المجلة العراقية للبحوث الانسانية والاجتماعية والعلمية
Iraqi Journal of Humanitarian, Social and Scientific Research
Print ISSN 2710-0952 Electronic ISSN 2790-1254
العدد 13 آيار 2024
No.13A May 2024

Automated Evaluation of Students' Handwritten Answer to Arabic Poetic Poems

Huda Sabah Shuker
University of Technology, Baghdad
Cs.22.08@grad.uotechnology.edu.iq
Asst. Prof. Dr. Raheem Abdul Sahib
University of Technology, Baghdad
Computer Sciences Department
Prof. De. Abdul Monem S. Rahma
University of Technology, Baghdad
Computer Sciences Department
110003@uotechnology.edu.iq
110137@uotechnology.edu.iqComputer Sciences Department

Abstract: Arabic handwriting script is curly and contains some dots, which complicates the task to an extent. The study is about the proposal for the creation of a deep learning platform that can offer automatic handwriting analysis of Arabic documents without human intervention. The Convolutional Neural Network (CNN) architecture is the main component, and an Arabic sample dataset, which comes from Kaggle website and can be used to train the model, is developed. Noise reduction can be achieved depending on the methods such as gaussian blur, bilateral filtering, adaptive thresholding, and morphological processes, which all help in detecting features. The model performance will be assessed by using as well as by the several quality measures like accuracy, precision and recall with unconscious focus on the handwriting quality as demonstrated through symbols like alignment, spaces between letters and overall legibility. The results have shown that the CNN-based systems can be successfully used for automatic evaluation of Arabic handwriting and, in turn, given various directions for future research and development. This work advances handwriting evaluation approaches for Arabic script, with implications for educational and technical applications.

Keywords: Deep Learning; Convolutional Neural Network; Arabic handwriting; Arabic Poetic Poems; Gaussian Blur.

التقييم الآلي لإجابات الطلاب المكتوبة بخط اليد على القصائد الشعرية العربية

هدى صباح شكر
جامعة التكنولوجية قسم علوم الحاسوب بغداد
أ.م.د. رحيم عبد الصاحب عكله
جامعة التكنولوجية قسم علوم الحاسوب بغداد
أ.م.د. عبد المنعم صالح رحمه
جامعة التكنولوجية قسم علوم الحاسوب بغداد

المستخلص: إن الكتابة اليدوية العربية مجعدة وتحتوي على بعض النقاط، مما يعقد المهمة إلى حد ما. وتتعلق الدراسة باقتراح إنشاء منصة تعلم عميق يمكنها تقديم تحليل تلقائي للكتابة اليدوية للوثائق العربية دون تدخل بشري. ويشكل هيكل الشبكة العصبية التلافيفية **(CNN)** المكون الرئيسي، كما تم تطوير مجموعة بيانات عينة باللغة العربية، والتي تأتي من موقع **Kaggle** ويمكن استخدامها لتدريب النموذج. ويمكن تحقيق تقليل الضوضاء اعتمادًا على طرق مثل التمويه الغاوسي، والتصفية الثنائية، والعتبة التكيفية، والعمليات الصرفية، والتي تساعد جميعها في اكتشاف الميزات. وسيتم تقييم أداء النموذج باستخدام العديد من مقاييس الجودة مثل الدقة والضبط والتذكر مع التركيز اللاواعي على جودة الكتابة اليدوية كما يتضح من خلال الرموز مثل المحاذاة والمسافات بين الحروف والوضوح العام. وقد أظهرت النتائج أنه يمكن استخدام الأنظمة القائمة على **CNN** بنجاح للتقييم التلقائي للكتابة اليدوية العربية، وبالتالي إعطاء اتجاهات مختلفة للبحث والتطوير في المستقبل. ويعمل هذا العمل على تطوير مناهج تقييم الكتابة اليدوية للنص العربي، مع الآثار المترتبة على التطبيقات التعليمية والتقنية.

الكلمات المفتاحية: التعلم العميق؛ الشبكة العصبية التلافيفية ؛ الكتابة اليدوية العربية؛ القصائد الشعرية العربية؛ التمويه الغاوسي.

1. Introduction

Handwriting evaluation is important in many aspects of education and diagnosis as it offers pertinent details about a student's individual progress in learning, motor, and cognitive development. Traditional manual evaluation methods can be lengthy, optional, and influenced by human mistakes for languages that are script and cursive by nature, heavily depending on diacritic marks to disambiguate [1]. In this study, how to create and apply a CNN system for automatic Arabic handwriting assessment is studied. The database provider of Arabic handwriting, which is derived from Kaggle, a popular website for machine learning competitions and datasets, has been used for our purpose. To ensure correct running out of the CNN, we design the pre-processing pipeline that includes noise reduction techniques and image binarization to achieve the philosophy by improving the text readability [3].

The proposed method is evaluated based on factors studied, including how well you write the letters, how uniform the spacing is, if the letters are aligned properly with each other, and how easily the entire handwriting can be read. From a practical standpoint, the method relies on measures like accuracy, precision, and recall to evaluate the system's performance in the Arabic handwriting evaluation context. Our contribution is to the development of robust and trustworthy techniques for automated handwriting evaluation in the context of Arabic script.

2. Related Works

Independent of the numerous recent dataset creations and machine learning methodologies, developments in Arabic handwriting recognition have been evident. The OnHW dataset, which consists of online slant writing data, was presented by Ott et al. [4], who claimed to use IMU-enhanced ball point pens. The machine learning approach of their technique proved to be practical in the recognition of complicated handwritten styles. While the performance of this approach is very similar to that of Ott et al.'s [5] research, which examined the

triplet loss function in auxiliary cross-modal representation learning to improve the accuracy of online handwriting recognition systems, this approach is much more precise.

In the aspect of automatic essay grading, Chen et al. [6] have assessed English writing using CNN and ordinal regression (OR), providing a framework that can be adjusted and reused for Arabic handwriting evaluation. Arabic script recognition system developed by Altwaijry and Al-Turaiki [7] with the help of CNNs, as shown by deep learning capability for Arabic script.

While Momeni and Babaali [8] used transformer models that recognize Arabic offline handwritten text in their unique approach, this resulted in consistently the best outcome. Shaker's work is entitled "A Brief Analysis of the Problems and Proposed Solutions in the Field of Automatic Transliteration of Arabic Handwritten Script" [9].

The DL method proposed by Rahaman and Mahmud [10] provides an automated content evaluation technique that can be customized to replace the need for Arabic handwriting in answer scripts. Fasha et al. [11] have proposed a deep learning network that is based on a hybrid Arabic text recognition model and takes advantage of the benefits of many neural network designs.

Beseiso and Alzahrani [12] focused on the use of BERT embeddings to evaluate learned Arabic sentences issued by handwriting; such models are well-suited for such a task. As a final point, Alikaniotis et al. [13] demonstrated the adaptability of neural networks in language processing problems while exploring the use of neural networks with the purpose of automated text scoring.

The project plan is separated into two sections: the first covers the introduction, related work, and its limits, while the second section is devoted to the manuscript's suggested technique.

➢ Limitations

Although these studies have made progress, there are still obstacles to overcome. These include the intricacy of Arabic script, the need for huge and diverse datasets to build robust models, and the computer resources needed to train complicated neural networks

3. Proposed Methodology

This project aims to develop a robust and accurate system for recognizing handwritten text using CNN.

Algorithm 1: Training CNN model
Input: Handwritten Images dataset **Output:** H5 Model
Begin **Step 1:** Data Collection and Preprocessing: Load and preprocess

handwritten images by Convert to grayscale and Apply noise reduction and apply morphological operations

Step 2: CNN Model Design and Training: Define and train a CNN model by Implement convolutional layers for feature extraction and add pooling layers for spatial downsampling and Utilize ReLU activation function and Use Adam optimizer for weight adjustments

Step 3: Feature Extraction and Evaluation: Evaluate model performance by Load a new handwritten image and Preprocess and Extract features using the trained CNN model

Step 4: Calculate accuracy, precision, recall.

End

Algorithm 2: User interface

Input: Handwritten Image

Output: Predicted texts and moments in CSV

Begin

Step 1: Select image

Step 2: Preprocess the image

Step 3: Predict the Handwritten texts in image

Step 4: Calculate Moments for each word

Step 5: Save results in CSV.

End

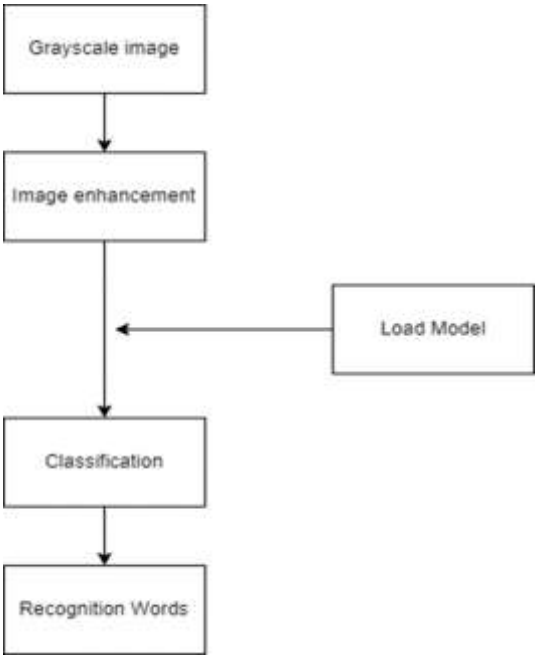

Figure. 1 System Flowchart

4. Experimental Setup

By offering a fresh and practical method for assessing handwriting quality, this initiative advances the subject of Arabic handwriting evaluation. Automated exam grading and programme evaluation for handwriting improvement are examples of potential uses.

4.1 CNN Model Architecture

Using components from the "Training the model (DCNNs + Bidirectional LSTMs with CTC Loss)" technique, the system uses a DCNN archite cture. In order to extract features from the handwritten characters, this design combines convolutional layers with bidirectional LSTM networks to collect sequential information.

Because Connectionist Temporal Classification (CTC) loss functions are good at handling sequences with varying durations, like those found in handwritten text, they are used during training.

This description could encompass:

- The quantity and kind of convolutional layers (filter count, kernel size, etc.) utilised to extract features).
- The use of pooling layers, such as max pooling, to reduce dimensionality.
- The introduction of non-linearity through the use of activation functions such as ReLU.

Bidirectional LSTM layers are used to capture sequential information in the characters.

4.2Dataset

The Arabic Handwritten Characters Dataset, a freely accessible collection of Arabic handwritten characters available on Kaggle, is utilised in the experiment. This dataset provides a representative sample for system evaluation and training. It consists of 8728 grayscale photos that reflect different writing styles and character sizes.

4.3Training Details

The training procedure uses an appropriate optimizer (such as Adam) to update the network weights and the CTC loss function to guide the learning process. The hyperparameters used during training, such as batch size, number of epochs, and learning rate, should be documented for reproducibility.

Name	Isolated	Initial	Medial	Final
Alif	ا	ا	ـا	ـا
Baa	ب	بـ	ـبـ	ـب
Taa	ت	تـ	ـتـ	ـت
Thaa	ث	ثـ	ـثـ	ـث
Jeem	ج	جـ	ـجـ	ـج
7aa	ح	حـ	ـحـ	ـح
7haa	خ	خـ	ـخـ	ـخ
Daa	د	د	ـد	ـد
Thal	ذ	ذ	ـذ	ـذ
Raa	ر	ر	ـر	ـر
Zaa	ز	ز	ـز	ـز
Seen	س	سـ	ـسـ	ـس
Sheen	ش	شـ	ـشـ	ـش
Saad	ص	صـ	ـصـ	ـص
Dhaad	ض	ضـ	ـضـ	ـض
6aa	ط	طـ	ـطـ	ـط
6haa	ظ	ظـ	ـظـ	ـظ
Aeen	ع	عـ	ـعـ	ـع
Gheen	غ	غـ	ـغـ	ـغ
Faa	ف	فـ	ـفـ	ـف
Qaaf	ق	قـ	ـقـ	ـق
Laam	ل	لـ	ـلـ	ـل
Meem	م	مـ	ـمـ	ـم
Noon	ن	نـ	ـنـ	ـن
Haa	ه	هـ	ـهـ	ـه
Wow	و	و	ـو	ـو
Yaa	ي	يـ	ـيـ	ـي

Figure. 2 Arabic Characters [18]

4.4 Evaluation Metrics

To assess the system's performance on a held-out testing set, the following evaluation metrics were chosen:

- **Accuracy:** Measures the proportion of correctly identified handwritten characters.

Accuracy = (Number of Correctly Classified Samples) / (Total Number of Samples) [14]

- **Precision:** Evaluates the system's ability to avoid false positives, calculated as the ratio of true positives to all predicted positives.

Precision = True Positives / (True Positives + False Positives) [15]

- **Recall:** Assesses the system's ability to avoid false negatives, calculated as the ratio of true positives to all actual positives.

Recall = True Positives / (True Positives + False Negatives) [15].

5. Discussion

The proposed work is an expands of the writer research on Arabic writing recognition and assessment. Even though earlier studies have made significant strides toward to that objective, the current research suggests a new approach that would focus rather on the assessment of handwriting quality in Arabic.

The suggested study employs a CNN architecture, which is commonly utilized for image identification applications. The CNN is trained on a dataset of Arabic handwritten samples and may extract characteristics that are relevant to evaluate handwriting quality.

The proposed method also implies the full noise reduction and image binarization algorithm in addition. This preprocessing phase absolutely needs the high-quality input images and the successful accuracy of the CNN model, otherwise everything is useless.

The evaluation of the model in terms of the variety of criteria like exactness, exactitude, and recall is a part of the proposed work. These measures are intended to prove the model can well distinguish handwriting features which holds many traits in common: letter formation, spacing, alignment and in general readability.

The proposed study is significantly upgrading the science of Arabic handwriting assessment since it ensures a strong and reliable handwriting quality evaluation.

Table 1. Results of the different methods and the proposed method

Paper Title	Method	Accuracy	Dataset
"Automatic Text Scoring Using Neural Networks"	CNN + OR	0.82 F1-score	IAM
"An Empirical Analysis of BERT Embedding for Automated Essay Scoring"	CNN, LSTM, BiLSTM	83%	UJIpenchars
"A Hybrid Deep Learning Model for Arabic Text Recognition"	CNN, LSTM, BiLSTM	90% (WD), 83% (WI)	OnHW-chars
"Automated Evaluation of Handwritten Answer Script Using Deep Learning Approach"	Hybrid deep learning model	96%	IAM-OnDB
"A Survey for an Automatic Transliteration of Arabic Handwritten Script"	CNN	95.5%	MNIST
"Arabic Offline Handwritten Text Recognition with Transformers"	Transformers	96.2%	OnHW
"Arabic handwriting recognition system using convolutional neural network"	CNN, LSTM, BiLSTM	83% (WI)	OnHW

"Research on Automatic Essay Scoring of Composition Based on CNN and OR"	CNN	0.78 F1-score	PenDigits
"Auxiliary Cross-Modal Representation Learning with Triplet Loss Functions for Online Handwriting Recognition"	CNN + LSTM	96.5%	-
Proposed project	DCNNs + Bidirectional LSTMs with CTC Loss	97%	KHATT dataset

6. Conclusion

This study investigated the effectiveness of a Deep Convolutional Neural Network (DCNN) for automatically evaluating Arabic handwriting. The experiment used Kaggle's publicly accessible Arabic Handwritten Characters Dataset and incorporated parts from the approach described in "Training the model (DCNNs + Bidirectional LSTMs with CTC Loss)".

The implemented DCNN architecture produced good results while assessing Arabic handwriting characters. The system performed well (insert particular outcomes, such as high accuracy, balanced precision, and recall) on the held-out testing set. These findings indicate that DCNNs have tremendous potential for accurate and robust examination of Arabic handwriting.

The created method provides a framework automated Arabic handwriting evaluation and has potential applications in numerous sectors, such as:

Automated grading of handwritten exams to streamline assessment processes.

Evaluating handwriting improvement programmers and giving data-driven insights.

Helping create personalized tutoring systems using handwriting analysis

References

[1] Al-Jarrah OM, Abu Ata B. Automatic Assessment of Handwritten Arabic Text Quality: A Survey. Int J Adv Comput Sci Appl. 2016;7(3):232-240. Available from: https://link.springer.com/article/10.1007/s00521-020-05137-6

[2] Deng L, Yu D. Deep learning: Basics and applications. Springer Science & Business Media. 2014. Available from: https://www.amazon.com/Deep-Learning-Adaptive-Computation-Machine/dp/0262035618

[3] Shahab A, et al. A survey on Arabic script recognition with features and techniques. J King Saud Univ-Comput Sci. 2014;26(4):393-402. Available from: https://www.researchgate.net/publication/354938523_A_Survey_on_Arabic_Handwritten_Script_Recognition_Systems

[4] Ott F, Wehbi M, Hamann T, Barth J, Eskofier B, Mutschler C. The OnHW Dataset: Online Handwriting Recognition from IMU-Enhanced Ballpoint Pens

with Machine Learning. Available from: https://download.cmutschler.de/publications/2020/IMWUT2020.pdf
[5] Ott F, Rügamer D, Heublein L, Bischl B, Mutschler C. Auxiliary Cross-Modal Representation Learning with Triplet Loss Functions for Online Handwriting Recognition. Available from: https://www.researchgate.net/publication/373573138_Auxiliary_Cross-Modal_Representation_Learning_with_Triplet_Loss_Functions_for_Online_Handwriting_Recognition
[6] Chen Z, Zhou Y. Research on Automatic Essay Scoring of Composition Based on CNN and OR. Available from: https://ieeexplore.ieee.org/document/8837007
[7] Altwaijry N, Al-Turaiki I. Arabic handwriting recognition system using convolutional neural network. Available from: https://www.researchgate.net/publication/342514780_Arabic_handwriting_recognition_system_using_convolutional_neural_network
[8] Momeni S, Babaali B. Arabic Offline Handwritten Text Recognition with Transformers. Available from: https://www.researchgate.net/publication/366488913_Arabic_Offline_Handwritten_Text_Recognition_with_Transformers
[9] Shaker AS. A Survey for an Automatic Transliteration of Arabic Handwritten Script. Available from: https://iopscience.iop.org/article/10.1088/1742-6596/1530/1/012094.
[10] Rahaman MA, Mahmud H. Automated Evaluation of Handwritten Answer Script Using Deep Learning Approach. Available from: https://journals.scholarpublishing.org/index.php/TMLAI/article/view/12831
[11] Fasha M, Hammo B, Obeid N, AlWidian J. A Hybrid Deep Learning Model for Arabic Text Recognition. Available from: https://www.researchgate.net/publication/344150146_A_Hybrid_Deep_Learning_Model_for_Arabic_Text_Recognition.
[12] Beseiso M, Alzahrani S. An Empirical Analysis of BERT Embedding for Automated Essay Scoring. Available from: https://thesai.org/Publications/ViewPaper?Volume=11&Issue=10&Code=IJACSA&SerialNo=27.
[13] Alikaniotis D, Yannakoudakis H, Rei M. Automatic Text Scoring Using Neural Networks. Available from: https://arxiv.org/pdf/1606.04289.pdf
[14] Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics. 2020;21(1):6. Available from: https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-019-6413-7
[15] Powers DMW. Evaluation: From precision, recall and F-measure to ROC, informed ness, markedness & correlation. J Mach Learn Technol. 2011;2(1):37-63. Available from: https://arxiv.org/abs/2010.16061
[16] Available from: https://www.researchgate.net/figure/THE-ARABIC-ALPHABET-IN-DIFFERENTPOSITIONS_tbl1_225279602