ZAMONAVIY ILM-FAN VA INNOVATSION TADQIQOTLARDA NATIJALAR MUHOKAMASI

XXIV-RESPUBLIKA ILMIY-AMALIY KONFERENSIYA

YAKUNLARI BO'YICHA ILMIY ISHLAR TO'PLAMI

Issue - 9(2026) Available at www.uznauka.uz

DEVELOPMENT OF AN AI-BASED MODEL FOR DETECTING ERRORS IN HANDWRITTEN ARABIC LETTERS FOR LANGUAGE LEARNING

Obidinov Rahmonjon Bahromjon o‘g‘li

Master’s graduate, International Islamic Academy of Uzbekistan, Tashkent, Uzbekistan

E-mail: obidinov02@gmail.com

Abstract. *Automated evaluation of handwritten exercises is essential for scaling personalized feedback in second-language (L2) Arabic acquisition. Traditional optical character recognition (OCR) systems map handwriting directly to standard character representations while explicitly ignoring minor structural deformations. In an educational context, however, identifying these deformations is critical for diagnostic feedback. This article presents a compact, educationally grounded deep learning model designed to recognize intended Arabic letters, evaluate their orthographic correctness and classify structural writing errors in isolated handwritten characters produced by L2 learners. A formal 7-category taxonomy is introduced, comprising the correct class and six error types: shape deformations, dot placement errors, proportion imbalances, baseline misalignments, incomplete strokes and connection faults. A multi-task neural network architecture is proposed to jointly optimize character classification, binary correctness determination and multi-class error diagnosis. Furthermore, an expert-annotated dataset collection protocol and an evaluation strategy based on standard classification metrics are outlined. By bridging the gap between character transcription and pedagogical analysis, the proposed model enables real-time visual feedback within intelligent tutoring systems.*

Keywords: *Arabic handwriting recognition, educational AI, deep learning, error detection, computer vision, second-language acquisition.*

Introduction

Mastering Arabic orthography presents distinct cognitive and motor challenges for beginner learners due to its cursive structure, reliance on diacritical dots, variable letter proportions and strict alignment relative to a horizontal baseline. In self-directed or digital learning environments, students lack immediate, fine-grained correction from instructors, which leads to the reinforcement of incorrect writing habits.

Manual evaluation of handwritten exercises by teachers is time-consuming and difficult to scale [4]. While deep learning has enabled high-accuracy offline Arabic handwriting recognition (AHR) [1, 2, 3], conventional AHR models are optimized for transcription robustness and therefore actively mask minor structural errors.

To address this limitation, the central research question of this study is formulated as follows: *how can a deep learning pipeline be designed to simultaneously recognize the intended Arabic letter, evaluate its orthographic correctness and categorize specific diagnostic handwriting errors for beginner language learners?*

The aim of the study is to develop a multi-task deep learning model and an accompanying data collection and evaluation protocol for the automated diagnosis of handwriting errors in isolated Arabic letters written by L2 learners.

Materials and Methods

Processing pipeline. The proposed system processes camera-captured images of handwritten Arabic letters written in physical exercise notebooks through a multi-stage computer vision pipeline (Figure 2). Input images undergo contrast-limited adaptive histogram equalization (CLAHE) to mitigate uneven ambient illumination, followed by adaptive binarization to separate ink strokes from the background paper grid. Bounding boxes are segmented around the primary letter, padded to preserve the aspect ratio and proportional integrity, and resized to a uniform 128 × 128 × 1 tensor.

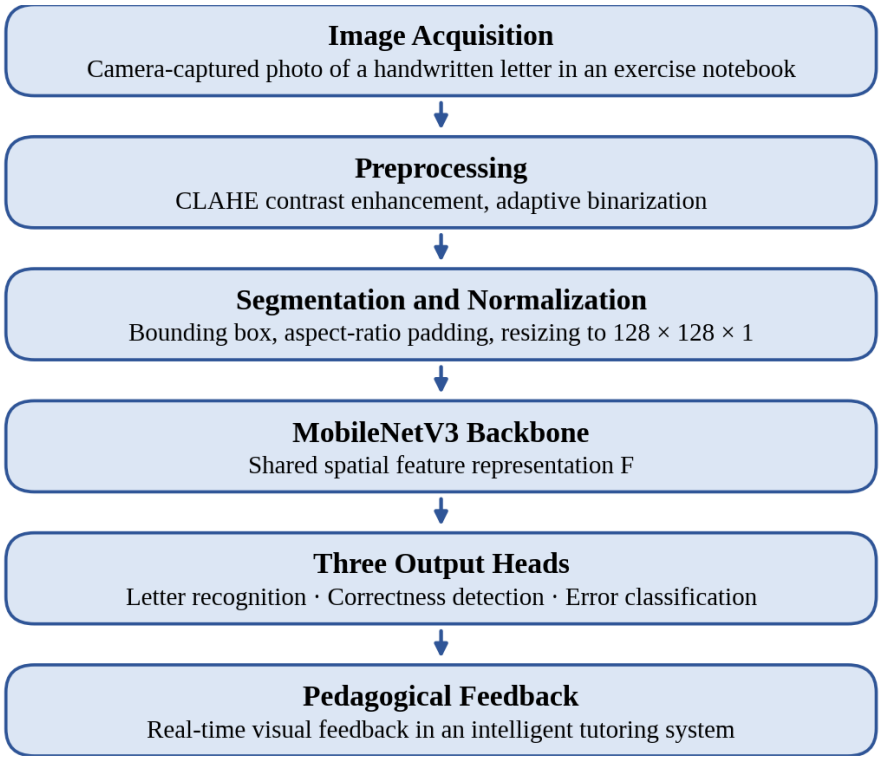

Figure 2. Multi-stage pipeline of the proposed handwriting error detection system

Model architecture. A parameter-efficient MobileNetV3 backbone is selected for feature extraction due to its high accuracy-to-latency ratio, which

makes it well suited for integration into mobile educational applications on edge devices. The shared backbone extracts spatial feature representations F, which feed into three task-specific output heads:

1. Intended character recognition head ($\hat{y}$): categorical prediction across the 28 primary Arabic letters using a Softmax activation.
2. Correctness detection head ($\hat{c}$): binary prediction (1 = correct, 0 = incorrect) using a Sigmoid activation.
3. Error classification head ($\hat{e}$): multi-class prediction across the pedagogical error types using a Softmax activation.

Loss function. The framework is optimized end to end using a joint multi-task loss function:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{rec}(\hat{y}, y) + \lambda_2 \mathcal{L}_{corr}(\hat{c}, c) + \lambda_3 (1 - c)\, \mathcal{L}_{err}(\hat{e}, e) \qquad (1)$$

where $\mathcal{L}_{rec}$ is the categorical cross-entropy for character identification:

$$\mathcal{L}_{rec} = -\sum_{k=1}^{28} y_k \log(\hat{y}_k) \qquad (2)$$

$\mathcal{L}_{corr}$ is the binary cross-entropy for correctness:

$$\mathcal{L}_{corr} = -[c\log(\hat{c}) + (1 - c)\log(1 - \hat{c})] \qquad (3)$$

and $\mathcal{L}_{err}$ is the categorical cross-entropy for the error class. The mask $(1 - c)$ ensures that correct handwriting samples ($c = 1$) do not generate loss gradients for the error classification head. The task weights $\lambda_1, \lambda_2, \lambda_3$ balance the gradient magnitudes during training.

Dataset design and error taxonomy. Existing public Arabic handwriting datasets, such as AHCD [5] and IFN/ENIT, consist almost entirely of fluent native writing and therefore do not represent the error distributions typical of beginner L2 learners. A structured dataset collection protocol targeting non-native Arabic learners is therefore proposed. Each sample is stored with a tuple metadata schema:

$$\mathcal{D} = \{(x_i, y_i, c_i, e_i)\}_{i=1}^{N} \qquad (4)$$

where x_i is the preprocessed image, $y_i \in \{1, \dots, 28\}$ is the intended letter, $c_i \in \{0,1\}$ is the correctness ground truth, and e_i is the designated error class. The error taxonomy used for annotation is given in Table 1.

Table 1. Pedagogical taxonomy of handwriting errors in isolated Arabic letters

Code	Category	Description
E0	Correct	The letter is written in accordance with orthographic norms
E1	Shape deformation	The overall form of the letter is distorted
E2	Dot placement error	Diacritical dots are missing, extra, misplaced or miscounted
E3	Proportion imbalance	The relative size of letter components is violated
E4	Baseline misalignment	The letter is incorrectly positioned relative to the baseline
E5	Incomplete stroke	Part of the letter stroke is missing

Code	Category	Description
E6	Connection fault	The connecting stroke is incorrect or missing

To maximize labeling reliability, all samples are evaluated by expert Arabic language educators, and inter-annotator disagreements are resolved by majority voting.

Evaluation protocol. Since this study introduces a proposed model architecture and experimental protocol, performance is to be evaluated empirically once the dataset has been collected. The evaluation framework comprises four components:

1. Letter recognition: overall Top-1 classification accuracy on unseen test participants.
2. Correctness detection: precision, recall and F1-score for binary correctness determination.
3. Error classification: macro-averaged F1-score across all error types ($E_1 \dots E_6$):

$$\text{Macro } F_1 = \frac{1}{M}\sum_{m=1}^{M}\frac{2 \cdot P_m \cdot R_m}{P_m + R_m} \qquad (5)$$

where M is the number of error classes, and P_m and R_m are the precision and recall for class m.

4. Error analysis: confusion matrices to evaluate cross-class confusion between closely related error types (for example, shape deformation and proportion imbalance).

Results and Discussion

The main result of this study is a methodological framework that reformulates Arabic handwriting recognition from a transcription task into a pedagogical diagnosis task. Unlike conventional OCR and AHR systems, which are trained to be invariant to small deformations [1, 3], the proposed model treats these deformations as the target signal. The multi-task design allows a single network to answer three questions that are relevant to the learner: which letter was intended, whether it was written correctly and, if not, what kind of error was made.

The masked loss term in Equation (1) is an important design decision: it prevents the error classification head from being trained on correct samples, so that the network learns error features only from genuinely erroneous handwriting. The choice of the lightweight MobileNetV3 backbone makes it possible to run the model directly on a smartphone, which is essential for providing feedback in real time while the learner works in a paper notebook.

The main limitation of the study is that the model has not yet been evaluated empirically, since a dataset of beginner L2 handwriting is still to be collected. In addition, the approach currently covers only isolated letters, whereas real writing involves connected letter forms whose analysis is considerably more complex [6].

Conclusion

This article presented an educationally oriented deep learning framework for detecting and diagnosing handwriting errors in isolated Arabic letters. By moving beyond conventional transcription OCR and introducing a 7-category pedagogical error taxonomy, an expert annotation schema and a multi-task MobileNetV3

architecture, the proposed approach addresses structural defects specific to beginner L2 handwriting. The system enables automated, actionable feedback for learners, bridging computer vision research and intelligent tutoring platforms.

Future work involves executing the empirical data collection protocol across diverse L2 student cohorts, refining the multi-task loss balance and extending the pipeline to continuous connected word handwriting.

References

1. Al-Badr, B., & Suen, C. Y. (2021). Survey of Arabic handwriting recognition strategies and applications. *Pattern Recognition, 112*, 107730.
2. Alrobai, A., Alnuaim, A., & Zakariah, M. (2021). Deep learning-based framework for offline handwritten Arabic character recognition. *IEEE Access, 9*, 122105–122120.
3. Al-Sarem, M., Ahsan, M. M., & Zeshan, F. (2019). Feature extraction methods for handwritten Arabic character recognition: A comparative study. *Journal of Computer Science, 15*(3), 341–352.
4. Djioua, M., Cheriet, M., & Suen, C. Y. (2021). Automated evaluation of handwriting quality: A pedagogical perspective. *Computers & Education, 168*, 104190.
5. El-Sawy, A., Hazem, M., & Loey, M. (2017). Arabic handwritten character dataset using convolutional neural network. *WSEAS Transactions on Computer Research, 5*, 11–19.
6. Zhang, X. Y., Yin, F., Zhang, Y. M., Liu, C. L., & Bengio, Y. (2020). Drawing and writing with deep generative models: A survey on handwriting

Issue - 9(2026) Available at www.uznauka.uz

analysis and synthesis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(11), 2753–2774.