters, diacritics, ligatures, and rich morphology, postprocessing is essential for achieving high accuracy and usability of the recognized text [25, 36]. The most common post-processing techniques used in AHTR are displayed in Table V and include:

- Spell Checking: Compares the recognized text against a dictionary to detect and correct the spelling errors.
- Grammar Checking: Evaluates the grammatical structure of sentences using predefined linguistic rules, correcting errors, such as incorrect verb tenses, subject-verb agreement, and misplaced modifiers.
- Contextual Analysis: Examines the recognized text within its surrounding context to resolve ambiguities and correct misrecognized words [37].
- Language Model Integration: Incorporates statistical, ML, and DL-based language models to ensure that the recognized text adheres to the syntactic and semantic structure of the target language. Traditional statistical models and rule-based systems are combined with advanced ML and DL techniques—such as n-gram models, Hidden Markov Models (HMMs), Support Vector Machines (SVMs), and neural networks (e.g., RNNs, LSTMs, Transformers)—to detect and correct contextual errors, enhance coherence, and improve the overall text accuracy [38].

TABLE V. POSTPROCESSING TECHNIQUES OF REVIEWED AHTR

| Ref. | Spell checking | Grammar checking | Contextual analysis | LM integration |
|---|---|---|---|---|
| [26-29, 33, 34] | × | × | × | × |
| [35] | × | ✓ | × | × |
| [16] | ✓ | × | ✓ | × |
| [30] | × | × | ✓ | ✓ |
| [31] | × | × | ✓ | ✓ |
| [20] | × | × | ✓ | ✓ |

### *F. Evaluation Metrics in AHTR*

In the context of Arabic handwritten text recognition, evaluation is a crucial step for measuring the accuracy and quality of the OCR output. This process involves comparing the recognized text with the ground truth, the original text in the input image or document, and calculating various performance metrics to assess the effectiveness of the OCR system [39]. Common evaluation metrics include:

- CER: Measures the percentage of characters that are incorrectly recognized, inserted, or deleted compared to the reference text. This is particularly significant for Arabic, where small changes in the character shape can affect the meaning.
- WER: Calculates the proportion of words that are incorrectly recognized. It helps in understanding how errors at the word level impact the overall text comprehension.
- Accuracy: Represents the overall percentage of correctly recognized characters or words compared to the total number of characters or words in the ground truth. High accuracy indicates that the OCR system is generally performing well at recognizing the input text.
- Precision: Indicates the proportion of the correctly recognized text elements (characters or words) out of all elements identified by the OCR system.
- Recall: Measures the proportion of the correctly recognized elements out of all actual (ground-truth) elements.
- F1 Score: The harmonic mean of precision and recall, providing a balanced measure of OCR performance.

These metrics provide a comprehensive framework for evaluating AHTR systems, as shown in Table VI. They help developers identify weaknesses and areas for improvement, and they enable comparisons between different OCR models or techniques [40, 41].

TABLE VI. EVALUATION METRICS OF REVIEWED AHTR

| Ref. | Accuracy | WER or CER | Precision | Recall | F1 score |
|---|---|---|---|---|---|
| [26-29, 34, 35] | ✓ | × | × | × | × |
| [33] | ✓ | ✓ | × | × | × |
| [16] | × | ✓ | × | × | ✓ |
| [30, 31] | × | ✓ | × | × | × |
| [20] | × | ✓ | ✓ | ✓ | ✓ |

While the reviewed methods show a promising performance, several persistent challenges hinder the broader applicability and generalization across historical Arabic manuscripts.

## IV. CHALLENGES

AHTR is a complex task within OCR due to both the intrinsic characteristics of the Arabic script and the practical limitations of working with historical documents [42]. These challenges can be summarized into the following points:

- Dot Placement: Arabic letters are distinguished by the number and position of dots (e.g., ب، ت، ث), which can appear above or below the baseline. Minor errors in skew correction can result in the misplacement or removal of dots, drastically changing the word meanings.
- High Cursiveness: Unlike non-cursive scripts (e.g., printed Latin), Arabic handwriting exhibits a continuous character flow, causing segmentation ambiguity and overlapping strokes. Incorrect segmentation severely reduces the recognition accuracy.
- Font and Style Variation: Each font (e.g., Naskh, Ruq'ah, Thuluth), as portrayed in Figure 1, displays different levels of ligature, stroke thickness, and alignment. The OCR models trained on one style often perform poorly on others.
- Handwriting Irregularities: Differences in slant, pressure, and spacing among individual writers or even within a single manuscript contribute to recognition challenges.
- Historical Document Degradation: Manuscripts often suffer from ink bleed-through, stains, tears, and fading, all of which can obscure or distort the text regions.
- Baseline Instability: Unlike modern printed text, historical handwriting lacks a clear and consistent baseline, making it hard to align lines for segmentation.
- Domain Gap: The models trained on synthetic or modern handwriting datasets (e.g., Naqsh, KHATT) frequently underperform when applied to real historical texts.

## V. CONCLUSION

This paper covers Deep Learning (DL)-based systems for Arabic Handwritten Text Recognition (AHTR) from historical manuscripts. These systems vary in terms of architecture (e.g., Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Transformer), recognition granularity (character, word, sub-word), and preprocessing strategies. The models in [26, 27, 35] use handcrafted or simplified features in conjunction with classical Machine Learning (ML) or CNN classifiers. These models emphasize the Classical Feature Engineering (Multiscale Convexity Concavity (MCC)) or straightforward CNN classifiers). While fast and computationally efficient, they typically lack robustness across diverse styles unless paired with strong preprocessing and domain-specific tuning. The hybrid CNN-RNN Models with Sequential Modeling used in [28, 16, 20] offer strong performance due to their ability to model sequential dependencies in complex cursive scripts. The use of You Only Look Once (YOLO) for segmentation before recognition in [30] is particularly helpful in degraded or overlapping texts. Transformer-based methods like HATFORMER [31] show strong generalization and domain adaptability, especially when pretraining is involved. These architectures outperform traditional RNNs when large annotated data or language models are available. Transfer Learning and specialized Optical Character Recognition (OCR) frameworks in [29, 33-34] excel in low-resource conditions or specific tasks. Transfer learning helps mitigate the challenge of scarce annotated data, while specialized architectures deliver state-of-the-art accuracy in curated settings.

Building upon the identified limitations in the current AHTR research, several directions can be pursued to advance the field both in theory and application. First and foremost, there is a need to develop larger and more diverse annotated datasets that capture the full spectrum of Arabic handwriting styles across different historical periods, regions, and calligraphic traditions. Collaborations between computational researchers and manuscript archives or libraries could facilitate the digitization and annotation of rare and complex texts, enabling a broader coverage and better model generalization.

Exploring self-supervised and semi-supervised learning techniques is another area of research, particularly to reduce the reliance on costly manual annotation. Techniques, such as masked image modeling, contrastive learning, and pretraining on unlabeled historical corpora, can help build more resilient models capable of adapting to unseen handwriting styles. Finally, future work could benefit from the development of robust evaluation frameworks and standardized benchmarks that account not only for recognition accuracy, but also for aspects, such as semantic correctness, historical fidelity, and usability in real-world archival applications.

## REFERENCES

[1] C. C. Tappert, C. Y. Suen, and T. Wakahara, "The state of the art in online handwriting recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 8, pp. 787–808, Aug. 1990, https://doi.org/10.1109/34.57669.

[2] B. Kada, A. Mohammed, and B. Abdelmajid, "An Optimized Approach for Handwritten Arabic Character Recognition based on the SVM Classifier," *Engineering, Technology & Applied Science Research*, vol. 15, no. 2, pp. 22232–22238, Apr. 2025, https://doi.org/10.48084/etasr.9292.

[3] K. Mohd, A. Adnan, A. Yusof, M. Ahmad, and M. Mohd Kamal, *Teaching Arabic Language to Malaysian University Students using Education Technologies based on Education 4.0 Principles.* International Invention, Innovative & Creative (InIIC), 2019.

[4] H. M. Balaha, H. A. Ali, and M. Badawy, "Automatic recognition of handwritten Arabic characters: a comprehensive review," *Neural Computing and Applications*, vol. 33, no. 7, pp. 3011–3034, Apr. 2021, https://doi.org/10.1007/s00521-020-05137-6.

[5] Metropolitan Museum of Art (New York, N.Y.), M. Ekhtiar, and C. Moore, Eds., *Art of the Islamic world: a resource for educators*. New York: The Metropolitan Museum of Art, 2012.

[6] D. Alashari and Mohd Azhar Abd. Hamid, "A Systematic Review on Arabic Calligraphy within Islamic Architecture," *Ulum Islamiyyah*, vol. 33, no. 1, pp. 1–15, Apr. 2021, https://doi.org/10.33102/uij.vol33no1.263.

[7] L. Mosbah, I. Moalla, T. M. Hamdani, B. Neji, T. Beyrouthy, and A. M. Alimi, "ADOCRNet: A Deep Learning OCR for Arabic Documents Recognition," *IEEE Access*, vol. 12, pp. 55620–55631, 2024, https://doi.org/10.1109/ACCESS.2024.3379530.

[8] M. Khayyat and L. Elrefaei, "A Deep Learning Based Prediction of Arabic Manuscripts Handwriting Style," *The International Arab Journal of Information Technology*, vol. 17, no. 5, pp. 702–712, Sep. 2020, https://doi.org/10.34028/iajit/17/5/3.

[9] M. J. Page *et al.*, "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *BMJ*, Mar. 2021, Art. no. n71, https://doi.org/10.1136/bmj.n71.

[10] M. S. Kasem, M. Mahmoud, and H.-S. Kang, "Advancements and Challenges in Arabic Optical Character Recognition: A Comprehensive Survey." arXiv, 2023, https://doi.org/10.48550/ARXIV.2312.11812.

[11] A. Nori, "Deep Learning Applications to Offline Arabic Handwriting Words Recognition Using Convolutional Neural Network," Jan. 2023.

[12] S. Faizullah, M. S. Ayub, S. Hussain, and M. A. Khan, "A Survey of OCR in Arabic Language: Applications, Techniques, and Challenges," *Applied Sciences*, vol. 13, no. 7, Apr. 2023, Art. no. 4584, https://doi.org/10.3390/app13074584.

[13] R. Farrahi Moghaddam, M. Cheriet, M. M. Adankon, K. Filonenko, and R. Wisnovsky, "IBN SINA: a database for research on processing and understanding of Arabic manuscripts images," in *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*, Boston Massachusetts USA, Jun. 2010, pp. 11–18, https://doi.org/10.1145/1815330.1815332.

[14] M. Kassis, A. Abdalhaleem, A. Droby, R. Alaasam, and J. El-Sana, "VML-HD: The historical Arabic documents dataset for recognition systems," in *2017 1st International Workshop on Arabic Script Analysis and Recognition (ASAR)*, Nancy, France, Apr. 2017, pp. 11–14, https://doi.org/10.1109/ASAR.2017.8067751.

[15] C. Clausner, A. Antonacopoulos, N. Mcgregor, and D. Wilson-Nunn, "ICFHR 2018 Competition on Recognition of Historical Arabic Scientific Manuscripts – RASM2018," in *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, Niagara Falls, NY, USA, Aug. 2018, pp. 471–476, https://doi.org/10.1109/ICFHR-2018.2018.00088.

[16] B. Hakim and B. Ahror, "Text line and word detection and recognition of historical Arabic manuscripts." In Review, May 05, 2023, https://doi.org/10.21203/rs.3.rs-2883455/v1.

[17] M. Saeed *et al.*, "Muharaf: Manuscripts of Handwritten Arabic Dataset for Cursive Text Recognition," 2024, https://doi.org/10.48550/ARXIV.2406.09630.

[18] S. A. Mahmoud *et al.*, "KHATT: Arabic Offline Handwritten Text Database," in *2012 International Conference on Frontiers in Handwriting Recognition*, Bari, Italy, Sep. 2012, pp. 449–454, https://doi.org/10.1109/ICFHR.2012.224.

[19] Lee, David, Ismael, Safa, Grimes, Stephen, Doermann, Dave, Strassel, Stephanie, and Chen, Song, "MADCAT Phase 1 Training Set." Linguistic Data Consortium, Sep. 17, 2012, Art. no. 7771127 KB, https://doi.org/10.35111/9BM5-NZ55.

[20] A. Waly *et al.*, "Invizo: Arabic Handwritten Document Optical Character Recognition Solution." arXiv, 2025, https://doi.org/10.48550/ARXIV.2502.05277.

[21] A. Zoizou, A. Zarghili, and I. Chaker, "MOJ-DB: A new database of Arabic historical handwriting and a novel approach for subwords extraction," *Pattern Recognition Letters*, vol. 159, pp. 54–60, Jul. 2022, https://doi.org/10.1016/j.patrec.2022.04.040.

[22] A. Ismail, Z. Kamel, and R. Mahmoud, "HICMA: The Handwriting Identification for Calligraphy and Manuscripts in Arabic Dataset," in *Proceedings of ArabicNLP 2023*, Singapore (Hybrid), 2023, pp. 24–32, https://doi.org/10.18653/v1/2023.arabicnlp-1.3.

[23] H. M. Al-Barhamtoshy, K. M. Jambi, M. A. Rashwan, and S. M. Abdou, "An Arabic Manuscript Regions Detection, Recognition and Its Applications for OCRing," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 22, no. 1, pp. 1–28, Jan. 2023, https://doi.org/10.1145/3532609.

[24] H. M. Al-Barhamtoshy and S. M. Abdou, "Arabic Manuscripts Alignment, Segmentation, Recognition, and Classification," *ACM Transactions on Asian and Low-Resource Language Information Processing*, Apr. 2025, Art. no. 3732939, https://doi.org/10.1145/3732939.

[25] O. Elharrouss, S. Al-Maadeed, J. M. Alja'am, and A. Hassaine, "A Robust Method for Text, Line, and Word Segmentation for Historical Arabic Manuscripts," in *Data Analytics for Cultural Heritage*, A. Belhi, A. Bouras, A. K. Al-Ali, and A. H. Sadka, Eds. Cham: Springer International Publishing, 2021, pp. 147–172.

[26] B. Alrehali, N. Alsaedi, H. Alahmadi, and N. Abid, "Historical Arabic Manuscripts Text Recognition Using Convolutional Neural Network," in *2020 6th Conference on Data Science and Machine Learning Applications (CDMA)*, Riyadh, Saudi Arabia, Mar. 2020, pp. 37–42, https://doi.org/10.1109/CDMA47397.2020.00012.

[27] S. Elaiwat and M. Abu-Zanona, "Arabic Word Recognition System for Historical Documents using Multiscale Representation Method," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 4, 2020, https://doi.org/10.14569/IJACSA.2020.01104107.

[28] H. Hassen, S. Al-Madeed, and A. Bouridane, "Subword Recognition in Historical Arabic Documents using C-GRUs," *TEM Journal*, pp. 1630–1637, Nov. 2021, https://doi.org/10.18421/TEM104-19.

[29] L. S. Al-homed, K. M. Jambi, and H. M. Al-Barhamtoshy, "A Deep Learning Approach for Arabic Manuscripts Classification," *Sensors*, vol. 23, no. 19, Sep. 2023, Art. no. 8133, https://doi.org/10.3390/s23198133.

[30] B. Hakim and B. Ahror, "Deep Learning for Accurate Recognition of Arabic Handwritten Words in Historical Documents.," *Procedia Computer Science*, vol. 244, pp. 57–65, 2024, https://doi.org/10.1016/j.procs.2024.10.178.

[31] A. Chan, A. Mijar, M. Saeed, C.-W. Wong, and A. Khater, "HATFormer: Historic Handwritten Arabic Text Recognition with Transformers." arXiv, 2024, https://doi.org/10.48550/ARXIV.2410.02179.

[32] S. Aabed and A. Khairaldin, "An End-to-End, Segmentation-Free, Arabic Handwritten Recognition Model on KHATT." arXiv, 2024, https://doi.org/10.48550/ARXIV.2406.15329.

[33] S. Faizullah, M. S. Ayub, T. Alghamdi, T. S. Ali, M. A. Khan, and E. Nabil, "Revolutionizing Historical Document Digitization: LSTM-Enhanced OCR for Arabic Handwritten Manuscripts," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 10, 2024, https://doi.org/10.14569/IJACSA.2024.01510120.

[34] K. Miloud, M. L. Abdelmounaim, B. Mohammed, and B. R. Ilyas, "Restoration of ancient Arabic manuscripts: a deep learning approach," *STUDIES IN ENGINEERING AND EXACT SCIENCES*, vol. 5, no. 2, Sep. 2024, Art. no. e7722, https://doi.org/10.54021/seesv5n2-183.

[35] A. Alqahtani and S. Alfahmi, "System for Detection and Recognition of Historical Arabic Manuscripts." In Review, Feb. 03, 2025, https://doi.org/10.21203/rs.3.rs-5936450/v1.

[36] R. Najam and S. Faizullah, "Analysis of Recent Deep Learning Techniques for Arabic Handwritten-Text OCR and Post-OCR Correction," *Applied Sciences*, vol. 13, no. 13, Jun. 2023, Art. no. 7568, https://doi.org/10.3390/app13137568.

[37] A. Mostafa *et al.*, "OCFormer: A Transformer-Based Model For Arabic Handwritten Text Recognition," in *2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, Cairo, Egypt, May 2021, pp. 182–186, https://doi.org/10.1109/MIUCC52538.2021.9447608.

[38] S. Momeni and B. BabaAli, "A Transformer-based Approach for Arabic Offline Handwritten Text Recognition." arXiv, 2023, https://doi.org/10.48550/ARXIV.2307.15045.

[39] S. Djaghbellou, A. Attıa, and A. Bouzıane, "A Survey on Text-Line Segmentation in Arab Historical Manuscripts," *International Journal of Informatics and Applied Mathematics*, vol. 7, no. 1, pp. 14–32, Jun. 2024, https://doi.org/10.53508/ijiam.1407236.

[40] S. Alghyaline, "Arabic Optical Character Recognition: A Review," *Computer Modeling in Engineering & Sciences*, vol. 135, no. 3, pp. 1825–1861, 2023, https://doi.org/10.32604/cmes.2022.024555.

[41] E. El-Awadly, A. Ebada, and A. Al-Zoghby, "Arabic Handwritten Text Recognition Systems and Challenges and Opportunities," *The Egyptian Journal of Language Engineering*, vol. 10, no. 2, pp. 84–103, Oct. 2023, https://doi.org/10.21608/ejle.2023.193993.1043.

[42] H. M. Al-Barhamtoshy, K. M. Jambi, S. M. Abdou, and M. A. Rashwan, "Arabic Documents Information Retrieval for Printed, Handwritten, and Calligraphy Image," *IEEE Access*, vol. 9, pp. 51242–51257, 2021, https://doi.org/10.1109/ACCESS.2021.3066477.