Khazanah Journal of Religion and Technology
Universitas Islam Negeri Sunan Gunung Djati Bandung, Indonesia
https://journal.uinsgd.ac.id/kjrt
Online ISSN: 2987-6060
Volume 3 Issue 1 (2025)
DOI: https://doi.org/10.15575/kjrt.v3i1.1619

Systematic Literature Review: Deep Learning Models in Arabic Script Classification

Muhamad Khansa Khalifaturohman
Robotics Student Club
Bandung, Indonesia
muhammadkhansa067@gmail.com

Abstract**—** Arabic calligraphy is an essential element of Islamic art, and is now widely developed in digital form. With the advancement of artificial intelligence technology, particularly Convolutional Neural Networks (CNNs), several studies have been conducted to classify the styles, characters, and authenticate Arabic calligraphy. This study aims to conduct a systematic literature review on the application of CNNs in the recognition and classification of Arabic calligraphy. The identification process was carried out by searching several scientific databases and screening 152 articles, but only five studies met the criteria for relevance and eligibility. The results of the study indicate that the application of CNNs in this domain is still limited and dominated by a focus on style or letter classification, while topics such as authenticity of original works and AI-generated calligraphy detection are still very rarely researched. The limited number of available studies indicates that this topic is an open area for further exploration in the academic realm and the development of digital Islamic art preservation technology

Keywords- AI-generated art, Convolutional Neural Network, Image Classification, Islamic Calligraphy

I. INTRODUCTION

Arabic script recognition, especially handwritten script, presents unique challenges due to its complexity and variety [1]. The Arabic script consists of 28 letters, each of which can appear in various forms depending on its position in the word, complicating the classification process. Consequently, there has been significant interest in applying deep learning techniques to improve the accuracy and efficiency of Arabic script classification systems.

As the field of Arabic script character recognition continues to grow, Systematic Literature Reviews (SLRs) play a vital role in synthesizing existing research, identifying gaps, and proposing future research directions [2]. This review aims to provide a comprehensive overview of deep learning models for Arabic script classification , highlighting key methodologies, performance metrics, and potential areas for further research.

This Systematic Literature Review (SLR) aims to systematically compare various deep learning architectures, including: Convolutional Neural Network (CNN) [3], [4], Convolutional Recurrent Neural Network (RNN) [5], Bidirectional Long Short-Term Memory (Bi-LSTM) [6], Attention-based CNN [7], and Capsule Network (CapsNet) [8], [9]. Highlighting their strengths and weaknesses in the context of Arabic script classification.

This literature review involves several datasets related to Arabic script, including: AHCD Dataset [10], IFN/ENIT-Database [11], KHATT Database [12], Hijja Dataset [13], and MADBase Dataset [14]. By integrating findings from various studies, this review will not only compare the strengths and weaknesses of these deep learning models but also provide recommendations for future research directions. By identifying gaps in the current literature and proposing potential avenues to explore, this systematic review aims to contribute to the development of an effective Arabic handwriting classification system.

II. RELATED WORKS

Several recent studies have focused on improving the performance of Arabic character recognition through deep learning architectures [15], [16]. The most common approach is a modified CNN, either as a single model or as a feature extraction component in a hybrid framework. Integrating CNNs with sequential mechanisms, such as RNN, BLSTM, and CTC, has been shown to improve temporal context modeling without requiring explicit segmentation [17], [18], [19] Lightweight CNN variants have also been explored for computational efficiency [20], while CapsNet has been proposed as an alternative that explicitly preserves hierarchical spatial relationships between features [21].

Early research based on pure CNN paved the way for architectural improvements [10], [22], [23], [24]. Subsequently, a CNN-RNN combination (CRNN) was applied to Arabic text document classification [25], while an end-to-end, segmentation-free architecture combining CNN, BLSTM, and CTC was equipped with an adaptive decoder to improve the decoding process [26]. A comparative study of several RNN-based models found BLSTM to be the most consistent variant on text classification tasks [18]. The overall literature confirms the trend towards hybrid models and attention mechanisms to balance accuracy, generalization, and computational burden in processing both handwritten and digital Arabic text.

III. RESEARCH METHODS

This study uses a Systematic Literature Review (SLR) approach [2], [27], a method for systematically and transparently identifying, evaluating, and interpreting all research relevant to a specific research question, topic, or phenomenon of interest. This SLR is designed to summarize findings from previous studies on the use of artificial intelligence models in Arabic script classification.

This study was conducted using the Systematic Literature Review (SLR) method, which was adapted to the Preferred Reporting Items for Systematic Reviews and Meta Analyses (PRISMA) guidelines. PRISMA is a useful software tool to support more interactive *literature review reporting* with easy access, which allows researchers to create systematic flowcharts [15].

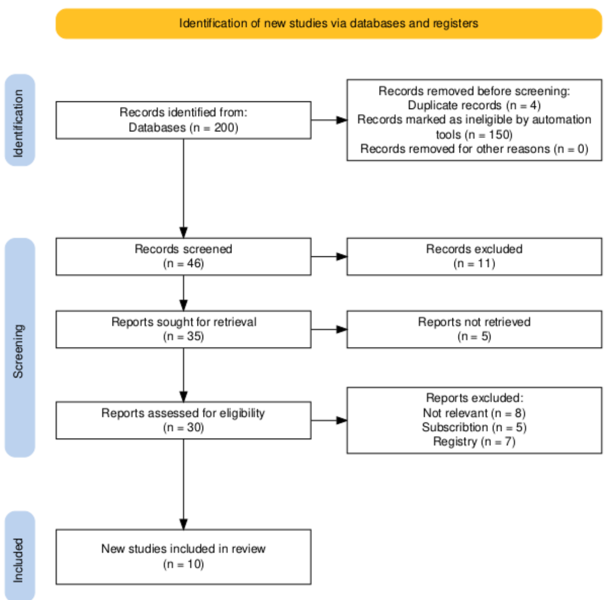

Fig. 1. PRISMA flow diagram

The literature search was conducted in scientific databases such as Mendeley, Google Scholar, IEEE Xplore, Springer, OpenAccess, Arxiv, and other repositories. The keywords used in the literature search were:

- “ Handwritten Arabic classification”
- “Deep learning models for Arabic: CNN, CRNN, BiLSTM , Attention-based CNN, Capsule Network (CapsNet)”

In the identification stage, a literature search was conducted from a database of 200 articles. A total of 154 articles were automatically removed before the screening process because they were deemed not to meet the criteria, resulting in a total of 46 articles. In the screening stage, 46 articles were selected, but 11 of them were excluded because they were not relevant in content or context, resulting in a total of 35 articles. Of the 35 articles that were further searched, 5 articles were not fully accessible (paid or unavailable), resulting in 30 articles. The screening continued, resulting in 8 irrelevant articles, 5 paid articles, and 7 registry articles. After screening, 10 articles were successfully obtained and assessed for eligibility, resulting in a total of 10 studies worthy of synthesis and inclusion in the final review.

A. *Dataset Related to Literature Study*

This literature study involves a number of datasets related to Arabic script, including:

1) Arabic Handwritten Character Dataset (AHCD)
   The dataset consists of 16,800 Arabic characters written by 60 participants. Each participant wrote each character (the letters "alif" to "ya") 10 times in two forms. This is illustrated in the Figure 2.

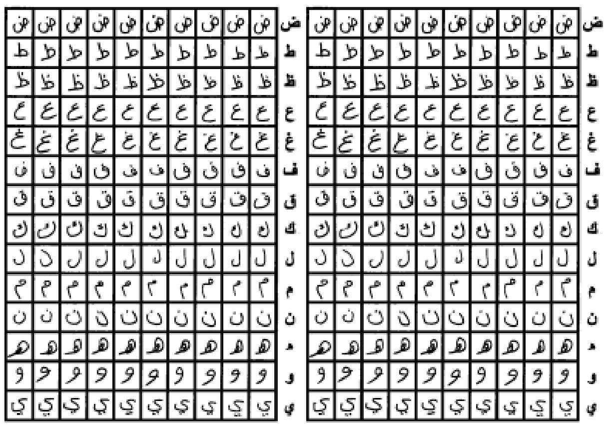
Fig. 2. Samples in the AHCD dataset [10]

2) IFN/ENIT Database
   The dataset consists of 26,400 names containing approximately 210,000 Arabic characters written by 411 participants. A sample of the database is as in Figure 3.

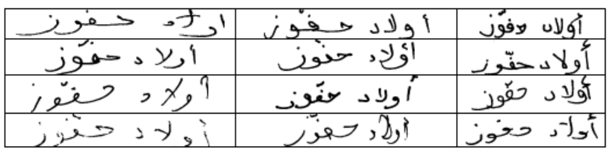
Fig. 3. Sample in the IFN/ENIT database [11]

3) KHATT Dataset
The dataset consists of 1,000 participants' paragraphs scanned at varying resolutions (200, 300, and 600 DPI). Participants come from different countries, genders, age groups, handedness, and education levels. A sample of the dataset is as in Figure 4.

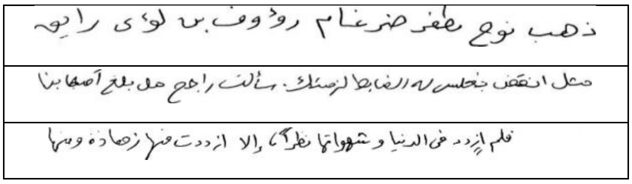

Fig. 4. Sample Arabic script in the KHATT dataset [12]

4) Hijja Dataset
Hijja is a dataset consisting of Arabic handwriting from children aged 7 to 12. The dataset, totaling 47,434 characters, was collected from 591 participants with different Arabic writing styles [13].

5) MADBase Dataset
Separate data between the test data and the training data. The training data contains a total of 60,000 photos, while the test data contains a total of 10,000 photos with different writing styles [14].

IV. RESULT AND DISCUSSION

A. *State of the Art*

A literature review found a total of 10 studies that align with the theme of this study, using similar or identical methods . is table from study previously .

Table 1. State of the Art

Paper	Objective	Method	Results	Limitation
Elagamy , M. et al. (2023), [17]	Developing an Arabic character classification model with modified CNN (HACR-MDL).	CNN with data augmentation on the AHCD dataset.	Accuracy 98.54%.	Difficulty in classifying similar characters (ر & ز), requires a more diverse dataset.
Gader, T. et al. (2022), [18]	Improve the accuracy and efficiency of Arabic text recognition.	CNN-RNN-CTC with attention and BLSTM, IFN-ENIT dataset.	Accuracy 97.1%.	Limited generalization , high complexity for real-time.
Mahdi, M. et al. (2024), [19]	Improving the model with a hybrid CNN and Bi-RNN approach.	CNN-BiLSTM/BiGRU on AHCD and Hijjaa datasets.	Accuracy 97.05% (AHCD), 91.78% (Hijjaa).	Common challenges: overfitting, big data requirements, high computation.
Saber, A., et al. (2024), [28]	Handles the complexities of Arabic calligraphy for better accuracy.	CNN + Bi-LSTM, rigorous preprocessing, KHATT dataset.	Character error 2.96%, accuracy 97.04%.	Need extensive data , risk of overfitting.
Mehdi, D., et al. (2022), [21]	Effectively capture the spatial structure of Arabic handwriting.	Capsule Network without preprocessing, 50,000 character dataset.	TOP-1 accuracy >97%.	Prone to against noise without preprocessing
El-Sawy , et al. (2017), [10]	Improving Arabic character recognition with CNN.	CNN on a dataset of 16,800 characters from 60 authors.	Accuracy 94.9%, average error 5.1%.	Similar character issue, need larger dataset.
Ameur, M. et al. (2020), [25]	Arabic text classification in NLP.	CNN + RNN, OSAC dataset.	Model outperforms method traditional .	Not explicitly stated, but implicit in the general challenges of Arabic NLP.
Kamal, M., et al. (2022), [20]	Handle Arabic handwriting variations with lightweight CNN.	18-layer CNN, AHCD & MadBase datasets .	Accuracy 96.93% (AHCD), 99.35% (MadBase).	Limited generalization , lightweight models are less suitable for complex tasks.
Mouhcine, et al. (2025), [26]	Arabic handwriting recognition without segmentation.	CNN + BLSTM + CTC + WBS Decoder, IFN/ENIT dataset produces better recognition results.	Accuracy 94.58%.	Dependence on dataset, difficulty handling complex text variations.
Al-Qerem, et al. (2024), [29]	Online Arabic text classification with multiple deep learning models.	CRNN, LSTM, CNN-LSTM, Bi-LSTM, news datasets.	The best Bi-LSTM with an accuracy of 94.02%.	Limited generalization , does not address class imbalance.

B. *Comparison of Deep Learning Models Based on Related Studies*

To provide a more comprehensive picture of the performance of various deep learning architectures in the task of Arabic script classification, a comparative analysis was conducted on several models frequently used in previous research.

Table 2. Comparison of deep learning models

Model	Studies	Dataset	Results	Lack
CNN	El- Sawy et al. (2017)	AHCD (16,800 characters)	94.9% accuracy (5.1% error rate)	Difficulty distinguishing similar characters; poor grasp of sequence
	Kamal et al. (2022)	AHCD & MadBase	96.93% (AHCD), 99.35% (MadBase)	

Model	Studies	Dataset	Results	Lack
CRNN	Gader & Echi (2022)	IFN-ENIT	97.1% accuracy	Dependence on specific datasets; high complexity
BiLSTM	Al-Qerem et al. (2024)	Article dataset news	94.02% accuracy	Low performance on noisy data, need lots of data
Attention-based CNN	Gader & Echi (2022)	IFN-ENIT	97.1% accuracy	Computationally intensive, still affected by data distribution
CapsNet	Mehdi & Abdelghani (2022)	50,000 Arabic characters	>97% accuracy (Top-1)	Prone to noise, without preprocessing less robust

C. *Discussion*

Based on a literature review, a significant number of studies have demonstrated the increasing complexity of artificial intelligence (AI) models, particularly deep learning architectures, in Arabic script classification. The dominant approach in recent studies is the application of Convolutional Neural Networks (CNNs), often combined with Recursive Neural Networks (RNNs), Long Short-Term Memory (LSTM), and Connectionist Temporal Classification (CTC) mechanisms to improve recognition accuracy and context awareness of handwritten Arabic text.

For example, Elagamy et al.[17], proposed a modified CNN architecture called HACR-MDL, which achieved a remarkable accuracy of 98.54% on the AHCD dataset by leveraging data augmentation and structural optimization. Their model rivaled conventional models in terms of accuracy and computational complexity. However, they noted that the system still struggled to distinguish visually similar characters such as ر (Rā') and ز (Zāy), suggesting the need for a more diverse dataset.

Similarly, Gader and Echi introduced a CNN-RNN-CTC model with an attention mechanism that captures sequence dependencies using a Bidirectional LSTM (BLSTM) [18]. Their model achieved 97.1% accuracy on the IFN-ENIT database, but highlighted limitations in generalization due to its reliance on a single dataset and concerns regarding the computational demands in real-time applications.

Hybrid models also demonstrated competitive performance. Mahdi et al. [19], developed a hybrid CNN-BiGRU model that combined spatial feature extraction with temporal modeling. Their model achieved the best results with 97.05% accuracy on AHCD and 91.78% on Hijjaa. Saber et al. [13], further supported the effectiveness of the hybrid model by integrating CNN and BiLSTM on the KHATT dataset, resulting in a character error rate of only 2.96% and an accuracy of 97.04%. Both studies highlight the advantages of combining spatial and sequence learning, although they acknowledge the risk of overfitting and high training costs.

Another approach was introduced by Mehdi and Abdelghani [21], who used a CapsNet to model the structural hierarchy of Arabic characters. With a top-1 accuracy exceeding 97%, their method offers a promising alternative to CNNs. However, the lack of preprocessing steps such as noise filtering may limit its real-world applications.

Previous research by El-Sawy et al [10], demonstrated the baseline effectiveness of CNNs on a dataset of 16,800 handwritten characters, achieving 94.9% accuracy and paving the way for further improvements through architectural innovations. Meanwhile, Ameur et al. focused on text classification using a hybrid CNN-RNN (CRNN) on an Arabic document corpus, highlighting the power of deep learning models in capturing local features and semantic sequences [25].

Efficiency-focused architectures were also explored. Kamal et al. [16], designed an 18-layer lightweight CNN, achieving 96.93% and 99.35% accuracy on AHCD and MadBase, respectively. While efficient, the limited dataset coverage of this model raises concerns about its generalizability across diverse handwriting styles.

In the case of segmentation-free recognition, Mouhcine and Amrouche, proposed a CNN-BLSTM-CTC architecture integrated with a WBS decoder, achieving 94.58% accuracy on IFN/ENIT [26]. This approach avoids the complexity of manual segmentation while maintaining high recognition performance, although it is still dataset-specific.

Finally, Al-Qarem et al. [29], compared several RNN-based architectures for Arabic text classification, finding Bi-LSTM to be the most effective with an accuracy of 94.02%. However, they noted that the reliance on a news article dataset may limit its applicability to broader contexts due to the rich morphological diversity of Arabic.

In summary, the literature shows that no single AI model is universally optimal for all Arabic script classification tasks. CNN-based models offer strong performance with relatively low complexity, making them suitable for real-time or resource-constrained environments. Hybrid models that combine CNNs with sequencing layers such as LSTMs or GRUs consistently achieve high accuracy, but often at the expense of higher computational requirements. Furthermore, dataset diversity and preprocessing remain key challenges across studies, impacting model generalization and robustness. Therefore, model selection should be tailored to the constraints of the target application, including accuracy requirements, hardware limitations, and data characteristics.

V. Conclusion

The CNN model developed to detect the authenticity of digital Islamic calligraphy yielded an F1-score of 0.464 with low prediction accuracy for both classes. This performance was affected by the limited data set of only 5 AI calligraphy images and their high visual similarity to authentic calligraphy, leading to the model's tendency to misclassify. These results indicate that the model is not yet able to effectively distinguish between human- and AI-generated calligraphy.

REFERENCES

[1] A. Ahmadian, K. Fouladi, and B. N. Araabi, "Model-based Persian calligraphy synthesis via learning to transfer templates to personal styles," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 23, no. 3, pp. 183–203, Sep. 2020, doi: 10.1007/s10032-020-00353-1.

[2] A. Nightingale, "A guide to systematic literature reviews," *Surgery (Oxford)*, vol. 27, no. 9, pp. 381–384, Sep. 2009, doi: 10.1016/j.mpsur.2009.07.005.

[3] Y. S. Chernyshova, A. V Sheshkus, and V. V Arlazarov, "Two-step CNN framework for text line recognition in camera-captured images," *IEEE Access*, vol. 8, pp. 32587–32600, 2020, doi: 10.1109/ACCESS.2020.297405.

[4] E. P. N. A. Wijaya, "Klasifikasi Akasara Jawa Dengan CNN," *Jurnal Teknika*, vol. 12, no. 2, pp. 61–64, 2020, doi: 10.30736/jt.v13i2.479.

[5] N. K. Manaswi and N. K. Manaswi, "RNN and LSTM," in *Deep Learning with Applications Using Python*, Apress, 2018, pp. 115–126. doi: 10.1007/978-1-4842-3516-4_9.

[6] F. Shahid, A. Zameer, and M. Muneeb, "Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM," *Chaos Solitons Fractals*, vol. 140, p. 110212, 2020.

[7] S. Majid, F. Alenezi, S. Masood, M. Ahmad, E. S. Gündüz, and K. Polat, "Attention based CNN model for fire detection and localization in real-world images," *Expert Syst Appl*, vol. 189, p. 116114, Mar. 2022, doi: 10.1016/j.eswa.2021.116114.

[8] J. Gu, V. Tresp, and H. Hu, "Capsule network is not more robust than convolutional network," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 14309–14317.

[9] S. Pande and M. S. R. Chetty, "Analysis of capsule network (Capsnet) architectures and applications," *J Adv Res Dynam Control Syst*, vol. 10, no. 10, pp. 2765–2771, 2018.

[10] A. El-Sawy, M. Loey, and H. El-Bakry, "Arabic handwritten characters recognition using convolutional neural network," *WSEAS Transactions on Computer Research*, vol. 5, no. 1, pp. 11–19, 2017.

[11] M. Pechwitz, S. S. Maddouri, V. Märgner, N. Ellouze, and H. Amiri, "IFN/ENIT-database of handwritten Arabic words," in *Proc. of CIFED*, Citeseer, 2002, pp. 127–136.

[12] S. A. Mahmoud *et al.*, "KHATT: An open Arabic offline handwritten text database," *Pattern Recognit*, vol. 47, no. 3, pp. 1096–1112, Mar. 2014, doi: 10.1016/j.patcog.2013.08.009.

[13] Mohamed Gresha Mahdi, "Hijja Dataset," Kaggle. Accessed: Jun. 24, 2025. [Online]. Available: https://www.kaggle.com/datasets/mohamedgreshamahdi/hijja-dataset?select=Hijja2.txt

[14] E. A. El-Sherif and S. Abdelazeem, "A Two-Stage System for Arabic Handwritten Digit Recognition Tested on a New Large Database.," in *Artificial intelligence and pattern recognition*, 2007, pp. 237–242.

[15] N. R. Haddaway, M. J. Page, C. C. Pritchard, and L. A. McGuinness, "PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis," *Campbell Systematic Reviews*, vol. 18, no. 2, Jun. 2022, doi: 10.1002/cl2.1230.

[16] M. N. AlJarrah, M. M. Zyout, and R. Duwairi, "Arabic Handwritten Characters Recognition Using Convolutional Neural Network," in *2021 12th International Conference on Information and Communication Systems (ICICS)*, IEEE, May 2021, pp. 182–188. doi: 10.1109/ICICS52457.2021.9464596.

[17] M. N. Elagamy, M. M. Khalil, and E. Ismail, "HACR-MDL: Handwritten Arabic Character Recognition Model using Deep Learning," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. X-1/W1-2023, pp. 123–128, Dec. 2023, doi: 10.5194/isprs-annals-X-1-W1-2023-123-2023.

[18] T. Ben Aïcha Gader and A. Kacem Echi, "Attention-based deep learning model for Arabic handwritten text recognition," *Machine Graphics and Vision*, vol. 31, no. 1/4, pp. 49–73, Dec. 2022, doi: 10.22630/MGV.2022.31.1.3.

[19] M. G. Mahdi, A. Sleem, I. M. Elhenawy, and S. Safwat, "Enhancing the Recognition of Handwritten Arabic Characters through Hybrid Convolutional and Bidirectional Recurrent Neural Network Models," *Sustainable Machine Intelligence Journal*, vol. 9, pp. 34–56, Oct. 2024, doi: 10.61356/SMIJ.2024.9382.

[20] M. Kamal, F. Shaiara, C. M. Abdullah, S. Ahmed, T. Ahmed, and Md. H. Kabir, "Huruf: An Application for Arabic Handwritten Character Recognition Using Deep Learning," in *2022 25th International Conference on Computer and Information Technology (ICCIT)*, IEEE, Dec. 2022, pp. 1131–1136. doi: 10.1109/ICCIT57492.2022.10054769.

[21] D. Mehdi and S. Abdelghani, "Handwritten Arabic characters recognition using Capsule Networks," in *2022 9th International Conference on Wireless Networks and Mobile Communications (WINCOM)*, IEEE, Oct. 2022, pp. 1–6. doi: 10.1109/WINCOM55661.2022.9966452.

[22] Prof. N. V. Gawali, "Image Classification Using Convolutional Neural Networks," *Int J Res Appl Sci Eng Technol*, vol. 11, no. 5, pp. 5176–5179, May 2023, doi: 10.22214/ijraset.2023.52435.

[23] A. Bychkov, T. Kiseleva, and E. Maslova, "Usage of Convolutional Neural Networks for Image Classification," *Bulletin of the Siberian State Industrial University*, vol. 1, no. 1, pp. 39–49, Mar. 2023, doi: 10.57070/2304-4497-2023-1(43)-39-49.

[24] X. Lin, "Research of Convolutional Neural Network on Image Classification," *Highlights in Science, Engineering and Technology*, vol. 39, pp. 855–862, Apr. 2023, doi: 10.54097/hset.v39i.6656.

[25] M. S. H. Ameur, R. Belkebir, and A. Guessoum, "Robust Arabic Text Categorization by Combining Convolutional and Recurrent Neural Networks," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 19, no. 5, pp. 1–16, Sep. 2020, doi: 10.1145/3390092.

[26] M. Rabi, "Convolutional Arabic handwriting recognition system based BLSTM-CTC using WBS decoder," *International Journal of Advanced Science and Computer Applications*, vol. 4, no. 1, Jan. 2024, doi: 10.47679/ijasca.v3i2.52.

[27] K. Richardson Dr. *et al.*, "A systematic review," 2013. doi: 10.7326/0003-4819-159-10-201311190-00007.

[28] A. Saber, A. Taha, and K. Abd El Salam, "A Comprehensive Approach to Arabic Handwriting Recognition: Deep Convolutional Networks and Bidirectional Recurrent Models for Arabic Scripts," *International Journal of Telecommunications*, vol. 04, no. 02, pp. 1–11, Jul. 2024, doi: 10.21608/ijt.2024.291347.1052.

[29] A. Al-Qerem, M. Raja, S. Taqatqa, and M. R. A. Sara, "Utilizing Deep Learning Models (RNN, LSTM, CNN-LSTM, and Bi-LSTM) for Arabic Text Classification," 2024, pp. 287–301. doi: 10.1007/978-3-031-43490-7_22.