

From Script to Digital: A Deep Learning Approach to Arabic Handwriting Recognition

Hamza Ahmed Abushahla*

*Department of Computer Science and Engineering
American University of Sharjah
Sharjah, United Arab Emirates
b00090279@aus.edu*

Layth Al-Khairulla*

*Department of Computer Science and Engineering
American University of Sharjah
Sharjah, United Arab Emirates
b00087225@aus.edu*

Ariel Justine Navarro Panopio does this work*

*Department of Computer Science and Engineering
American University of Sharjah
Sharjah, United Arab Emirates
b00088568@aus.edu*

Alex Aklson†

*Department of Computer Science and Engineering
American University of Sharjah
Sharjah, United Arab Emirates
aaklson@aus.edu*

Abstract—Handwritten Text Recognition (HTR) for Arabic script is crucial for enabling digital accessibility and automating the conversion of handwritten documents into searchable digital formats. The cursive nature of Arabic script, with its positional letter shapes and diacritical marks, presents significant challenges that require specialized recognition systems. These challenges are compounded by the variability in handwriting styles and the limitations of techniques developed for other languages. In this study, we leverage the KHATT dataset to develop an end-to-end deep learning-based HTR system. Our approach effectively addresses the complexities of Arabic cursive handwriting using a segmentation-based model. This work demonstrates the potential of deep learning in advancing Arabic HTR, enabling the digitization of both contemporary and historical texts and supporting broader applications in cultural preservation and digital workflows.

Keywords—Arabic Handwriting Recognition, KHATT Dataset, Arabic Handwriting, Optical Character Recognition, Deep Learning, Cursive Text Recognition

I. INTRODUCTION

The Arabic language is the official language of 24 sovereign countries and is spoken by over 400 million people worldwide. It is also one of the six official languages of the United Nations, reflecting its international importance [?]. Beyond its widespread use, Arabic holds profound cultural, literary, and religious significance, serving as a cornerstone of the heritage of Arab and Muslim communities globally. This extensive reach emphasizes the importance of developing accurate Arabic handwritten text recognition (HTR) systems, which have diverse applications in educational, governmental, and cultural preservation contexts. For instance, the ability to accurately convert both historical and contemporary handwritten Arabic documents into digital text is vital for the digitization of archives, streamlining administrative processes, and supporting large-scale data analysis efforts.

Recognizing Arabic cursive handwriting presents unique challenges due to the inherent complexities of the script. Arabic is written from right to left, with letters assuming different shapes based on their positional context within words. Diacritical marks, known as “حركات” (harakat), and other special symbols such as the “همزة” (hamza) and “مَدَّة” (madda) further complicate recognition, as many letters share identical base shapes but are distinguished by one or more dots placed above or below the character [?]. These characteristics, combined with the variability in individual handwriting styles, make the task of Arabic handwriting recognition particularly demanding [?]. Furthermore, because of these unique challenges, techniques developed for recognizing handwriting in other languages—such as Latin-based scripts—cannot be directly applied to Arabic, necessitating tailored approaches and specialized models.

Traditional Arabic HTR systems have predominantly relied on shallow learning techniques. These methods typically involve handcrafted feature extraction processes that are sensitive to noise and degradation, limiting their effectiveness and scalability [?]. Furthermore, most existing systems adopt a segmentation-based approach, requiring the segmentation of words into individual characters before recognition. This segmentation process is particularly challenging for Arabic due to the cursive connections between letters and the presence of similar character shapes, making it difficult to accurately isolate each character [?], [?]. In contrast, segmentation-free models (holistic approaches) recognize words as whole-word images without any segmentation processes, which can be more effective for specific applications with limited vocabularies [?], [?].

Deep learning has revolutionized HTR, offering superior performance and generalizability compared to traditional approaches. Convolutional Neural Networks (CNNs) have emerged as the foundation of modern HTR systems due to their ability to extract complex spatial features, making them

*These authors contributed equally to this work.

†Author to whom correspondence should be addressed.

particularly effective for cursive scripts like Arabic [?], [?]. Combined with Bidirectional Long Short-Term Memory (BiLSTM) networks and Connectionist Temporal Classification (CTC), deep learning models can effectively handle sequence modeling for continuous text recognition [?]. More recently, Transformers and attention-based models have introduced new paradigms [?], [?], [?], enabling HTR systems to focus on specific regions of an image and capture long-range dependencies. These advances have significantly improved accuracy and generalizability, particularly for complex and diverse handwriting styles. Additionally, the integration of language models during post-processing has further improved prediction accuracy by leveraging contextual language information [?].

To advance the development of robust HTR systems for Arabic, leveraging comprehensive and diverse datasets is paramount. While historical datasets like Muharaf [?] have provided valuable resources for analyzing historical manuscripts, there is a growing need to shift focus towards more general and contemporary datasets to ensure broader applicability. In this study, we adopt the KHATT dataset [?], a benchmark dataset renowned in the field of Arabic handwriting recognition. The KHATT dataset encompasses a wide range of modern handwritten Arabic text, covering diverse writing styles from different regions and writers. This shift from historical to contemporary data allows for the development of more generalized HTR systems capable of handling both historical and modern handwriting styles.

This study presents a segmentation-based HTR model utilizing advanced deep learning techniques to achieve high accuracy in Arabic handwriting recognition. The proposed system employs ResNet as a backbone for feature extraction, followed by a BiLSTM-CTC architecture for sequence modeling. To further enhance performance, a language model is incorporated during the post-processing phase, improving prediction accuracy by leveraging contextual information. In general, our contributions can be summarized as follows:

- We develop a CNN-based deep learning approach combined with BiLSTM-CTC and a language model to accurately recognize Arabic handwritten text.
- We design a robust preprocessing pipeline to address handwriting variability and optimize input data for recognition.
- We rigorously evaluate our system using the KHATT dataset, demonstrating its effectiveness across diverse handwriting styles.

The remainder of this paper is organized as follows: Section 2 provides the background, including an overview of HTR, the unique characteristics of Arabic handwriting, and essential terminology. Section 3 reviews related work, covering traditional methods, advancements in deep learning, and the importance of datasets in Arabic handwriting recognition. Section 4 outlines our methodology, detailing the proposed solution, algorithms, and techniques used to address the problem. Section 5 presents the results and evaluation, discussing the findings and their implications. Finally, Section 6 concludes the study

by summarizing key contributions, addressing limitations, and proposing directions for future research.

II. BACKGROUND

A. Handwritten Text Recognition

Optical Character Recognition (OCR) laid the foundation for modern text recognition systems by enabling the automated extraction of text from printed documents. This field gained momentum in the 1990s [?], with early neural network models like LeNet [?], showcasing significant promise in character classification tasks. These systems primarily targeted machine-printed text and laid the groundwork for extending recognition capabilities to handwritten text, evolving into what is now known as HTR.

HTR has progressed significantly since its early focus on isolated character-level recognition [?], which remains prevalent for logographic languages like Japanese [?] and Chinese [?]. For alphabetical languages such as English and Arabic, handwriting recognition expanded to word-level tasks, where single words are transcribed from handwritten images [?], [?]. More advanced techniques have since moved towards line-level HTR, where entire lines of text, including spaces, are transcribed. Line-level HTR can either rely on pre-segmented input or integrate segmentation and recognition into a unified framework. In recent developments, systems also tackle paragraph- and page-level recognition, incorporating layout analysis for handling complex document structures.

Furthermore, recent advancements focus on line-level HTR, where the aim is to transcribe entire text lines, including spaces, which were often omitted in word-level systems. Line-level recognition can be performed on pre-segmented text [23]–[27] or integrated into joint detection and recognition frameworks where both line segmentation and transcription are done at the same [29].

More advanced systems now address paragraph or page-level recognition [28], combining layout analysis techniques such as paragraph and line segmentation [30]–[33].

The sequential nature of handwriting makes its recognition uniquely challenging. Characters within words are connected, influenced by their context, and require a holistic understanding of the sequence for accurate transcription. Cutting-edge HTR algorithms leverage recurrent architectures, such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, which are designed to capture and model these sequential dependencies effectively. These architectures have been instrumental in advancing HTR by processing input as a series of interconnected data points rather than isolated units, enabling robust handling of cursive and complex scripts.

Among recurrent architectures, Multidimensional Long Short-Term Memory (MDLSTM) networks have emerged as a powerful tool for HTR. Unlike standard LSTMs, which operate along a single axis, MDLSTM networks extend recurrence along two axes, making them particularly effective for two-dimensional inputs like line images in HTR. They excel in

extracting features from line-level handwritten text, converting 2D data into 1D sequences for transcription. This capability lies at the core of many state-of-the-art line-level HTR systems. However, MDLSTMs are computationally intensive compared to Convolutional Neural Networks (CNNs), which have become the go-to architecture for efficient and scalable feature extraction.

Recent innovations combine CNNs with RNNs or MDLSTMs to capitalize on their complementary strengths. For example, CNNs extract spatial features from input images, which are then processed by recurrent layers to capture the sequential dependencies of text. These hybrid architectures have achieved significant breakthroughs, with deformable convolutions further enhancing CNN-based models by allowing the convolutional kernel to adapt geometrically. Such approaches address the variability in handwriting styles, reducing transcription errors and setting new benchmarks in HTR accuracy.

Advances in deep learning have driven much of the progress in HTR. Recurrent Neural Networks (RNNs) and their multidimensional variant, MDLSTM (Multidimensional Long Short-Term Memory), are particularly well-suited for sequential data and two-dimensional inputs. MDLSTM networks introduce recurrence along two axes, making them highly effective for line-level HTR, where 2D data is converted into 1D sequences for character transcription. However, their computational complexity has shifted the focus toward Convolutional Neural Networks (CNNs), which offer a balance of efficiency and accuracy. CNNs extract features from images, often in combination with RNNs, to generate robust text predictions. Recent innovations, such as deformable convolutions, have further enhanced CNN-based models by allowing kernels to adapt geometrically, addressing the variability in handwriting styles. These advancements have established CNNs and hybrid architectures as state-of-the-art approaches in HTR.

The foundation of cutting-edge HTR algorithms lies in recurrent architectures, such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM), which capture the sequential nature of text. Over the last decade, CNNs and, more recently, Transformers have gained prominence for their superior feature extraction and context-awareness capabilities.

B. Characteristics of Arabic Handwriting

Characteristics of Arabic handwriting: cursive nature, positional letter shapes, and diacritics.

C. Challenges in Arabic HTR

Challenges in Arabic HTR and its historical significance.
Importance of datasets for advancing HTR research.

III. RELATED WORK

In this section, we review previous research on (), including traditional methods, advancements in deep learning, and the role of Arabic handwriting datasets in advancing the field.

Review of existing HTR systems for Arabic text. Limitations of traditional feature-based approaches in HTR. Advances in deep learning for handwriting recognition. Datasets

of Arabic Handwriting: Overview of existing datasets: WAHD, KHATT, and Balamand. Introduction of the Muharaf dataset, its features, and its advantages.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

A.

B.

C. Datasets of Arabic Handwriting

Several datasets have been developed for Arabic handwriting research, (). These datasets differ in focus, size, and availability. Below, we summarize the most relevant ones including WAHD, KHATT, Balamand, and Muharaf.

- **WAHD Dataset [?]:**

The WAHD dataset is the first dataset explicitly designed for writer analysis tasks in Arabic historical documents. It consists of 353 manuscripts, 333 from the Islamic Heritage Project (IHP) and 20 from the National Library

in Jerusalem. Written by 302 writers (23 of them identified), it includes 2,313 pages authored by 11 scribes contributing multiple manuscripts. WAHD is freely available, making it a valuable resource for historical handwriting analysis.

- **KHATT Dataset [?]:**

The KHATT dataset is a modern (non-historical) Arabic handwriting dataset comprising 4,000 paragraphs written by 1,000 scribes, with six paragraphs contributed by each. It was created mainly for writer identification and was developed jointly by researchers from KFUPM (Saudi Arabia), TU Dortmund (Germany), and TU Braunschweig (Germany). However, its restricted access limits its usability, as it is not publicly available.

- **Balamand Dataset [?]:**

The Balamand dataset contains 567 historic manuscripts collected from 14 repositories in Lebanon and Syria, including Antiochian Orthodox monasteries and bishoprics, which have been digitized at the University of Balamand. The dataset, spanning the 13th to the 19th century, identifies 256 copyists who produced 329 manuscripts. However, it is not publicly available.

- **Muharaf Dataset [?]:**

The Muharaf dataset, used in this study, is the largest publicly available Arabic dataset with fully annotated and transcribed historical manuscripts at the text-line level. It includes 1,644 pages (1,216 public and 428 restricted), spanning the early 19th to the early 21st century, with 36,311 text lines (24,495 public). Line-level images, stored in PNG format, were generated using line warping software to create consistent horizontal grids, and extensive metadata is provided in JSON format.

Although Muharaf was primarily explored for handwriting text recognition (HTR) in previous works [?], [?], our project utilizes it for writer identification, leveraging the existing writer labels (about 25% of the public portion). The dataset contains a rich diversity of handwriting styles, including writings by Arab Levantines living in America, as well as contributors from various other regions. This mix of historical and contemporary writing makes Muharaf a valuable resource for writer identification.

IV. METHODOLOGY

A. Data Preparation and Preprocessing

To prepare the data for model training, we utilized the filtered line-by-line images. The following preprocessing steps were applied to ensure consistency and enhance the model's performance...

- Resizing, Normalization, Augmentation

B. Supervised Learning

1) *Model Architecture:* We implemented a CNN-based architecture...

2) *Model Training and Evaluation:* The model was trained using a categorical cross-entropy loss function and the Adam optimizer. Early stopping was employed to prevent overfitting.

V. RESULTS AND DISCUSSION

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

VI. CONCLUSIONS AND FUTURE WORK

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et

neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

REFERENCES

- [1] N. Altwaijry and I. Al-Turaiki, "Arabic handwriting recognition system using convolutional neural network," *Neural Computing and Applications*, vol. 33, no. 7, pp. 2249–2261, 2021.
- [2] M. T. Parvez and S. A. Mahmoud, "Offline arabic handwritten text recognition: a survey," *ACM Computing Surveys (CSUR)*, vol. 45, no. 2, pp. 1–35, 2013.
- [3] L. Mosbah, I. Moalla, T. M. Hamdani, B. Neji, T. Beyrouthy, and A. M. Alimi, "Adocnet: A deep learning ocr for arabic documents recognition," *IEEE Access*, 2024.
- [4] N. Alrobah and S. Albahli, "Arabic handwritten recognition using deep learning: A survey," *Arabian Journal for Science and Engineering*, vol. 47, no. 8, pp. 9943–9963, 2022.
- [5] M. Li, T. Lv, J. Chen, L. Cui, Y. Lu, D. Florencio, C. Zhang, Z. Li, and F. Wei, "Trocr: Transformer-based optical character recognition with pre-trained models," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 11, 2023, pp. 13 094–13 102.
- [6] T. Wang, Y. Zhu, L. Jin, C. Luo, X. Chen, Y. Wu, Q. Wang, and M. Cai, "Decoupled attention network for text recognition," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 07, 2020, pp. 12 216–12 224.
- [7] A. K. Bhunia, S. Ghose, A. Kumar, P. N. Chowdhury, A. Sain, and Y.-Z. Song, "Metaht: Towards writer-adaptive handwritten text recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 15 830–15 839.
- [8] M. Saeed, A. Chan, A. Mijar, J. Moukarzel, G. Habchi, C. Younes, A. Elias, C.-W. Wong, and A. Khater, "Muharaf: Manuscripts of handwritten arabic dataset for cursive text recognition," *arXiv preprint arXiv:2406.09630*, 2024.
- [9] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [10] N. D. Cilia, C. De Stefano, F. Fontanella, and A. S. di Freca, "A ranking-based feature selection approach for handwritten character recognition," *Pattern Recognition Letters*, vol. 121, pp. 77–86, 2019.
- [11] T. Clanuwat, A. Lamb, and A. Kitamoto, "Kuronet: Pre-modern japanese kuzushiji character recognition with deep learning," in *2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2019, pp. 607–614.
- [12] M. Jaderberg, K. Simonyan, A. Zisserman *et al.*, "Spatial transformer networks," *Advances in neural information processing systems*, vol. 28, 2015.
- [13] A. K. Bhunia, A. Das, A. K. Bhunia, P. S. R. Kishore, and P. P. Roy, "Handwriting recognition in low-resource scripts using adversarial learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4767–4776.
- [14] F. P. Such, D. Peri, F. Brockler, H. Paul, and R. Ptucha, "Fully convolutional networks for handwriting recognition," in *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*. IEEE, 2018, pp. 86–91.
- [15] A. Abdelhaleem, A. Droby, A. Asi, M. Kassis, R. Al Asam, and J. El-sanaa, "Wahd: a database for writer identification of arabic historical documents," in *2017 1st International workshop on arabic script analysis and recognition (ASAR)*. IEEE, 2017, pp. 64–68.
- [16] S. A. Mahmoud, I. Ahmad, W. G. Al-Khatib, M. Alshayeb, M. T. Parvez, V. Märgner, and G. A. Fink, "Khatt: An open arabic offline handwritten text database," *Pattern Recognition*, vol. 47, no. 3, pp. 1096–1112, 2014.
- [17] A. Chan, A. Mijar, M. Saeed, C.-W. Wong, and A. Khater, "Hatformer: Historic handwritten arabic text recognition with transformers," *arXiv preprint arXiv:2410.02179*, 2024.