International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Convolutional Arabic Handwriting Recognition System Based BLSTM-CTC Using WBS Decoder

Mouhcine Rabi*[1], Mustapha Amrouche[2]

Submitted: 08/12/2023 **Revised**: 15/01/2024 **Accepted**: 29/01/2024

Abstract: Arabic handwriting recognition (AHR) poses major challenges for pattern recognition due to the cursive script and visual similarity of Arabic characters. While deep learning demonstrates promise, architectural enhancements may further improve performance. This study presents an offline AHR approach using a convolutional neural network (CNN) with bidirectional long short-term memory (BLSTM) and connectionist temporal classification (CTC). By enhancing temporal modelling and context representations without segmentation requirements, this BLSTM-CTC-CNN framework with an integrated Word Beam Search (WBS) decoder achieved 94.58% accuracy on the IFN/ENIT database. Results highlight improved efficiency over prior works. This demonstrates continued advancement in sophisticated deep learning techniques for accurate AHR through specialized modelling of Arabic script cursive properties and decoding constraints. This research represents an advancement in the continuous development of progressively intricate and precise systems for handwriting recognition.

Keywords: *Arabic handwriting recognition (AHR), Bi-Dimensional Long Short-Term Memory (BLSTM), Convolutional Neural Networks (CNN), Connectionist Temporal Classification (CTC), Word Beam Search (WBS).*

1. Introduction

Handwriting recognition, a distinct subset of Optical Character Recognition (OCR), holds significant importance in various contexts, such as industry, education, government, and healthcare [8, 9, 10, 11] and many other fields [1, 14], identifying Arabic handwriting in sequential images presents a promising research opportunity. Despite its complexity, real-time application of OCR enhances its practicality, making it more applicable in real-world scenarios [98]. The significance of handwriting recognition systems is underscored by their capacity for efficient data storage, expedited information retrieval, enhanced accessibility, and superior customer service [6, 7, 12, 13].

Arabic, spoken by over 375 million individuals globally, is not the only language that uses Arabic letters; they are also used in several other languages including Persian, Urdu, and Jawi [15]. Despite this extensive usage, the development of a robust offline automated handwriting recognition system for Arabic scripts remains a big challenge. This is primarily attributed to the cursive nature of the Arabic script, the substantial variation in individual handwriting styles, and the positional-dependent shape alterations of Arabic letters within a word [16], the Fig 1 provides an overview of some characteristic traits of Arabic handwriting. Despite these challenges, recent years have witnessed significant strides towards enhancing the accuracy of Arabic handwriting recognition systems.

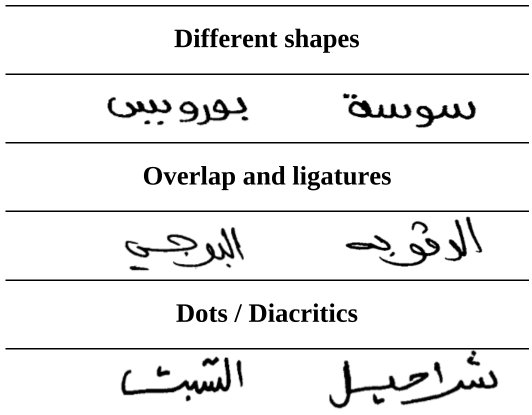

Fig 1. Some examples of Arabic handwriting characteristics

Handwritten word recognition typically employs preprocessing techniques to mitigate the extensive variation in word arrangement. Standard preprocessing methods rely on the normalization of character size and the correction of letter slopes in the given image [17,18, 19]. Traditional offline handwriting recognition systems often involve a segmentation step prior to feature extraction. However, manual feature extraction in Arabic handwriting recognition presents several challenges, primarily due to distortions and pattern variability.

Hidden Markov Models (HMMs) are employed in sequence learning tasks to model outputs as character sequences. Despite their acceptable effectiveness, HMMs are constrained by the Markovian assumption, which restricts

[1] *IMIS Laboratory, Faculty of Applied Sciences, Ait Melloul, Ibn Zohr University, Morocco*
ORCID ID 0009-0009-6864-0825
[2] *IRF-SIC Laboratory, Faculty of Sciences, Agadir, Ibn Zohr University Morocco*
** Corresponding Author Email: m.rabi@uiz.ac.ma*

contextual learning to the present state, as indicated in references. This limitation can impact the depth of text analysis that HMMs can achieve. [20, 21].

Deep learning techniques have proven their efficacy across a multitude of real-world applications, including autonomous driving, natural language processing, finance, and healthcare [22, 23]. In pattern recognition, hierarchical deep neural networks enable end-to-end systems, eliminating the need for segmentation and manual feature engineering.

Studies indicate that Recurrent Neural Networks (RNNs) outperform Hidden Markov Models in tasks related to sequence labeling. This superior performance is particularly noticeable in areas like speech and handwriting recognition [24, 25]. Unlike HMMs, RNNs consider the interdependencies between neighboring states. While RNNs are trained using discriminative methods, HMMs uses generative methods. It has been noted that discriminative methods frequently yield superior outcomes in pattern recognition tasks when a substantial dataset is available.

In the realm of Optical Character Recognition (OCR) for handwriting, the implementation of Recurrent Neural Network (RNN) architectures, which utilize raw pixel data as input, has brought about a significant paradigm shift. This innovative methodology has revolutionized the processing and interpretation of handwritten data, providing a more comprehensive and efficient solution. Initially, RNNs were notorious for their training difficulties, primarily due to complications such as the vanishing gradient problem during backpropagation training. However, the advent of the Long Short-Term Memory (LSTM) architecture has breathed new life into the application of RNNs. The LSTM, a highly non-linear recurrent network equipped with multiplicative gates and additive feedback, empowers the network to leverage the contextual information embedded in the data. [26].

While the application of RNNs to offline handwriting presents its own set of challenges due to the multi-dimensional nature of the input, a viable solution has been found in presenting the images to the network one vertical line at a time, effectively transforming them into 1D sequences. A bidirectional variant of the LSTM architecture has been proposed to access context in both forward and backward directions, which are then connected to a single output layer. Building on this foundation, recent advancements in offline handwriting recognition have seen the successful integration of Convolutional Neural Networks with RNNs [27, 28, 29, 30, 31, 32]. CNNs have proven to be highly effective in discerning the structure of handwritten characters or words, thereby facilitating the automatic extraction of distinct features. To address the challenge of segmented training data, the Connectionist Temporal Classification (CTC) has been employed to align transcripts with the output of the neural network, further enhancing the efficacy of this combined approach.

In this study, we develop an advanced system for offline Arabic handwriting recognition, exploiting the capabilities of CNNs, BLSTM and CTC. The proposed system is specifically engineered to tackle the distinct challenges associated with Arabic handwriting recognition. The BLSTM, a particular instance of directed acyclic graph networks, extends the functionality of standard RNNs by incorporating recurrent connections across all spatio-temporal dimensions inherent in the data. These connections equip the BLSTM with robustness against local distortions across any combination of input dimensions, thereby enabling the flexible modeling of two-dimensional context.

The organization of this paper is as follows: Section 2 provides an exhaustive survey of the current state of the art in the field of Arabic handwriting recognition. Our proposed system, which integrates Convolutional Neural Networks, Bidimensional Long Short-Term Memory, and Connectionist Temporal Classification with a Word Beam Search decoder, is detailed in Section 3. Section 4 is dedicated to the presentation of our experimental results and a comprehensive discussion, offering a systematic evaluation of the system’s performance. The paper culminates in Section 5, where we summarize our key findings and suggest potential directions for future exploration in this research area.

2. Related Work

Offline handwritten recognition has been extensively explored, with various technological solutions being developed over time. Before the advent of deep learning, Hidden Markov Models (HMMs) were the primary method used for handwriting recognition [33, 34, 35, 36]. HMMs are statistical models that manage state transitions and observations. Their goal is to identify the most likely state sequence that maximizes the posterior probability given a sequence of observations. In more formal terms, the aim is to optimize the posterior probability $P(S|X)$, considering these probabilistic models of state transitions and output generation [37, 38, 39, 40].

Following the initial research in offline handwriting recognition, which primarily utilized a sliding window technique to extract hand-crafted features, and Hidden Markov Models for generating character sequences [44], the field began to explore neural network-based approaches. These methods combined Hybrid HMMs with Artificial Neural Network (ANN) models to enhance performance. [41, 42, 43]. The process involves extracting features based on a sliding window technique, subsequently transforming a line of text into a series of temporal features through the application of Hidden Markov Models [45].

Handwriting text recognition models based HMMs display a significant constraint in their use of contextual information. According to the Markovian assumption, these models concentrate solely on the current time step, which can limit their effectiveness. In contrast, techniques based on statistical language models are employed to recognize entire lines of text [46]. These models are often supplemented with external language models to enhance their performance. Additionally, research has been conducted to examine the impact of lexicon during the decoding process [47, 48].

Deep learning methods have proven their efficacy, overcoming the constraints of Hidden Markov Models (HMMs). With the recent advancements in computational power, models based on deep learning, specifically Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and a combination of the two have achieved unprecedented accuracy levels in the field of handwritten text recognition. Rather than depending on manually engineered image features, these models incorporate a Convolutional Neural Network to produce more resilient and effective image features [49, 51, 53, 54, 56, 57, 58]. The process begins with the CNN processing the input handwriting to generate features sequence, which are then passed through an RNN network to obtain the final transcription [27, 50, 52, 55].

An advanced methodology involves the use of a Convolutional Recurrent Neural Network architecture in conjunction with Connectionist Temporal Classification (CTC). This combined approach has been applied to both online [59, 60, 61, 62] and offline handwriting recognition [63, 64, 65, 66]. The Connectionist Temporal Classification (CTC) approach, which originated in the field of speech recognition over the past decade, is a key component of this methodology [71, 72, 73]. The CTC algorithm processes a sequence of probability distributions and outputs a sequence of recognizable characters. More particularly [28, 31, 67, 68, 69, 70] proposed a models-based CNN-LSTM-CTC in a unified framework for Arabic script recognition and achieved a better recognition performance over approaches based on HMMs or hybrid HMM-RNN [74, 75, 76].

Recently, other works employed Transformer architectures, a solution to the gradient explosion/vanishing problem in Recurrent Neural Networks, employs an encoder-decoder structure with an attention mechanism. Transformers have been introduced into the field of text recognition, offering a novel approach to this complex task [78, 79].

The Transformer architecture, known for its capacity to model complex language rules and enable enhanced parallelization, has gained extensive acceptance in the field of text recognition [80, 81, 82]. In the context of historical document transcription, fine-tuning pre-trained transformer models has shown promising results and outperform existing state-of-the-art methods on several datasets [77]. The use of transformers in handwriting recognition presents a significant advancement, particularly due to their ability to handle style-content entanglement and capture both global and local writing style patterns [83, 84, 85].

Generative Adversarial Networks (GANs) can be an adopted approach in pattern recognition. (GANs) comprise two components, a generator to create examples that closely resemble real data, and the discriminator to differentiate between real and generated images. The generator uses a random input vector to create examples, and the discriminator evaluates both real and generated examples to compute the loss, leading to model updates. GANs are commonly used to augment data and generate samples that mirror the real dataset. This helps to prevent overfitting and reduce generalization errors in deep learning models. Recent research papers have explored this potential further [86, 87, 88, 89, 93]. Given the current research gap in the field of pattern recognition, particularly in the context of handwriting recognition, a CDCGAN (Conditional Deep Convolutional Generative Adversarial Networks) model is designed specifically for the guided generation of isolated handwritten Arabic characters. The empirical findings indicate that the CDCGAN is capable of synthesizing handwritten Arabic characters that bear a striking resemblance to the original samples [92].

Table 1. AHR systems: Comparative results.

Work-Year	Architecture	Dataset	RR* (%)
[67] - 2022	CNN-BLSTM	IFN/ENIT	91.79
[28] - 2023	ConvLSTM- CTC	IFN/ENIT	95.05
[31] - 2023	DE-GAN with CNN-LSTM	AcTiV-R	97.95
[70] - 2023	CNN- RNN	IFN/ENIT	87.4
[54] - 2023	CNN-SVM	HACDB	89.7
	CNN-SVM	HIJJA	88.8
[79] - 2021	CNN-Transformer	KHATT	91.71
[80] - 2022	Transformer with cross attention	KHATT	CER*: 18.45
[83] - 2023	Transformer Encoder and Transformer Decoder	KHATT	CER: 13.48
[90] - 2022	GANs- CNN	AHCD	99.78

[91] - 2022	GANs- BiLSTM	IFN/ENIT	95.87
[92] - 2022	CDCGAN	AHCD	95.08

RR: Recognition Rate, CER: Character Error Rate

A review of the literature *(Table1)* reveals that various models such as CNN-RNN, Transformers and GANs have been employed in the field of handwriting recognition. Each with their own strengths and challenges. The combination of Convolutional Neural Networks and Long Short-Term Memory in its variants across directions has shown proficiency in handling spatial relations and textual dependencies. However, these models may overfit on non-challenging datasets. Transformer-based models, on the other hand, offer robust results but require more resources and have latency issues. GANs are useful for data augmentation but their performance is notably sensitive to the balance of datasets. Employing GANs serves as a strategy to counteract the constraints of Arabic datasets. However, this introduces a dependency where the overall performance is contingent upon the specific model applied for classification.

3. Proposed Approach

The following section outlines the development of an advanced Convolutional Recurrent Neural Network model for Arabic handwritten script recognition. The approach is a composite of three key components (Fig 2.), a Convolutional Neural Network based feature extractor, which extracts pertinent features from a text image, recurring layers that utilize Bidimensional Long Short-Term Memory to predict a pre-frame from an input sequence, and a transcription layer that incorporates a Connectionist Temporal Classification. The CTC is particularly effective in calculating the loss function, enabling the model to learn more effectively.

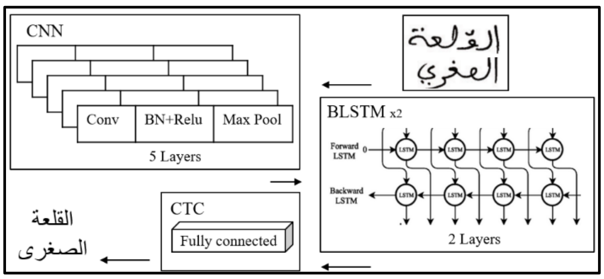

Fig 2. Model architecture

3.1. CNN: Feature extraction

Convolutional Neural Networks introduced by LeCun and Bengio in 1995, have emerged as a powerful tool for feature extraction in the field of Arabic handwriting recognition. CNNs, with their architecture and ability to learn hierarchical representations, can automatically and adaptively learn spatial hierarchies of features. This makes them particularly suited for the task of recognizing complex patterns in Arabic handwriting. By leveraging the multilayered structure of CNNs, low-level features in the initial layers can be transformed into high-level features in deeper layers. These high-level features provide a robust representation of the input data, capturing intricate details that are crucial for accurate recognition. Thus, the use of CNNs as feature extractors offers a promising approach to enhance the performance of Arabic handwriting recognition systems.

A CNN architecture comprises convolution layers with learnable filters that generate feature maps. Pooling layers subsequently reduce the spatial size of these maps, aiming to decrease parameters, computation time, and exercise control over potential overfitting. The adopted CNN architecture is presented and detailed in our previous work [49], with some hyperparameter modifications aimed at improving the feature extraction process. The Fig 3. illustrates the implemented CNN architecture.

The convolution operation identifies pertinent features from the input image using filters, which are then subjected to a batch normalization process. A Rectified Linear Unit (ReLU) activation function is employed to discard negative values and mitigate the impact of vanishing gradients. The outputs from the activation function are forwarded to a max-pooling layer, which conducts sub-sampling by choosing the most significant features.

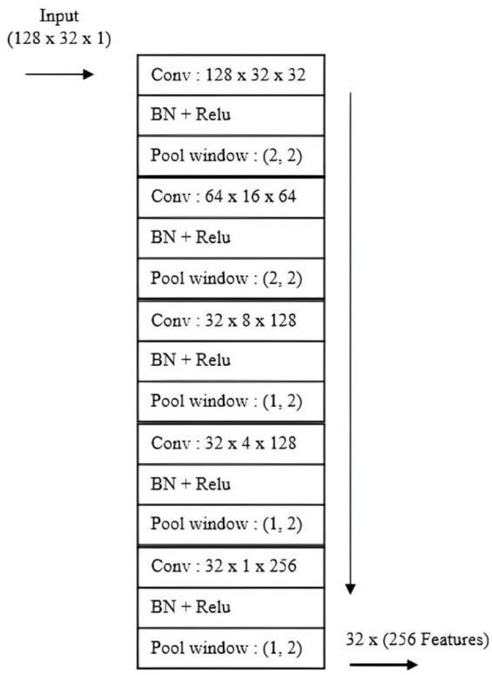

Fig 3. CNN architecture

Five convolutional layers are used, the first two layers are designed with filters of dimension (5, 5) units, a strategic choice to expand the receptive field of the network's early stages. The subsequent layers employ smaller filters of dimension (3, 3) units. The pooling Layers has been incorporated, the initial two layers use a sliding window of size (2x2), while the remaining layers utilize a window of size (1, 2). This variation in window sizes is a deliberate design decision within the network's architecture. Following the automatic feature extraction process ensured by CNN, the output of the final pooling layer is an input of the BLSTM network, which is the second component of our architecture model.

3.2. LSTM vs BLSTM

Recurrent Neural Networks have gained prominence as a powerful model for processing sequential data due to their ability to capture dynamic temporal behavior. A notable evolution within the domain of RNNs is the advent of Long Short-Term Memory (LSTM) networks. These networks possess the ability to learn dependencies over extended periods. LSTM networks incorporate a memory cell that allows them to maintain information in memory for extended periods.

The block diagram (Fig 4.) illustrates the internal structure of an LSTM cell, which consists of three main gates: the input gate, the forget gate, and the output gate. These gates regulate the flow of information through the cell, controlling how much of the previous state (Ct-1) is forgotten, how much of the new input (Xt) is added, and how much of the updated state (Ct) is outputted.

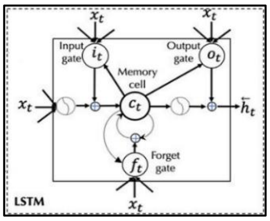

Fig 4. LSTM architecture

- Input gate (i_t): determines how much of the new input (X_t) should be incorporated into the cell state. It takes the current input (X_t) and the previous hidden state (h_{t-1}) as inputs and applies a sigmoid function to produce a value between 0 and 1. A value closer to 1 indicates that more of the new input should be added, while a value closer to 0 indicates that less of the new input should be added.
- Forget gate (f_t): decides how much of the previous cell state (C_{t-1}) should be retained or forgotten. It takes the current input (X_t) and the previous hidden state (h_{t-1}) as inputs and applies a sigmoid function to produce a value between 0 and 1. A value closer to 1 indicates that more of the previous state should be kept, while a value closer to 0 indicates that more of the previous state should be forgotten.
- Output gate (O_t): controls how much of the updated cell state (C_t) should be outputted as the hidden state (h_t). It takes the current input (X_t) and the previous hidden state (h_{t-1}) as inputs and applies a sigmoid function to produce a value between 0 and 1. A value closer to 1 indicates that more of the cell state should be outputted, while a value closer to 0 indicates that less of the cell state should be outputted.

The LSTM cell also includes a tanh function that scales the cell state (Ct) to a value between -1 and 1. This scaled cell state is then multiplied by the output gate (O_t) to determine the final hidden state (h_t).

Bidirectional LSTM, an extensive network has been developed to enhance performance in tasks requiring contextual information from both preceding and succeeding time steps. This improvement is achieved by training two distinct LSTMs, forward and backward states (Fig 5).

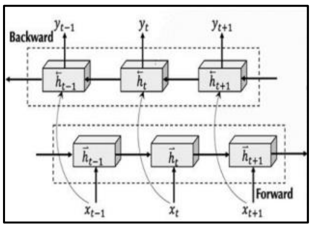

Fig 5. BLSTM architecture

The combined outputs provide a more comprehensive representation, capturing a broader spectrum of temporal dependencies. Two BLSTM layers are used containing two LSTM cells for executing the forward and backward passes of inputs within the network. Each LSTM cell was equipped with 256 hidden units.

3.3. The Connectionist Temporal Classification (CTC)

Upon completion of the feature extraction sequence by the CNN, the BLSTM will propagate the output as a matrix to the CTC layer. The CTC is a type of output for a neural network that typically follows the recurrent neural network layer, which is usually tasked with scoring functions. It presents an exceptional solution for handwritten image recognition as it eliminates the need for pre-segmentation to transform the RNN outputs into label sequences. The CTC performs two tasks: Firstly, it calculates the loss function to train the model, secondly, it decodes the data to obtain the recognized text included in the input image. While CTC is effective in calculating the loss function, other methods such as, best path decoder, Beam Search, or Word Beam Search [99] offer improved decoding performance.

- The Best Path decoder selects the most probable label sequence by considering the entire output sequence's probability. It takes into account the dependencies between consecutive labels, providing a more accurate prediction but at a higher computational cost.
- Beam Search, maintains a set number of alternative sequences (the 'beam width') at each time step. This method balances between computational efficiency and prediction accuracy.

- Word Beam Search is a specialized version of Beam Search that incorporates language model information. It can recognize whole words at a time, making it particularly effective for languages with clear word boundaries like Arabic.

In terms of effectiveness, while Best Path and Beam Search provide good results, Word Beam Search often outperforms them in AHR systems. This is due to its ability to leverage language model information and recognize whole words, which aligns well with the characteristics of the Arabic language.

4. Experiments Results

This section presents a comprehensive analysis of the results derived from our proposed model. It includes an exploration of the IFN-ENIT dataset, data augmentation techniques, and discusses the performance of various model variants under diverse parameters and methods. Each of these elements plays a pivotal role in enhancing the overall efficacy of our system.

4.1. IFN-ENIT Database

The IFN/ENIT dataset, which originates from the Institut fuer Nachrichtentechnik Germany /Ecole Nationale d'Ingénieur de Tunis and was published by Pechwitz et al. [94], consists of 32,492 images of Arabic words written by over 1000 writers. These words represent the names of 937 towns and villages in Tunisia. The dataset is divided into five distinct sets, as detailed in Table 2. For the purpose of training our model, we use sets 'a', 'b', and 'c', while set 'd' for validation and set 'e' for testing.

Table 2. The IFN-ENIT database

Sets	Words	Characters
a	6537	51984
b	6710	53862
c	6477	52155
d	6735	54166
e	6033	45169
Total	**32492**	**257336**

The scarcity of data can pose a significant impediment to the performance of deep neural networks employed in each field research, including our model. To mitigate this issue, we have incorporated the strategy of Data Augmentation. This approach amplifies the size and diversity of our dataset, thereby bolstering the performance and robustness of the system.

4.2. Data augmentation

The application of data augmentation methods is pivotal in boosting the efficacy of handwriting recognition systems, especially when it comes to intricate scripts such as Arabic. For the IFN-ENIT database, which is a widely used resource for Arabic handwriting recognition, data augmentation can involve various transformations such as rotations, translations, scaling, and shearing [95, 96]. These techniques amplify the size and infuse a variety of writing styles and character orientations that the model may encounter, thereby improving its ability to generalize and perform robustly on unseen data. Moreover, data augmentation helps in mitigating overfitting by providing a broader spectrum of data for the model to learn from. This is especially crucial for our system as it is designed to recognize a wide array of Arabic handwriting styles.

In our work, we employ a variety of data augmentation techniques including geometric transformations, morphological operations, and cropping. By introducing variations such as rotation, scaling and alterations in writing thickness, we significantly increase the diversity and size of our training data. The application of these techniques has led to a significant increase in the robustness and accuracy of our models, demonstrating their efficacy in improving model performance. It is important to mention that the choice of data augmentation techniques should be carefully made considering the attributes of the Arabic script.

4.3. Results and discussion

To evaluate the performance of our proposed Arabic handwriting recognition system, we have carried out an exhaustive examination encompassing three separate facets. This evaluation not only takes into account the performance of the system but also considers the results of our previous works for comparative analysis. Furthermore, the elaboration of different systems using various techniques has been incorporated into our study, thereby providing a holistic view of the system's performance. This rigorous approach ensures that our evaluation is both thorough and robust, providing valuable insights into the system's strengths and potential areas for improvement. Table 3 encapsulates the comparative analysis of our proposed system with our preceding models. These include the Arabic handwriting recognition systems based on HMM, HMM-MLP, and CNN-HMM.

Table 3. Comparative results of our previous and present work

System	Approach	RR (%)
RABI [38]	HMM	87.93
RABI [43]	HMM-MLP	88.74
AMROUCHE [49]	CNN-HMM	88.95

Present work (without data augmentation)	**CNN-BLSTM-CTC/WBS**	**91.69**

The comparative analysis of our various Arabic handwriting recognition systems, as presented in Table 3, reveals in one hand the progression from HMM to neural networks architecture. Each subsequent method demonstrates an improvement in the recognition rate, indicating that the incorporation of advanced techniques and architectures enhances performance. In other hand, the CNN-BLSTM-CTC/WBS model that outperforms all other models by a significant margin, achieving a recognition rate of 91.69% without data augmentation step. This highlights the effectiveness of this approach for Arabic handwriting recognition.

To explore the effectiveness of each method, we conducted a series of experiments. Table 4 provides a more detailed evaluation of our model under various scenarios and techniques. The use of data augmentation has shown to enhance the model's accuracy by 2.89 %. This indicates that the model has effectively learned more generalized features, leading to improved performance on unseen data.

In terms of decoding strategies, the Word Beam Search (WBS) decoder yielded the best results among all decoders. This highlights the significance of incorporating linguistic knowledge into the decoding process, which can substantially enhance the system's performance. The distinct characteristics of each decoder influence their effectiveness. The Best Path decoder, while simple, selects the most probable output at each time step without considering dependencies between consecutive characters. Beam Search improves upon this by considering a 'beam' of the top 'k' probable sequences at each time step, allowing for consideration of more potential sequences. However, it is the Word Beam Search that proves to be the most effective. By incorporating language models to favor sequences that form valid words, WBS significantly enhances the decoding process and leads to superior results.

Table 4. Comparative results under various techniques

RR (%)	Best Path	Beam Search	Word Beam Search
Without data augmentation	89.02	90.22	91.69
With data augmentation	89.72	92.17	94.58

These results shows that the WBS technique not only performs better overall, but also benefits more from data augmentation compared to the other techniques. This could be due to the fact that WBS is a more flexible decoding technique, which can better leverage the additional diversity provided by data augmentation. This highlights the significance of incorporating linguistic knowledge into the decoding process, which can substantially enhance the system's performance. The distinct characteristics of each decoder influence their effectiveness.

Table 5. Comparative state of the art recognition systems on IFN/ENIT dataset.

Work	Approach	RR (%)
[67]	CNN-BLSTM	91.79
[28]	ConvLSTM- CTC	95.05
[70]	CNN- RNN	87.4
[91]	GANs- BiLSTM	95.87
[97]	CNN-SVM	92.95
Our work	**CNN-BLSTM-CTC/WBS**	**94.58**

In the last phase of our evaluation, we juxtapose our system with existing methodologies. Our model sets itself apart through its effectiveness, which is a result of the incorporation of various combined techniques. As depicted in Table 5, our approach achieves a recognition rate (RR) of 94.58%. This performance is competitive when compared to other approaches such as the ConvLSTM-CTC and GANs-BLSTM methods. However, our model significantly surpasses the performance of standard CNN-RNN and CNN-SVM methods.

The findings from our study show the effectiveness of our proposed model for Arabic handwriting recognition. The robust performance of our model, with hybrid deep learning approach using CNN-BLSTM with a connectionist temporal classification loss function, using WBS decoder, presents a promising strategy for this task. Future work could explore ways to further optimize our model to close the gap with the highest performing methods.

5. Conclusion & Perspectives

This research has presented an Arabic handwriting recognition method using a Convolutional Neural Network for feature selection, a Bidirectional LSTM for classification, and a CTC layer for label alignment, decoded with a Word Beam Search decoder. The model's effectiveness is demonstrated through its performance on the IFN/ENIT database.

The use of data augmentation techniques has further enhanced the model's performance by enriching the diversity of the training data, it has led to a more resilient classifier capable of accommodating a broad spectrum of handwriting styles and variations. The comparative analysis with different techniques has provided valuable insights into the effectiveness of each method used, further validating the proposed model's robustness and accuracy. This research contributes to the ongoing efforts in handwriting recognition for complexed cursive script, and the findings suggest promising directions for future work. The model's architecture and the strategies employed open up possibilities for further optimizations and improvements, potentially leading to even more accurate and efficient Arabic handwriting recognition systems

Looking ahead, the integration of Generative Adversarial Networks could be a promising avenue for further enhancing the performance of our model. GANs have shown remarkable success in generating new data that is similar to the input data. In the context of our work, GANs could be used to generate additional synthetic training samples, thereby increasing the diversity and volume of our training data. This could potentially lead to improved model robustness and generalization capabilities. Future research could explore the integration of GANs into our existing framework and investigate the impact on Arabic handwriting recognition performance. In addition to the aforementioned strategies, another promising direction for future work is the exploration of Transformer models. Transformers, which have been highly successful in various natural language processing tasks, could potentially be adapted for the task of Arabic handwriting recognition. The self-attention mechanism in Transformers allows them to capture dependencies in the data regardless of their distance, which could be particularly useful for recognizing handwriting where the context plays a crucial role.

To conclude, the field of Arabic handwriting recognition continues to present numerous opportunities for exploration and innovation. The potential to push the boundaries of what is currently achievable is an exciting direction for future work.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Porwal, U., Fornés, A. & Shafait, F. Advances in handwriting recognition. IJDAR 25, 241–243 (2022). https://doi.org/10.1007/s10032-022-00421-8

[2] M. M. Al-Taee, S. B. H. Neji and M. Frikha, "Handwritten Recognition: A survey," *2020 IEEE 4th International Conference on Image Processing, Applications and Systems (IPAS)*, Genova, Italy, 2020, pp. 199-205, doi: 10.1109/IPAS50080.2020.9334936.

[3] Alrobah, N., Albahli, S. Arabic Handwritten Recognition Using Deep Learning: A Survey. Arab J Sci Eng 47, 9943–9963 (2022). https://doi.org/10.1007/s13369-021-06363-3.

[4] Altwaijry, N., Al-Turaiki, I. Arabic handwriting recognition system using convolutional neural network. Neural Comput & Applic 33, 2249–2261 (2021). https://doi.org/10.1007/s00521-020-05070-8.

[5] R. Raj and A. Kos, "A Comprehensive Study of Optical Character Recognition," *2022 29th International Conference on Mixed Design of Integrated Circuits and System (MIXDES)*, Wrocław, Poland, 2022, pp. 151-154, doi: 10.23919/MIXDES55591.2022.9837974.

[6] De Oliveira, L.L., Vargas, D.S., Alexandre, A.M.A. et al. Evaluating and mitigating the impact of OCR errors on information retrieval. Int J Digit Libr 24, 45–62 (2023). https://doi.org/10.1007/s00799-023-00345-6.

[7] Ravi, S., Chauhan, S., Yadlapallii, S.H., Jagruth, K., Manikandan, V.M. (2022). A Novel Educational Video Retrieval System Based on the Textual Information. In: Abraham, A., et al. Proceedings of the 13th International Conference on Soft Computing and Pattern Recognition. Lecture Notes in Networks and Systems, vol 417. Springer, Cham. https://doi.org/10.1007/978-3-030-96302-6_47.

[8] Monteiro G, Camelo L, Aquino G, Fernandes RdA, Gomes R, Printes A, Torné I, Silva H, Oliveira J, Figueiredo C. A Comprehensive Framework for Industrial Sticker Information Recognition Using Advanced OCR and Object Detection Techniques. *Applied Sciences*. 2023; 13(12):7320. https://doi.org/10.3390/app13127320.

[9] Akhtar, Zainab; Lee, Jong Weon;Attique Khan, Muhammad;Sharif, Muhammad;Ali Khan, Sajid;Riaz, Naveed Less. « Optical character recognition (OCR) using partial least square (PLS) based feature reduction: an application to artificial intelligence for biometric identification » *Journal of enterprise information management*, 24 Apr 2023, Vol. ahead-of-print, Issue ahead-of-print, pages 767 – 789. ISSN**:** 17410398. DOI: 10.1108/JEIM-02-2020-0076

[10] S. Tangkawanit, J. Pooksook, J. Ieamsaard and P. Sornkhom, "OCR Application for Cancer Care," 2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Chiang Mai, Thailand, 2022, pp. 489-493, doi: 0.23919/APSIPAASC55919.2022.9980078

[11] Daniela Gifu, AI-backed OCR in Healthcare, Procedia Computer Science, Volume 207, 2022, Pages 1134-1143, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2022.09.169

[12] Yuan, S. et al. MCIC: Multimodal Conversational Intent Classification for E-commerce Customer Service. In: Lu, W., Huang, S., Hong, Y., Zhou, X.

(eds) Natural Language Processing and Chinese Computing. NLPCC 2022. Lecture Notes in Computer Science(), vol 13551. Springer, Cham. https://doi.org/10.1007/978-3-031-17120-8_58

[13] Sandeep Dwarkanath Pande, Pramod Pandurang Jadhav, Rahul Joshi, Amol Dattatray Sawant, Vaibhav Muddebihalkar, Suresh Rathod, Madhuri Navnath Gurav, Soumitra Das, -Digitization of handwritten Devanagari text using CNN transfer learning – A better customer service support, Neuroscience Informatics, Volume 2, Issue 3, 2022, 100016,ISSN 2772-5286, https://doi.org/10.1016/j.neuri.2021.100016

[14] Faizullah, S.; Ayub, M.S.; Hussain, S.; Khan, M.A. A Survey of OCR in Arabic Language: Applications, Techniques, and Challenges. *Appl. Sci.* 2023, *13*, 4584. https://doi.org/10.3390/app13074584

[15] Imane Guellil, Houda Saâdane, Faical Azouaou, Billel Gueni, Damien Nouvel, Arabic natural language processing: An overview, Journal of King Saud University - Computer and Information Sciences, Volume 33, Issue 5, 2021, Pages 497-507, ISSN 1319-1578, https://doi.org/10.1016/j.jksuci.2019.02.006

[16] Alghyaline, S. (2023). Arabic Optical Character Recognition: A Review. CMES-Computer Modeling in Engineering & Sciences, 135(3), 1825–1861

[17] E. Chammas, C. Mokbel and L. Likforman-Sulem, "Arabic handwritten document preprocessing and recognition," *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, Tunis, Tunisia, 2015, pp. 451-455, doi: 10.1109/ICDAR.2015.7333802

[18] H. Boukerma and N. Farah, "Preprocessing Algorithms for Arabic Handwriting Recognition Systems," *2012 International Conference on Advanced Computer Science Applications and Technologies (ACSAT)*, Kuala Lumpur, Malaysia, 2012, pp. 318-323, doi: 10.1109/ACSAT.2012.59

[19] Sahlol, A.T., Suen, C.Y., Elbasyoni, M.R., Sallam, A.A. (2014). Investigating of Preprocessing Techniques and Novel Features in Recognition of Handwritten Arabic Characters. In: El Gayar, N., Schwenker, F., Suen, C. (eds) Artificial Neural Networks in Pattern Recognition. ANNPR 2014. Lecture Notes in Computer Science(), vol 8774. Springer, Cham. https://doi.org/10.1007/978-3-319-11656-3_24

[20] Amlan Kundu, Yang He, Paramvir Bahl, Recognition of handwritten word: First and second order hidden Markov model based approach, Pattern Recognition, Volume 22, Issue 3, 1989, Pages 283-297, ISSN 0031-3203, https://doi.org/10.1016/0031-3203(89)90076-9

[21] Gilloux, M. (1994). Hidden Markov Models in Handwriting Recognition. In: Impedovo, S. (eds) Fundamentals in Handwriting Recognition. NATO ASI Series, vol 124. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-78646-4_15

[22] Alzubaidi, L., Zhang, J., Humaidi, A.J. et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data 8, 53 (2021). https://doi.org/10.1186/s40537-021-00444-8

[23] Sarker, I.H. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. SN COMPUT. SCI.2, 420 (2021).https://doi.org/10.1007/s42979-021-00815-1

[24] O. Morillot, L. Likforman-Sulem and E. Grosicki, "Comparative Study of HMM and BLSTM Segmentation-Free Approaches for the Recognition of Handwritten Text-Lines," *2013 12th International Conference on Document Analysis and Recognition*, Washington, DC, USA, 2013, pp. 783-787, doi: 10.1109/ICDAR.2013.160

[25] Wang, Xin ; Takaki, Shinji ; Yamagishi, Junichi. / A Comparative Study of the Performance of HMM, DNN, and RNN based Speech Synthesis Systems Trained on Very Large Speaker-Dependent Corpora. Proceedings of 9th ISCA Speech Synthesis Workshop. 2016. pp. 125-128

[26] Yong Yu, Xiaosheng Si, Changhua Hu, Jianxun Zhang; A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. *Neural Comput* 2019; 31 (7): 1235–1270. doi: https://doi.org/10.1162/neco_a_01199

[27] Firat Kizilirmak, Berrin Yanikoglu. CNN-BiLSTM model for English Handwriting Recognition: Comprehensive Evaluation on the IAM Dataset. Computer Vision and Pattern Recognition https://doi.org/10.48550/arXiv.2307.00664

[28] Gader, T.; Chibani, I. and Echi, A. (2023). Arabic Handwriting off-Line Recognition Using ConvLSTM-CTC. In Proceedings of the 12th International Conference on Pattern Recognition Applications and Methods - ICPRAM; ISBN 978-989-758-626-2; ISSN 2184-4313, SciTePress, pages 529-533. DOI: 10.5220/0011794700003411

[29] [29] M. Geetha, R. C. Suganthe, S. K. Nivetha, S. Hariprasath, S. Gowtham and C. S. Deepak, "A Hybrid Deep Learning Based Character Identification Model Using CNN, LSTM, And CTC To Recognize Handwritten English Characters And Numerals," *2022*

International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2022, pp. 1-6, doi: 10.1109/ICCCI54379.2022.9740746.

[30] Bisht, M., Gupta, R. Offline Handwritten Devanagari Word Recognition Using CNN-RNN-CTC. SN COMPUT. SCI. 4, 88 (2023). https://doi.org/10.1007/s42979-022-01461-x.

[31] Mars, A., Dabbabi, K., Zrigui, S., Zrigui, M. (2023). Combination of DE-GAN with CNN-LSTM for Arabic OCR on Images with Colorful Backgrounds. In: Nguyen, N.T., et al. Advances in Computational Collective Intelligence. ICCCI 2023. Communications in Computer and Information Science, vol 1864. Springer, Cham. https://doi.org/10.1007/978-3-031-41774-0_46.

[32] Xiaohui Huang, Lisheng Qiao, Wentao Yu, Jing Li, Yanzhou Ma. « End-to-End Sequence Labeling via Convolutional Recurrent Neural Network with a Connectionist Temporal Classification Layer » International Journal of Computational Intelligence Systems 2020 10.2991/ijcis.d.200316.001

[33] Gilloux, M. (1994). Hidden Markov Models in Handwriting Recognition. In: Impedovo, S. (eds) Fundamentals in Handwriting Recognition. NATO ASI Series, vol 124. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-78646-4_15

[34] Roman Bertolami, Horst Bunke. Hidden Markov model-based ensemble methods for offline handwritten text line recognition, Pattern Recognition, Volume 41, Issue 11, 2008, Pages 3452-3460, ISSN 0031-3203, https://doi.org/10.1016/j.patcog.2008.04.003

[35] Plötz, T., Fink, G.A. Markov models for offline handwriting recognition: a survey. IJDAR 12, 269–298 (2009). https://doi.org/10.1007/s10032-009-0098-4

[36] Mor, B., Garhwal, S. & Kumar, A. A Systematic Review of Hidden Markov Models and Their Applications. Arch Computat Methods Eng 28, 1429–1448 (2021). https://doi.org/10.1007/s11831-020-09422-4

[37] Núria Cirera; Alicia Fornés; Josep Lladós. Hidden Markov model topology optimization for handwriting recognition. Published in: 2015 13th International Conference on Document Analysis and Recognition (ICDAR) IEEE. DOI: 10.1109/ICDAR.2015.7333837

[38] Rabi Mouhcine, Amrouch Mustapha, Mahani Zouhir. Recognition of cursive Arabic handwritten text using embedded training based on HMMs, Journal of Electrical Systems and Information Technology, Volume 5, Issue 2, 2018, Pages 245-251,ISSN,2314-7172, https://doi.org/10.1016/j.jesit.2017.02.001

[39] Natarajan, P., Saleem, S., Prasad, R., MacRostie, E., Subramanian, K. (2008). Multi-lingual Offline Handwriting Recognition Using Hidden Markov Models: A Script-Independent Approach. In: Doermann, D., Jaeger, S. (eds) Arabic and Chinese Handwriting Recognition. SACH 2006. Lecture Notes in Computer Science, vol 4768. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-78199-8_14

[40] Adrià Giménez, Ihab Khoury, Jesús Andrés-Ferrer, Alfons Juan. Handwriting word recognition using windowed Bernoulli HMMs, Pattern Recognition Letters, Volume 35, 2014, Pages 149-156, ISSN 0167-8655, https://doi.org/10.1016/j.patrec.2012.09.002

[41] Frinken, V., Peter, T., Fischer, A., Bunke, H., Do, TMT., Artieres, T. (2009). Improved Handwriting Recognition by Combining Two Forms of Hidden Markov Models and a Recurrent Neural Network. In: Jiang, X., Petkov, N. (eds) Computer Analysis of Images and Patterns. CAIP 2009. Lecture Notes in Computer Science, vol 5702. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-03767-2_23

[42] Y. Bengio, Y. LeCun, C. Nohl and C. Burges, "LeRec: A NN/HMM Hybrid for On-Line Handwriting Recognition," in *Neural Computation*, vol. 7, no. 6, pp. 1289-1303, Nov. 1995, doi: 10.1162/neco.1995.7.6.1289

[43] Rabi, M., Amrouch, M., Mahani, Z. (2018). Hybrid HMM/MLP Models for Recognizing Unconstrained Cursive Arabic Handwritten Advanced Information Technology, Services and Systems. AIT2S 2017. Lecture Notes in Networks and Systems, vol 25. Springer, Cham. https://doi.org/10.1007/978-3-319-69137-4_39

[44] Rabi, M. Amrouch,M and Mahani.Z A Survey of Contextual Handwritten Recognition Systems based HMMs for Cursive Arabic and Latin Script. International Journal of Computer Applications 160(2):31-37, February 2017. publisher = {Foundation of Computer Science (FCS), NY, USA} Doi :10.5120/ijca2017912982

[45] V. Marti and H. Bunke, "Handwritten sentence recognition," *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, Barcelona, Spain, 2000, pp. 463-466 vol.3, doi: 10.1109/ICPR.2000.903584

[46] Marti and H. Bunke. Using a statistical language model to improve the performance of an hmm-based

cursive handwriting recognition system. International Journal of Pattern Recognition and Artificial IntelligenceVol. 15, No. 01, pp. 65-90 (2001)SPECIAL ISSUE: Hidden Markov Models in Vision; Edited by H. Bunke and T. Caelli https://doi.org/10.1142/S0218001401000848

[47] Wassim Swaileh, Yann Soullard, Thierry Paquet. A unified multilingual handwriting recognition system using multigrams sub-lexical units. Pattern Recognition Letters, 2019, 121, pp.68-76.10.1016/j.patrec.2018.07.027. hal-02075654

[48] Lalita Kumari, Sukhdeep Singh, Vaibhav Varish Singh Rathore, Anuj Sharma. Lexicon and attention based handwritten text recognition system. Machine Graphics & Vision 2022 Pages: 75—92 DOI: 10.22630/MGV.2022.31.1.4

[49] Mustapha Amrouch, Mouhcine Rabi & Youssef Es-Saady . Convolutional Feature Learning and CNN Based HMM for Arabic Handwriting Recognition Image and Signal Processing: 8th International Conference, ICISP 2018, Cherbourg, France, July 2-4, 2018, Proceedings Jul 2018 Pages 265–274 https://doi.org/10.1007/978-3-319-94211-7_29

[50] Porwal, U., Fornés, A. & Shafait, F. Advances in handwriting recognition. International Journal on Document Analysis and Recognition (IJDAR) volume 25, pages241–243 (2022) https://doi.org/10.1007/s10032-022-00421-8

[51] Savita Ahlawat, Amit Choudhary , Anand Nayyar, Saurabh Singh and Byungun Yoon. Improved Handwritten Digit Recognition Using Convolutional Neural Networks (CNN) Sensors 2020, 20, 3344; doi:10.3390/s20123344

[52] A. A. Rangari, S. Das and R. D, "Cursive Handwriting Recognition Using CNN with VGG-16," *2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF)*, Chennai, India, 2023, pp. 1-6, doi: 10.1109/ICECONF57129.2023.10083561

[53] Chahi, A., El-merabet, Y., Ruichek, Y. et al. An effective DeepWINet CNN model for off-line text-independent writer identification. Pattern Anal Applic 26, 1539–1556 (2023). https://doi.org/10.1007/s10044-023-01186-4

[54] Mamouni El Mamoun An Effective Combination of Convolutional Neural Network and Support Vector Machine Classifier for Arabic Handwritten Recognition. Aut. Control Comp. Sci. 57, 267–275 (2023). https://doi.org/10.3103/S0146411623030069

[55] M. M. Al-Taee, S. B. H. Neji and M. Frikha, "Handwritten Recognition: A survey," 2020 IEEE 4th International Conference on Image Processing, Applications and Systems (IPAS), Genova, Italy, 2020, pp. 199-205, doi: 10.1109/IPAS50080.2020.9334936

[56] Albattah, W.; Albahli, S. Intelligent Arabic Handwriting Recognition Using Different Standalone and Hybrid CNN Architectures. Appl. Sci. 2022, 12, 10155. https://doi.org/10.3390/app121910155

[57] Islam, M. et al. (2023). Efficient Approach to Using CNN-Based Pre-trained Models in Bangla Handwritten Digit Recognition. In: Smys, S., Tavares, J.M.R.S., Shi, F. (eds) Computational Vision and Bio-Inspired Computing. Advances in Intelligent Systems and Computing, vol 1439. Springer, Singapore. https://doi.org/10.1007/978-981-19-9819-5_50

[58] Alzubaidi, L., Zhang, J., Humaidi, A.J. et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data 8, 53 (2021). https://doi.org/10.1186/s40537-021-00444-8

[59] Chen, Y., Zhang, H., Liu, CL. (2023). Improved Learning for Online Handwritten Chinese Text Recognition with Convolutional Prototype Network. In: Fink, G.A., Jain, R., Kise, K., Zanibbi, R. (eds) Document Analysis and Recognition - ICDAR 2023. ICDAR 2023. Lecture Notes in Computer Science, vol 14190. Springer, Cham. https://doi.org/10.1007/978-3-031-41685-9_3

[60] D. Teng, D. Fan, F. Bai and Y. Pan, "End-to-End Model Based on Bidirectional LSTM and CTC for Online Handwritten Mongolian Word Recognition," *2022 12th International Conference on Information Science and Technology (ICIST)*, Kaifeng, China, 2022, pp. 271-275, doi: 10.1109/ICIST55546.2022.9926844

[61] F. Kreß *et al.*, "Hardware-aware Workload Distribution for AI-based Online Handwriting Recognition in a Sensor Pen," *2022 11th Mediterranean Conference on Embedded Computing (MECO)*, Budva, Montenegro, 2022, pp. 1-4, doi: 10.1109/MECO55406.2022.9797131

[62] Alwajih, F., Badr, E., & Abdou, S. (2022). Transformer-based Models for Arabic Online Handwriting Recognition. International Journal of Advanced Computer Science and Applications, 13(5). DOI :10.14569/IJACSA.2022.01305102

[63] S. Wu, Y. Li and Z. Wang, "Improving CTC-based Handwritten Chinese Text Recognition with Cross-Modality Knowledge Distillation and Feature Aggregation," *2023 IEEE International Conference on Multimedia and Expo (ICME)*, Brisbane, Australia,

2023, pp. 792-797, doi: 10.1109/ICME55011.2023.00141

[64] M. B, S. S, T. M. V. K and V. Ramanan P, "Offline Recognition Of Handwritten Text Using Combination Of Neural Networks," *2023 8th International Conference on Communication and Electronics Systems (ICCES)*, Coimbatore, India, 2023, pp. 865-870, doi: 10.1109/ICCES57224.2023.10192850

[65] M. S. Akter, H. Shahriar, A. Cuzzocrea, N. Ahmed and C. Leung, "Handwritten Word Recognition using Deep Learning Approach: A Novel Way of Generating Handwritten Words," *2022 IEEE International Conference on Big Data (Big Data)*, Osaka, Japan, 2022, pp. 5414-5423, doi: 10.1109/BigData55660.2022.10021025

[66] Sehr Zia, N., Naeem, M.F., Raza, S.M.K. et al. A convolutional recursive deep architecture for unconstrained Urdu handwriting recognition. Neural Comput & Applic 34, 1635–1648 (2022). https://doi.org/10.1007/s00521-021-06498-2

[67] Boualam, M.; Elfakir, Y.; Khaissidi, G.; Mrabti, M. Arabic Handwriting Word Recognition Based on Convolutional Recurrent Neural Network. In WITS 2020, Proceeding of the 6th International Conference on Wireless Technologies, Embedded, and Intelligent Systems, Fez, Morocco, 14–16 October 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 877–885

[68] R. E. Shtaiwi, G. A. Abandah and S. A. Sawalhah, "End-to-End Machine Learning Solution for Recognizing Handwritten Arabic Documents," *2022 13th International Conference on Information and Communication Systems (ICICS)*, Irbid, Jordan, 2022, pp. 180-185, doi: 10.1109/ICICS55353.2022.9811155

[69] Mohammad Fasha, Bassam Hammo, Nadim Obeid and Jabir AlWidian, “A Hybrid Deep Learning Model for Arabic Text Recognition” International Journal of Advanced Computer Science and Applications(IJACSA), 11(8), 2020. http://dx.doi.org/10.14569/IJACSA.2020.0110816

[70] Khaled, M. Moneb and Alzebdeh, Ala' and Lataifeh, Mohammad and Lulu, Leena and Elnagar, Ashraf M., A Hybrid Deep Learning Approach for Arabic Handwritten Recognition: Exploring the Complexities of the Arabic Language. Rochester: New York, NY, USA, 2023. Available at SSRN: http://dx.doi.org/10.2139/ssrn.4399243

[71] S. Watanabe, T. Hori, S. Kim, J. R. Hershey and T. Hayashi, "Hybrid CTC/Attention Architecture for End-to-End Speech Recognition," in *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 8, pp. 1240-1253, Dec. 2017, doi: 10.1109/JSTSP.2017.2763455

[72] Müller, M., Stüker, S., Waibel, A. (2017). Language Adaptive Multilingual CTC Speech Recognition. In: Karpov, A., Potapova, R., Mporas, I. (eds) Speech and Computer. SPECOM 2017. Lecture Notes in Computer Science(), vol 10458. Springer, Cham. https://doi.org/10.1007/978-3-319-66429-3_47

[73] Kürzinger, L., Winkelbauer, D., Li, L., Watzel, T., Rigoll, G. (2020). CTC-Segmentation of Large Corpora for German End-to-End Speech Recognition. In: Karpov, A., Potapova, R. (eds) Speech and Computer. SPECOM 2020. Lecture Notes in Computer Science(), vol 12335. Springer, Cham. https://doi.org/10.1007/978-3-030-60276-5_27

[74] Deshmukh, A.M. 2020. Comparison of Hidden Markov Model and Recurrent Neural Network in Automatic Speech Recognition. European Journal of Engineering and Technology Research. 5, 8 (Aug. 2020), 958–965. DOI: https://doi.org/10.24018/ejeng.2020.5.8.2077

[75] Y. Zhang and X. Lu, "A Speech Recognition Acoustic Model Based on LSTM -CTC," *2018 IEEE 18th International Conference on Communication Technology (ICCT)*, Chongqing, China, 2018, pp. 1052-1055, doi: 10.1109/ICCT.2018.8599961

[76] T. Raissi, W. Zhou, S. Berger, R. Schlüter and H. Ney, "HMM vs. CTC for Automatic Speech Recognition: Comparison Based on Full-Sum Training from Scratch," *2022 IEEE Spoken Language Technology Workshop (SLT)*, Doha, Qatar, 2023, pp. 287-294, doi: 10.1109/SLT54892.2023.10022967

[77] Parres, D., Paredes, R. (2023). Fine-Tuning Vision Encoder–Decoder Transformers for Handwriting Text Recognition on Historical Documents. In: Fink, G.A., Jain, R., Kise, K., Zanibbi, R. (eds) Document Analysis and Recognition - ICDAR 2023. ICDAR 2023. Lecture Notes in Computer Science, vol 14190. Springer, Cham. https://doi.org/10.1007/978-3-031-41685-9_16

[78] Barrere, K., Soullard, Y., Lemaitre, A., Coüasnon, B. (2022). A Light Transformer-Based Architecture for Handwritten Text Recognition. In: Uchida, S., Barney, E., Eglin, V. (eds) Document Analysis Systems. DAS 2022. Lecture Notes in Computer Science, vol 13237. Springer, Cham. https://doi.org/10.1007/978-3-031-06555-2_19

[79] A. Mostafa *et al*., "OCFormer: A Transformer-Based Model For Arabic Handwritten Text Recognition," *2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, Cairo, Egypt, 2021,

pp. 182-186, doi: 10.1109/MIUCC52538.2021.9447608

[80] Saleh Momeni, Bagher BabaAli. A Transformer-based Approach for Arabic Offline Handwritten Text Recognition. Computer Vision and Pattern Recognition 2023. https://doi.org/10.48550/arXiv.2307.15045

[81] Yongping Dan , Zongnan Zhu, Weishou Jin, and Zhuo Li. PF-ViT: Parallel and Fast Vision Transformer for Offline Handwritten Chinese Character Recognition. Computational Intelligence and Neuroscience Volume 2022, Article ID 8255763, 11 pages. https://doi.org/10.1155/2022/8255763

[82] Geng, S.; Zhu, Z.; Wang, Z.; Dan, Y.; Li, H. LW-ViT: The Lightweight Vision Transformer Model Applied in Offline Handwritten Chinese Character Recognition. *Electronics* 2023, *12*, 1693. https://doi.org/10.3390/electronics12071693

[83] Marwa Dhiaf, Ahmed Cheikh Rouhou, Yousri Kessentini, Sinda Ben Salem. MSdocTr-Lite: A lite transformer for full page multi-script handwriting recognition, Pattern Recognition Letters, Volume 169, 2023, Pages 28-34, ISSN 0167-8655, https://doi.org/10.1016/j.patrec.2023.03.020

[84] Ganai, A.F., Khurshid, F. (2023). Handwritten Urdu Recognition Using BERT with Vision Transformers. In: Parah, S.A., Hurrah, N.N., Khan, E. (eds) Intelligent Multimedia Signal Processing for Smart Ecosystems. Springer, Cham. https://doi.org/10.1007/978-3-031-34873-0_15

[85] Riaz, N., Arbab, H., Maqsood, A. et al. Conv-transformer architecture for unconstrained off-line Urdu handwriting recognition. IJDAR 25, 373–384 (2022). https://doi.org/10.1007/s10032-022-00416-5

[86] Kaur, S., Bawa, S., Kumar, R. (2023). Evaluating Generative Adversarial Networks for Gurumukhi Handwritten Character Recognition (CR). In: Sisodia, D.S., Garg, L., Pachori, R.B., Tanveer, M. (eds) Machine Intelligence Techniques for Data Analysis and Signal Processing. Lecture Notes in Electrical Engineering, vol 997. Springer, Singapore. https://doi.org/10.1007/978-981-99-0085-5_41

[87] Jha, G., Cecotti, H. Data augmentation for handwritten digit recognition using generative adversarial networks. Multimed Tools Appl 79, 35055–35068 (2020). https://doi.org/10.1007/s11042-020-08883-w

[88] Li, J., Song, G. & Zhang, M. Occluded offline handwritten Chinese character recognition using deep convolutional generative adversarial network and improved GoogLeNet. Neural Comput & Applic 32, 4805–4819 (2020). https://doi.org/10.1007/s00521-018-3854-x

[89] Yeleussinov, A.; Amirgaliyev, Y.; Cherikbayeva, L. Improving OCR Accuracy for Kazakh Handwriting Recognition Using GAN Models. *Appl. Sci.* 2023, *13*, 5677. https://doi.org/10.3390/app13095677

[90] Alwaqfi, Y.; Mohamad, M.; Al-Taani, A. Generative Adversarial Network for an Improved Arabic Handwritten Characters Recognition. Int. J. Adv. Soft Comput. Its Appl. 2022, 14, 177–195. DOI: 10.15849/IJASCA.220328.12

[91] Eltay M, Zidouri A, Ahmad I, Elarian Y. 2022. Generative adversarial network based adaptive data augmentation for handwritten Arabic text recognition. PeerJ Computer Science 8:e861 https://doi.org/10.7717/peerj-cs.861

[92] Mustapha, I.B., Hasan, S., Nabus, H. et al. Conditional Deep Convolutional Generative Adversarial Networks for Isolated Handwritten Arabic Character Generation. Arab J Sci Eng 47, 1309–1320 (2022). https://doi.org/10.1007/s13369-021-05796-0

[93] Sana Khamekhem Jemni, Mohamed Ali Souibgui, Yousri Kessentini, Alicia Fornés, Enhance to read better: A Multi-Task Adversarial Network for Handwritten Document Image Enhancement, Pattern Recognition, Volume 123, 2022, 108370, ISSN 0031-3203, https://doi.org/10.1016/j.patcog.2021.108370

[94] M. Pechwitz, S. Snoussi Maddouri, V. Märgner, N. Ellouze , H. Amiri, "IFN/ENIT-DATABASE OF HANDWRITTEN ARABIC WORDS" , in the 7th Colloque International Francophone sur l'Ecrit et le Document , CIFED 2002, Oct. 21-23, 2002, Hammamet, Tunis, (2002)

[95] Eltay, M., Zidouri, A., Ahmad, I., Elarian, Y. (2021). Improving Handwritten Arabic Text Recognition Using an Adaptive Data-Augmentation Algorithm. In: Barney Smith, E.H., Pal, U. (eds) Document Analysis and Recognition – ICDAR 2021 Workshops. ICDAR 2021. Lecture Notes in Computer Science(), vol 12916. Springer, Cham. https://doi.org/10.1007/978-3-030-86198-8_23

[96] Heil, R., Breznik, E. (2023). A Study of Augmentation Methods for Handwritten Stenography Recognition. In: Pertusa, A., Gallego, A.J., Sánchez, J.A., Domingues, I. (eds) Pattern Recognition and Image Analysis. IbPRIA 2023. Lecture Notes in Computer Science, vol 14062. Springer, Cham. https://doi.org/10.1007/978-3-031-36616-1_11

[97] M. Elleuch, R. Maalej and M. Kherallah, “A new design based-SVM of the CNN classifier architecture with dropout for offline Arabic handwritten

recognition," Procedia Computer Science, vol. 80, pp. 1712– 1723, 2016

[98] Chen, HJ., Fathoni, H., Wang, ZY., Lien, KY., Yang, CT. (2023). A Real-Time Streaming Application for License Plate Recognition Using OpenALPR. In: Deng, DJ., Chao, HC., Chen, JC. (eds) Smart Grid and Internet of Things. SGIoT 2022. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 497. Springer, Cham. https://doi.org/10.1007/978-3-031-31275-5_33

[99] H. Scheidl, S. Fiel, R. Sablatnig, Word beam search: A connectionist tem- poral classification decoding algorithm, in: 2018 16th International Confer- ence on Frontiers in Handwriting Recognition (ICFHR), IEEE, 2018, pp. 253–258