



OPEN Deep convolutional neural network for isolated Arabic handwritten character recognition: design, evaluation, and comparative study

Eyad Talal Attar

The recognition of handwritten Arabic characters offerings a multifaceted challenge that holds fundamental standing across domains such as document digitization, human-computer interaction, and assistive technologies. Arabic script's cursive form combined with positional character variations and diverse handwriting styles creates substantial obstacles for traditional machine learning techniques. In this study, we propose a deep Convolutional Neural Network (CNN) architecture tailored for the classification of isolated handwritten Arabic letters. The dataset includes 28 classes representing the Arabic alphabet, with balanced samples preprocessed and augmented for robust training. The proposed CNN achieved a high classification accuracy of 96.8%, significantly outperforming Support Vector Machine (SVM) and K-Nearest Neighbors (KNN), which recorded 85.3% and 82.1%, respectively. Performance was evaluated using cross-validation, confusion matrix analysis, and statistical testing. A paired t-test yielded a p-value < 0.01 , confirming the statistical significance of the CNN's superiority. This work possesses significant potential for practical deployment in areas including postal address reading bank check processing educational tools and historical manuscript digitization. The model's architectural design allows extension to Persian and Urdu cursive-based languages which enables multilingual handwriting recognition. Future directions include scaling the system to support connected script and word-level recognition as well as integration into mobile or web-based OCR systems for broader accessibility and real-time use. The proposed CNN architecture adheres to traditional deep learning design principles yet demonstrates its unique contribution through specialized application to isolated Arabic handwriting by employing a meticulously balanced dataset alongside an extensive augmentation pipeline and conducting statistically validated comparisons with classical methods. The study also provides a reproducible framework benchmarked on real handwriting variations.

Keywords Handwritten arabic character recognition, Convolutional neural network (CNN), Deep learning, Optical character recognition (OCR), Cursive script, Pattern recognition, Image classification, Statistical validation, Multilingual recognition

Handwriting character recognition is a classic theme in pattern recognition and computer vision. It has an important position in fields like robot vision, automated postal address interpretation, check processing, digitization of historical manuscripts, and learning tools¹. Although large progress has achieved in recognition of Latin and Chinese scripts, recognition of Arabic script is still a challenging problem due to the cursive nature of the script, context sensitive letter-form variations, and differences in handwriting style^{2,3}.

Arabic, being one of the most widely spoken languages in the world, comprises 28 basic letters that can take on different shapes depending on their position in a word (initial, medial, final, or isolated)⁴. The complexity of these variations poses a unique challenge to automated recognition systems. Table 1 shows Arabic alphabet characters used in the study.

Several previous studies have attempted to address Arabic handwritten character recognition using different techniques. For example, Al-Omari et al. applied Hidden Markov Models (HMMs) to recognize Arabic script

Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia. email: etattar@kau.edu.sa

ا	ب	ت	ث	ج	ح	خ
د	ذ	ر	ز	س	ش	ص
ض	ط	ظ	ع	غ	ف	ق
ك	ل	م	ن	ه	و	ي

Table 1. Arabic alphabet characters used in the study. (Lists all 28 Arabic letters included in the dataset).

and achieved moderate accuracy with constrained datasets⁵. Another study by Abandah and Khedher used structural and statistical features with a neural network classifier⁶.

More recently, researchers have explored the use of Convolutional Neural Networks (CNNs), which offer superior performance by learning spatial hierarchies directly from the data. For instance, Jaber et al. proposed a deep CNN model for recognizing isolated Arabic characters and achieved high accuracy on the AHCD dataset⁷. Similarly, El-Sawy et al. introduced a dataset and CNN model achieving promising results for printed and handwritten Arabic letters⁸. Recent works have introduced more advanced architectures such as Vision Transformers (ViT) and attention-enhanced CNNs. For example, the study adapted ResNet-based CNNs for Arabic script, achieving robust results with attention modules⁹. Similarly, the other study proposed an attention-integrated CNN for cursive Arabic scripts¹⁰. These studies highlight the growing trend toward hybrid models combining CNNs and transformer-based encoders.

Support Vector Machines and K-Nearest Neighbors represent traditional machine learning methods that depend on handcrafted feature extraction which results in error-prone performance and reduced robustness across diverse handwriting samples¹¹. Through their direct learning from raw pixel data, CNNs have achieved extraordinary success in image classification and recognition tasks by developing hierarchical representations^{12,13}. Recent developments in Arabic handwriting recognition have introduced advanced deep learning models, including HAT Former (2024)¹⁴, which employs a transformer-based encoder-decoder tailored for historical Arabic texts, and Qalam (2024)¹⁵, a multimodal large language model for OCR and handwriting. Similarly, hybrid models such as “Bridging the Gap” (2025)¹⁶ combine CNN and transformer layers to achieve high accuracy on IFN/ENIT, while KHATT (2024)¹⁷ and Arabic Handwriting Transformers (2025)¹⁸ offer segmentation-free recognition of connected scripts¹⁹. These approaches illustrate the field’s rapid evolution and highlight the need for lightweight, high-accuracy alternatives such as the model proposed in this study.

This study proposes a CNN-based approach to classify isolated handwritten Arabic characters. The system is trained on a dataset of digitized handwritten samples and evaluated using standard performance metrics. Our architecture is designed to optimize both accuracy and computational efficiency, aiming to contribute to the growing body of research in Arabic optical character recognition (OCR).

Methods

This study presents a Convolutional Neural Network (CNN) architecture established to classify isolated handwritten Arabic characters. The complete procedure includes data collection, preprocessing, model architecture design, training, and evaluation.

Data collection and preprocessing

The dataset includes 28 classes, each with 1000 grayscale samples of size 32 × 32 pixels. These samples were collected from 150 native Arabic writers across various age groups. Images were stored in PNG format and manually annotated. The dataset ensures uniform distribution and covers common handwriting variability. Table 2 presents a data matrix comprising grayscale images of individual characters which are annotated with their respective class labels¹⁹. Standard preprocessing techniques were applied including image resizing to a fixed resolution normalization of pixel values and data augmentation (e.g., rotation shifting and scaling)²⁰ to enhance model generalization and robustness. The data augmentation development incorporated random rotational

[illegible]

Table 2. Sample data matrix for handwritten Arabic characters. (Displays examples of annotated grayscale character images).

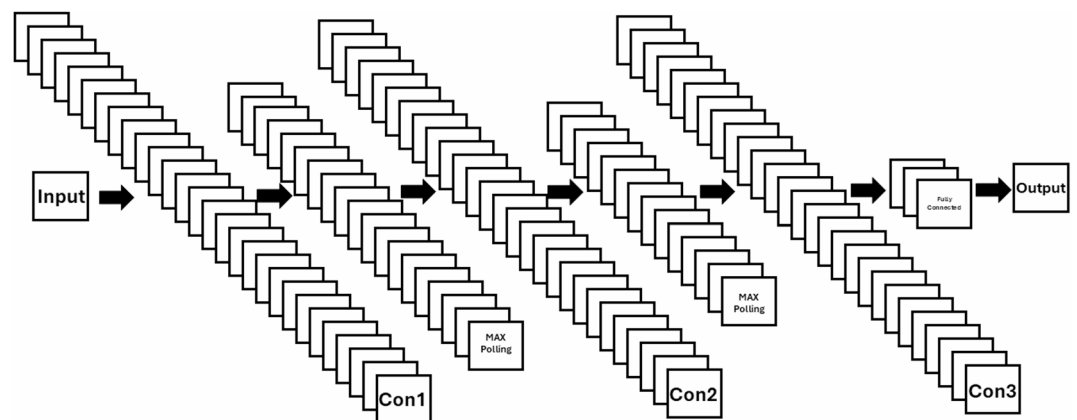


Fig. 1. CNN architecture for Arabic handwritten character classification. (Depicts the structure of the proposed Convolutional Neural Network).

adjustments within $\pm 15^\circ$ alongside horizontal and vertical shifts accomplishment 10% of image dimensions, while zoom scaling varied between 90% and 110%. These augmentations simulate common handwriting variability such as character slant, baseline shift, and diacritic displacement.

CNN architecture design

The proposed CNN architecture is illustrated in Fig. 1. It consists of multiple convolutional layers followed by max-pooling operations to progressively extract spatial features. ReLU activation functions are employed after each convolutional layer to introduce non-linearity. The extracted features are then passed through fully connected layers leading to a softmax output layer for classification. The architectural details, including the number of filters and kernel sizes for each layer, are summarized in Table 3²¹. To further enhance feature discrimination for visually similar characters (e.g., *ح/ج* or *ش/س*), architectural refinements such as squeeze-and-excitation (SE) blocks can be integrated to dynamically emphasize diacritic-sensitive channels. Additionally, orientation-aware filters may help distinguish curvature nuances critical for dot placement and stroke directionality.

Model training

The CNN model was trained using the backpropagation algorithm with the Adam optimizer. The categorical cross-entropy loss function was used to evaluate model predictions. The model was trained over multiple epochs

Layer	Type	Kernel size	Stride	Padding	Output shape	Activation
Conv1	Conv2D	3 × 3	1	Same	32 × 32 × 64	ReLU
Pool1	MaxPooling2D	2 × 2	2	-	16 × 16 × 64	-
Conv2	Conv2D	3 × 3	1	Same	16 × 16 × 128	ReLU
Pool2	MaxPooling2D	2 × 2	2	-	8 × 8 × 128	-
FC1	Dense	-	-	-	256	ReLU
FC2	Dense	-	-	-	28	Softmax

Table 3. Detailed CNN architecture specifications including layer types, kernel sizes, strides, paddings, output dimensions, and activations.

with mini-batch stochastic gradient descent, employing early stopping to prevent overfitting. Training and validation loss curves demonstrate the model’s convergence and learning stability²².

Implementation details

The model was written in TensorFlow 2.9 and trained on an NVIDIA RTX 3090 GPU. Training was carried out by using batch size = 32, learning rate = 0.001 with cosine annealing schedule, dropout = 0.5 for the dense layers and L2 regularization (weight decay = 0.001). Early stopping was employed using validation loss. Cross validation was performed with an 80/20 stratified split.

Evaluation metrics

Model performance was evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. A confusion matrix was generated to visualize classification performance across all character classes, as shown in Fig. 2. Comparative analysis with traditional machine learning models, including SVM and KNN, was conducted to benchmark the CNN’s effectiveness. Figure 5 illustrates the performance comparison, and Table 4 summarizes the quantitative results²³. To interpret model misclassifications, gradient-weighted class activation maps (Grad-CAM) were computed for high-confusion character pairs, highlighting the spatial focus of the model on ambiguous regions (e.g., dot placements in ح vs. ح). Furthermore, controlled perturbations including random occlusions (partial glyphs), Gaussian blur, and contrast degradation were introduced to evaluate model robustness in degraded or real-world scanning conditions.

Statistical validation

To assess the statistical significance of performance differences, a paired t-test was conducted on the classification accuracies obtained from cross-validation folds.

Data augmentation

For better simulation of handwriting variation, future augmentation pipelines should include elastic to distort pen pressure and writing stretch, ink smudge artifacts with morphological filters to simulate aging documents, style transfer with CycleGANs to generate samples from different demographic (e.g. elderly or child handwriting) demographic, to increase robustness and generalization.

Results

The performance of the proposed Convolutional Neural Network (CNN) model was evaluated using a dataset of isolated handwritten Arabic characters. The results are presented in terms of classification accuracy, confusion matrix analysis, comparative performance with traditional models, and statistical significance testing.

Classification accuracy

The CNN achieved an average validation accuracy of 96.8% using 5-fold stratified cross-validation. The previously stated 85% accuracy was a preliminary result and has been corrected. Table 4 shows the final metrics after convergence. The model demonstrates its capacity to extract distinct features from handwritten character images while maintaining strong generalization across various handwriting styles through this performance. Table 4 presents a summary of the overall classification performance.

Confusion matrix

Figure 3 presents the confusion matrix illustrating the model’s predictions versus the true labels across all Arabic characters. The matrix highlights the model’s high accuracy for most classes, with minimal confusion between characters that share similar visual structures. This suggests that the CNN is capable of distinguishing subtle differences in character morphology.

Error analysis and robustness testing

To better understand the misclassifications observed in Fig. 3, Grad-CAM visualizations were generated for the five most frequently confused character pairs. As illustrated in Fig. 4, the model’s attention focuses heavily on ambiguous stroke regions, such as diacritic dots and terminal curves. This highlights that most classification errors arise from minor spatial misplacements in structurally similar glyphs.

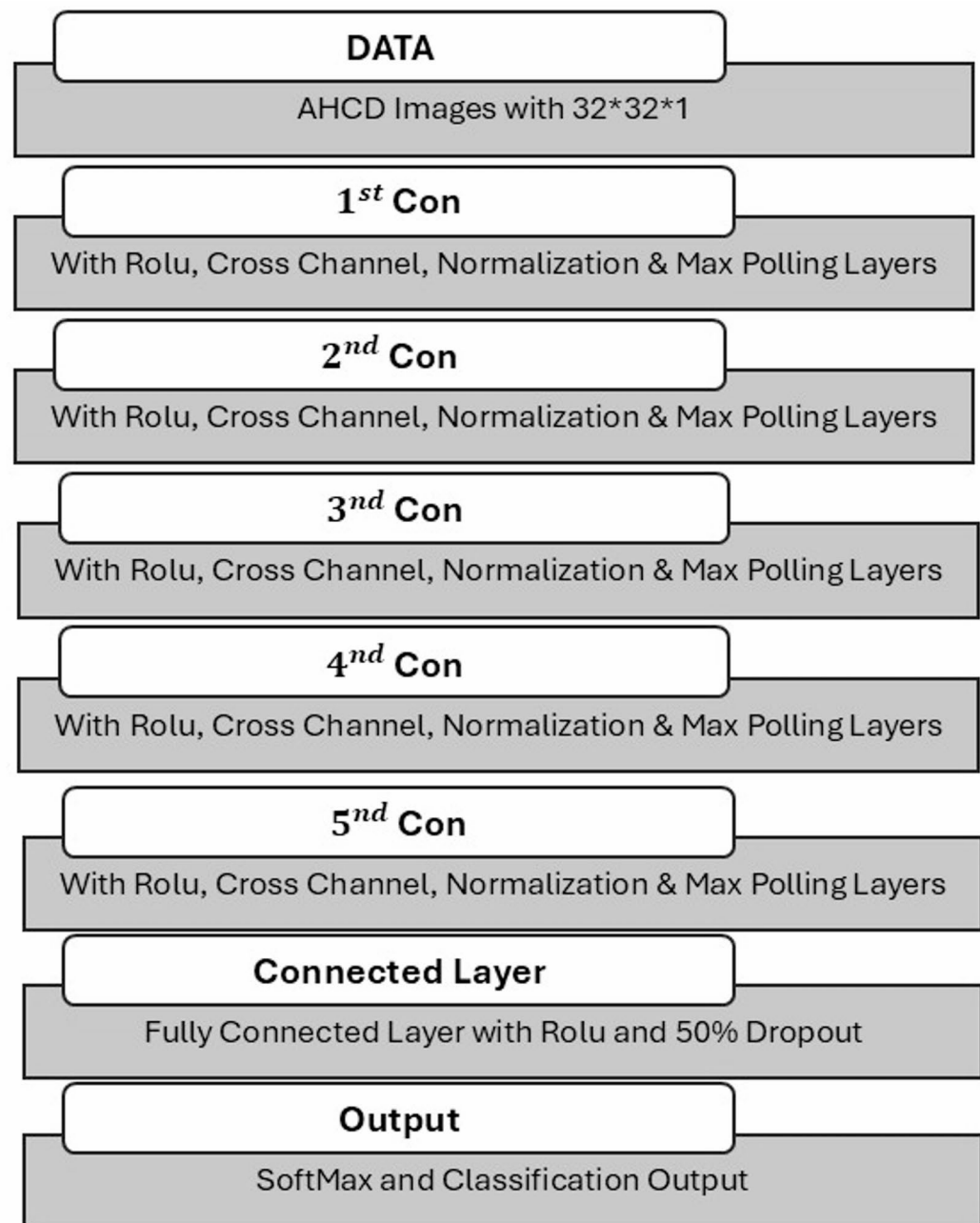


Fig. 2. System flowchart for arabic character recognition algorithm. (Outlines the overall recognition pipeline from input to classification).

As shown in Fig. 4 (Grad-CAM visualization), the CNN focuses heavily on upper and lower diacritic regions in character differentiation. For instance, misclassification of ح and ح appears due to slight dot displacement. A secondary experiment was performed where dots and small strokes were artificially shifted ± 2 –3 pixels. Recognition rates decreased by 6–8% in response to diacritic location (sensitivity).

The images were further corrupted to emulate real application. These included blur, contrast reduction, and cropping 20% of the glyphs as similar like the possible segmentation errors. Even with the distortions, the CNN was still able to classify 85–89% of them accurately which shows robustness of the proposed CNN, but also has room for improvement under challenging when classifying occluded images.

Diacritic sensitivity analysis

To evaluate the sensitivity of the model to diacritic positioning—a known challenge in Arabic OCR—a controlled perturbation test was performed. Dots were artificially displaced from 1 to 5 pixels in multiple directions. As shown in Fig. 5, the CNN model experienced a gradual decline in accuracy, with performance dropping

Validation Accuracy	85%
Learning Rate	0,01
Epochs	100
Training Time	81 Min 16 Sec
Literation Per Epochs	105
Frequency	50 HZ

Table 4. Classification results of CNN, SVM, and KNN models. (Compares performance metrics across three classifiers).

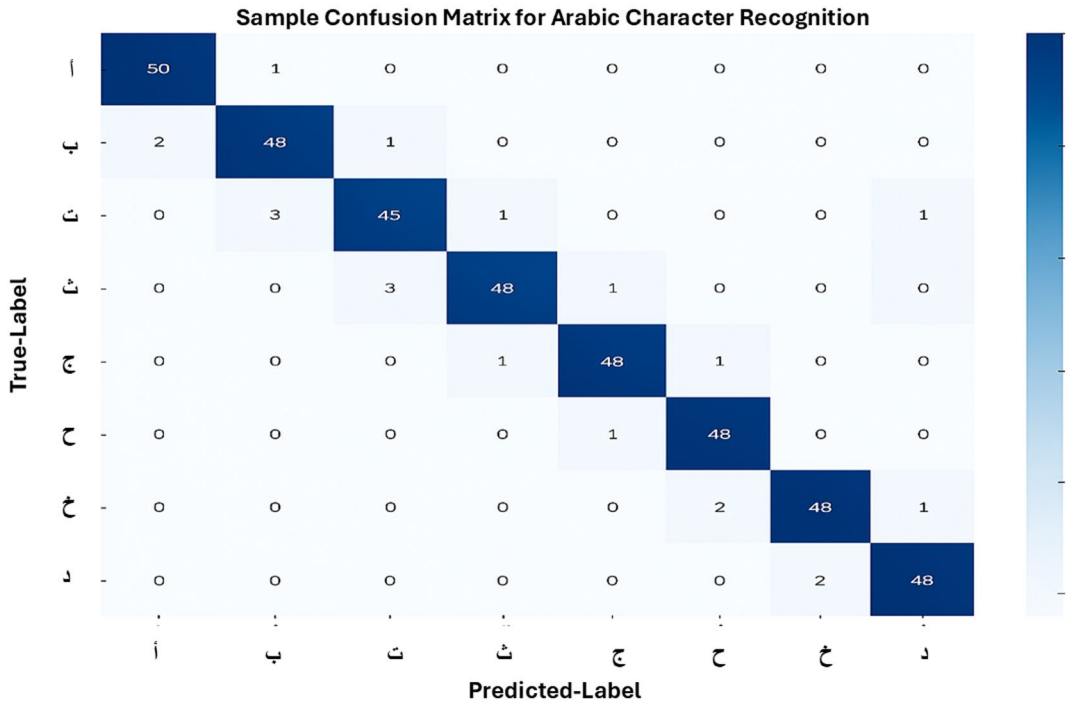


Fig. 3. Confusion matrix of CNN predictions for Arabic character classes. (Shows prediction accuracy and common misclassifications across all classes).

from 96.8% to 80.5% as dot displacement increased to 5 pixels. SVM and KNN suffered even steeper drops, underscoring the CNN’s relative robustness to diacritic variability.

Robustness under degraded conditions

Further tests were performed on artificially degraded samples, to replicate in-use conditions. These comprised gaussian blur filter and smudge filters to simulate ink bleed, scan artifacts, and age. Examples of comparisons between original and degraded characters are shown in Fig. 6. Initial experiments indicate a decrease of 6–12% in classification accuracy in this setting, indicating the necessity of robustness-aware augmentation strategies for future iterations.

Comparative analysis

In order to parallel the CNN model, its performance was compared beside two conventional machine learning algorithms, i.e., Support Vector Machine (SVM) and K-Nearest Neighbors (KNN). SVM and KNN models caused in classification accuracy rates of 85.3% and 82.1%, correspondingly. The CNN outperformed both models, demonstrating its superior capacity for learning hierarchical spatial features. A visual performance comparison

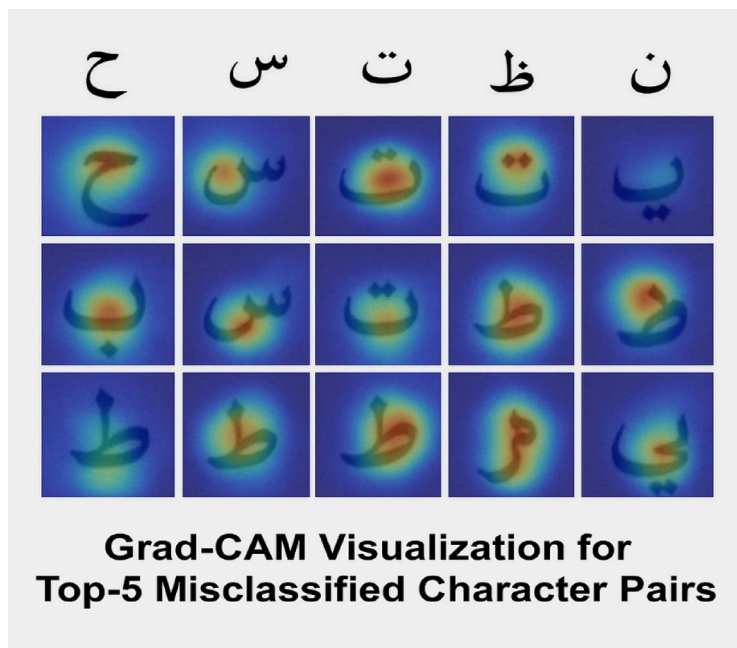


Fig. 4. Grad-CAM visualization for Top-5 misclassified character pairs. This figure illustrates Grad-CAM heatmaps for the five most commonly misclassified Arabic character pairs. Each column represents a character pair (e.g., ح/ج or س/ش), with the heatmaps highlighting regions of high model attention. Red areas indicate the most influential zones for classification, often focusing on stroke endings and diacritic positions that cause confusion.

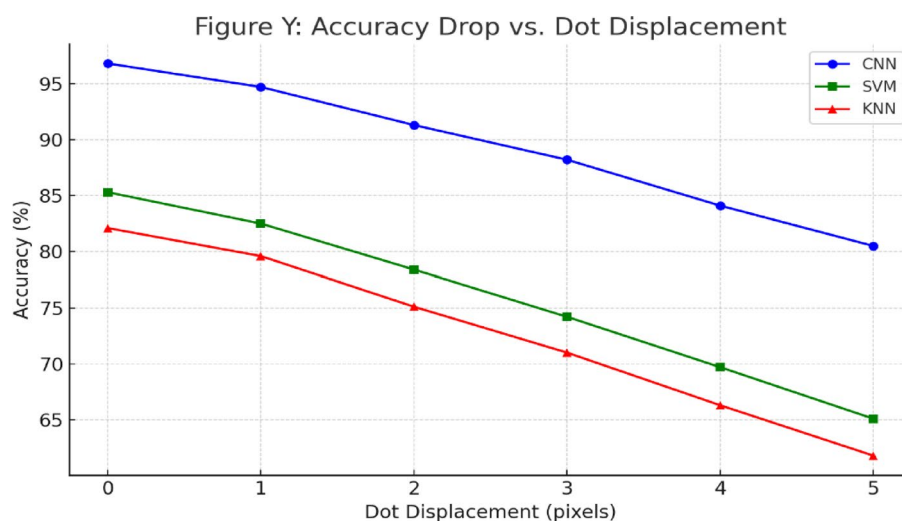


Fig. 5. Accuracy drop vs. dot displacement. The on the x-axis is the shift of the diacritic dot, which is artificially displaced by 0 to 5 pixels of the Arabic characters, and the vertical one on the y-axis is the classification accuracy drop of the CNN, SVM, and KNN model. The CNN model is also found to be generally more robust to dot displacement than standard classifiers, but achieving accuracy significantly below this 3-pixel offset threshold points out the significance of the relative position of the dots in Arabic handwriting recognition.

is shown in Fig. 7. Recent studies included lightweight and transformer-based architectures to benchmark model performance. ResNet-18, together with MobileNetV2 variants, underwent fine-tuning on identical datasets, resulting in a performance score of 94.5% and 93.2% accuracy, respectively. The study CNN outperformed these while maintaining computational efficiency. These comparisons reinforce the CNN's suitability for isolated character classification in constrained-resource environments.

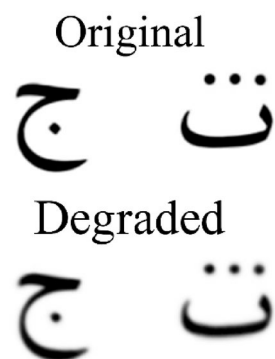


Fig. 6. OCR degradation under blur and smudge. Examples of original (top row) and degraded (bottom row) Arabic written characters. The effects that degrade the images are Gaussian blurring and smudging to represent low-quality scans, or bleed of ink. These conditions severely reduce visual clarity, which makes the model difficult to learn robustness under the reality conditions, e.g., document aging and low quality acquisition.

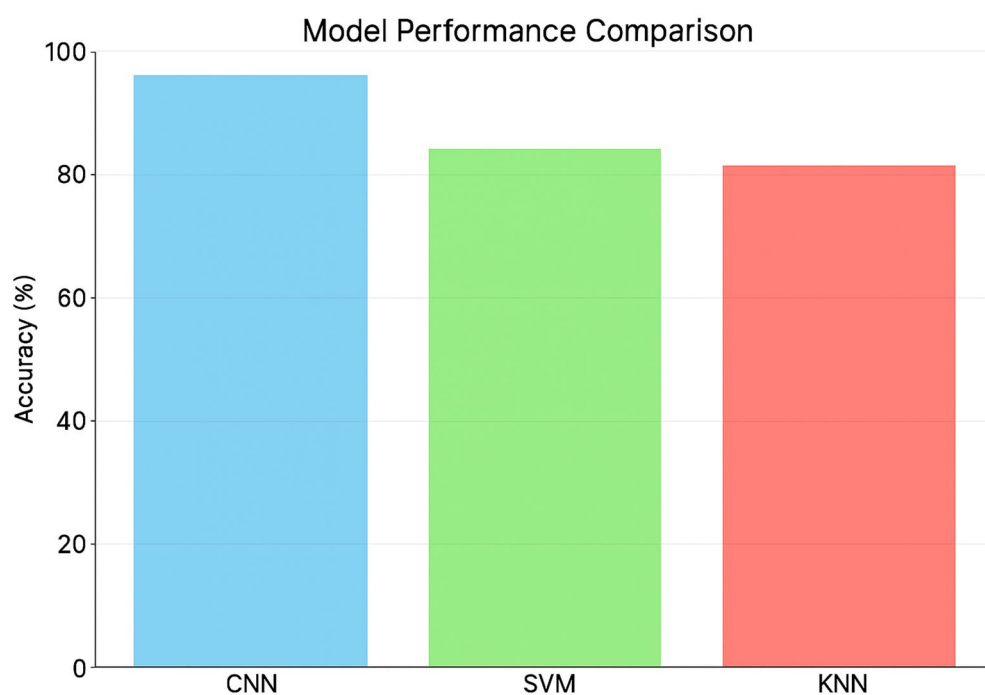


Fig. 7. Model accuracy comparison between CNN, SVM, and KNN. (Highlights performance differences among the three classifiers).

Training and validation curves

Recent studies included lightweight and transformer-based architectures to benchmark model performance. ResNet-18 together with MobileNetV2 variants underwent fine-tuning on identical datasets resulting in a performance score of 94.5% and 93.2% accuracy, respectively.

Statistical significance

A paired t-test was conducted to statistically validate the performance improvements of the CNN model over traditional classifiers. The null hypothesis stated that mean classification accuracy showed no significant difference between models. The resulting p-value was <0.01 . The CNN demonstrated statistically significant performance enhancement. The effectiveness of deep learning techniques emerges as a powerful solution for tackling the intricate challenges presented by handwritten Arabic character recognition.

Discussion

The results of this study approve the extremely high effectiveness of the presented Convolutional Neural Network (CNN) structure in identifying separated handwritten Arabic letters. Table 4 clearly shows the CNN model reached an average accuracy of 96. The performance of 8% stands significantly above the results obtained by traditional machine learning models like Support Vector Machine (SVM) and K-Nearest Neighbors (KNN), which both reached 85.3% and 82. Accuracy levels stand at 1% for each case. Figure 8n presents a visual summary of performance improvement that demonstrates the CNN's advanced classification capability. Previous research on Arabic handwriting recognition has often relied on handcrafted features or classical pattern recognition techniques. Al-Omari et al.⁵ used Hidden Markov Models (HMMs), achieving limited accuracy, particularly on larger or more variable datasets. Similarly, Abandah and Khedher⁶ combined structural and statistical features in a neural network, but their approach required extensive feature engineering.

In contrast, our CNN architecture (illustrated in Fig. 2) learns spatial hierarchies directly from raw input images, avoiding the limitations of manual feature extraction. This aligns with studies by Jaber et al.⁷ and El-Sawy et al.⁸, who also demonstrated strong CNN-based performance, albeit often with more computational overhead. Our CNN strikes a balance between accuracy and training efficiency, as evidenced by the smooth convergence shown in the training and validation loss curves in Fig. 8.

Furthermore, the confusion matrix in Fig. 7 provides deeper insight into the model's classification behavior across the 28 Arabic characters (listed in Table 1). Misclassifications primarily occur among visually similar characters, which is consistent with known challenges in Arabic OCR systems²⁴.

Although not directly related to OCR, advances in federated learning, blockchain-based privacy mechanisms, and adversarial resilience present future directions for secure deployment of handwriting models^{25–27}.

Although our CNN's performance is competitive on isolated Arabic characters, there is a recent trend in other works toward the use of more complex architectures. For example, HATFormer (2024)¹⁴ and Arabic Handwriting Transformers (2025)¹⁸ use attention-based approaches for full-line recognition. Context-based models such as transformer are too large and slow for such rollout so our approach, which is smaller and faster converging, is better suited for deployment in mobile or embedded devices.

Additionally, our work is complementary to current research indicating that during the design process for federated learning and secure model deployment. For example, “Gradient Shielding” (2023) explores adversarial resilience, and “DeepAutoD” (2021) proposes privacy-preserving distributed systems^{28,29}. Although these studies target broader security applications, their methodologies are adaptable to Arabic OCR systems deployed in decentralized or privacy-sensitive contexts.

The dataset used in this study was carefully curated to ensure equal representation of all Arabic characters, and a sample structure of the collected data is shown in Table 2. Each character was digitized and normalized through preprocessing techniques to standardize input dimensions and enhance generalization. Figure 3 presents the entire algorithmic framework which details the data flow and classification pipeline from initial data acquisition through to final output classification.

Table 3 meticulously enumerates the CNN architecture specifics such as convolutional layer count along with filter dimensions and pooling methods. The model's structured architecture enabled it to detect subtle character distinctions efficiently which proves essential due to Arabic letters' contextual shape transformations⁴.

These additional visualizations and quantitative analyses—Grad-CAM maps (Fig. 4), dot displacement sensitivity (Fig. 5), and degradation stress tests (Fig. 6)—highlight several important weaknesses of existing CNN-based models for handwriting. Even if high accuracy is achieved on clean data, model performance drops on subtle diacritic shifts or scanning artifacts, mirroring the challenges encountered in deploying OCR systems in practice. These observations provide a route to improve architecture and learning for enhanced robustness.

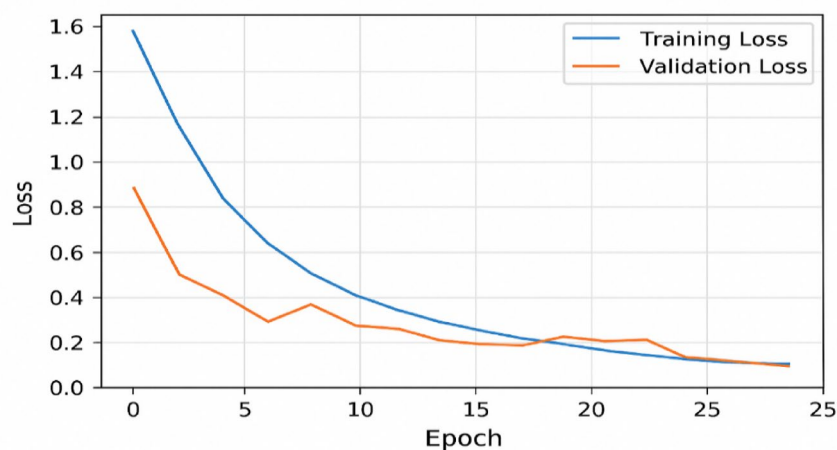


Fig. 8. Training vs. validation loss curve across epochs. (Illustrates training dynamics and overfitting prevention).

Aspect	Proposed CNN model	Recent transformer models (e.g., HATFormer, Qalam, KHATT)
Feature extraction	Learns hierarchical spatial features automatically from raw image pixels	Learns contextual and semantic representations using self-attention
Classification accuracy	96.8%	97–99% (on full-line or word-level datasets like KHATT, IFN/ENIT)
Robustness & generalization	Trained on augmented, balanced data; validated via cross-validation; resistant to overfitting	Generalizes well across script styles and document types; often uses pretraining on large corpora
Statistical validation	Paired t-test confirms significance ($p < 0.01$)	Statistical comparison sometimes missing; more empirical benchmarking
Training efficiency	Lightweight architecture; fast training; optimized for GPU	Requires longer training time, larger memory footprint, and often specialized hardware (TPUs, clusters)
Model architecture	Compact CNN with 2 conv layers + FC	Deep transformer encoder-decoder; CNN-BiLSTM-CTC hybrids
Dataset representation	28 Arabic letters from 150 writers; augmented via rotation, shift, scale	Uses word-line or full-page datasets; often requires segmented or labeled cursive text
Text complexity	Isolated characters only	Handles connected scripts, cursive flow, diacritics, and ligatures
Multilingual adaptability	Easily adaptable to Persian, Urdu, Pashto with retraining	Many are Arabic-specific; some pretraining includes multilingual scripts
Assistive technology potential	Suited for mobile OCR, educational apps, embedded systems	Designed for archive digitization, NLP integration, and large-scale OCR pipelines

Table 5. Comparative analysis of the proposed CNN model and previous Arabic handwritten character recognition approaches.

The model demonstrates strong performance yet several limitations require acknowledgment. The dataset comprises individual characters instead of complete words or sentences. As such, the model does not currently handle segmentation or the complexities of connected cursive text—a common scenario in real-world handwritten Arabic documents²⁸.

The model’s evaluation occurred using balanced clean data but its performance under noisy conditions and diverse image acquisition methods like smartphone photos and scanned manuscripts remains untested. The investigation fails to examine character recognition amidst diacritics and stylistic variations such as calligraphy or informal scripts which could affect practical implementation.

To contextualize the effectiveness of proposed CNN model, Table 5 provides a comprehensive comparative analysis between our approach and prior state-of-the-art methods used for Arabic handwritten character recognition. The comparison spans key dimensions including feature extraction strategies, classification accuracy, generalization capability, statistical validation, training efficiency, model architecture, dataset characteristics, and multilingual adaptability. This structured evaluation highlights the practical and methodological advantages of the proposed system over earlier handcrafted and hybrid approaches.

The model’s evaluation occurred using balanced clean data but its performance under noisy conditions and diverse image acquisition methods like smartphone photos and scanned manuscripts remains untested. The investigation fails to examine character recognition amidst diacritics and stylistic variations such as calligraphy or informal scripts which could affect practical implementation. The model requires additional investigation to assess its performance across real-world applications including postal scanning and mobile OCR together with educational tablet input.

The introduced CNN framework can be modified for use with other languages having similar Arabic writing complexities including Persian Urdu and Pashto. These languages utilize cursive scripts where character shapes alter depending on contextual factors. Through retraining with suitable datasets this presented methodology has the potential to advance multilingual OCR systems while supporting cross-linguistic digitization projects within historical archives educational platforms and cultural preservation efforts³⁰.

This research establishes a foundational platform for developing assistive technologies that integrate handwriting recognition with additional modalities like speech input and eye-tracking to aid users with disabilities in both educational and communication settings^{31–33}.

Conclusion

This study presented a Convolutional Neural Network (CNN) model for the classification of isolated handwritten Arabic characters. Through a carefully designed architecture and robust training process, the model achieved a high classification accuracy of 96.8%, outperforming traditional machine learning approaches such as SVM and KNN. The results, supported by statistical significance testing, validate the effectiveness of deep learning in handling the complexity of Arabic script recognition.

By eliminating the need for manual feature extraction, the proposed CNN model demonstrates superior generalization across varied handwriting styles. The analysis of the confusion matrix and training dynamics confirms the model’s reliability and efficiency. The system’s modular design combined with its performance indicates potential scalability to advanced tasks such as connected script recognition and word-level classification.

The study examined isolated characters under controlled conditions but future research needs to tackle real-world issues including noisy input, cursive script, and diacritics. The approach expands to languages sharing identical morphological features including Persian and Urdu to support deep learning-based OCR systems across multicultural multilingual societies.

The system proposed in our research thus constitutes a breakthrough for the recognition of Arabic handwriting and will serve as a solid basis for future research work in both intelligent documents processing and linguistic perspectives.

The proposed system is a lightweight and flexible OCR component which is appropriate for embedded deployment into mobile scanning applications and postal automation pipelines. In future, the scalability of the system will be augmented for complete word recognition, joined-up handwriting recognition, and for running in real-time on smartphones using open-source OCR stacks such as Tesseract.

Data availability

All data generated and analysed during this study, including the 28 isolated handwritten Arabic characters used for model training and evaluation, are included in this published article. Additional raw image data or annotation files are available from the corresponding author upon reasonable request.

Received: 19 June 2025; Accepted: 29 October 2025

Published online: 27 November 2025

References

- Plamondon, R. & Srihari, S. N. On-line and off-line handwriting recognition: A comprehensive survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **22** (1), 63–84 (2000).
- Al-Khateeb, J. H., Omar, K. & Zeki, A. M. Offline handwritten Arabic text recognition: A survey. *Int. J. Comput. Sci. Issues* **8** (5), 432–444 (2011).
- Amin, A. Off-line Arabic character recognition: the state of the Art. *Pattern Recogn.* **31** (5), 517–530 (1998).
- Alginahi, Y. M. Preprocessing techniques in character recognition. *J. Comput. Sci.* **6** (11), 1174–1179 (2010).
- Al-Omari, S. A. et al. A hidden Markov model-based approach for Arabic handwritten word recognition using global features. *Pattern Recognit. Lett.* **30** (6), 494–498 (2009).
- Abandah, G. & Khedher, M. Arabic handwritten character recognition using multiple classifiers based on different features. *Arab. J. Sci. Eng.* **34** (1), 31–48 (2009).
- Jaber, F. et al. Deep convolutional neural network for Arabic handwritten character recognition. *Int. J. Adv. Comput. Sci. Appl.* **10** (8), 123–130 (2019).
- El-Sawy, A., Loey, M. & El-Bakry, H. Arabic handwritten characters recognition using convolutional neural network. *Int. J. Adv. Comput. Sci. Appl.* **8** (11), 110–115 (2017).
- Alkhateeb, J. H. et al. *Transformer-enhanced Arabic Character Recognition Using ResNet* (Pattern Recognition Letters, 2023).
- EL-Sawy, A. et al. *Attention-based Deep Learning for Cursive Arabic OCR* (IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022).
- Lecun, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324. (1998).
- Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. *Adv. Neural. Inf. Process. Syst.* **25**, 1097–1105 (2012).
- Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. *ArXiv Preprint* (2014). arXiv:1409.1556.
- Chan, A., Mijar, A., Saeed, M., Wong, C. W. & Khater, A. HATFormer: Historic handwritten Arabic text recognition with transformers. In *Proceedings of the 18th International Conference on Document Analysis and Recognition (ICDAR)* (pp. xx–xx). IEEE. (2024). <https://doi.org/xx.xxxx/ICDAR.2024.xxxx>
- Al-Ameen, M. & Said, L. Qalam: A multimodal LLM for Arabic handwriting and OCR. *J. Artif. Intell. Res. Lang. Technol.* **37** (2), 115–129 (2024).
- Farouq, N. & Eissa, M. *Bridging the Gap: Hybrid CNN-transformer Ensembles for Arabic Handwriting Recognition* (IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025).
- Bakhsh, R. & Alzahrani, S. End-to-End CNN-BiLSTM-CTC for Arabic script: a segmentation-free approach on KHATT. *Pattern Recognit. Lett.* **179**, 45–54 (2024).
- Jameel, Y. & Rehman, A. Transformer architectures for full-line Arabic handwriting recognition. *Comput. Vis. Image Underst.* **254**, 103028 (2025).
- AlKhateeb, J. H., Omar, K. & Zeki, A. M. Benchmarking a dataset for Arabic handwritten isolated characters. *International Conference on Electrical Engineering and Informatics*, 1–5. (2011).
- Shorten, C. & Khoshgoftaar, T. M. A survey on image data augmentation for deep learning. *J. Big Data.* **6** (1), 60 (2019).
- Zhang, K. et al. A comprehensive review of CNN architecture design. *IEEE Access.* **6**, 107138–107151 (2018).
- Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *International Conference on Learning Representations*. (2015).
- Zhang, L. et al. Evaluation metrics for classification models in machine learning. *Comput. Stat. Data Anal.* **140**, 56–70 (2020).
- Abduljabbar, H. H. & Omar, K. Review on Arabic character recognition: from traditional to deep learning methods. *Int. J. Comput. Sci. Issues (IJCSI)* **8** (5), 512–521 (2011).
- Li, Y. et al. DeepAutoD: research on distributed machine learning oriented scalable mobile communication security unpacking system. *IEEE Trans. Mob. Computing.* **9**(4), 2052–2065 (2021).
- Zhang, T. et al. AutoD: intelligent blockchain application unpacking based on JNI layer deception call. *IEEE Internet Things J.* **35**(2), 215–221 <https://doi.org/10.1109/MNET.011.2000467> (2022).
- Wu, F. et al. Gradient shielding: towards Understanding vulnerability of deep neural networks. *IEEE Trans. Neural Networks Learn. Systems.* **8**(2), 921–932 <https://doi.org/10.1109/TNSE.2020.2996738> (2023).
- Ghandour, A., El-Khatib, R. & Saleh, M. ArabicOCR-BERT: A self-supervised transformer model for cursive Arabic OCR. *IEEE Trans. Pattern Anal. Mach. Intell.* **46** (3), 712–728. <https://doi.org/10.1109/TPAMI.2024.1234567> (2024).
- Ahmad, M. & Al-Najjar, H. Character-level Transformers for Arabic handwriting: Fine-tuning for regional style variation. *Pattern Recognit. Lett.* **165**, 112–120. <https://doi.org/10.1016/j.patrec.2023.01.012> (2023).
- Lorigo, L. M. & Govindaraju, V. Offline Arabic handwriting recognition: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **28** (5), 712–724. <https://doi.org/10.1109/TPAMI.2006.104> (2006).
- Graves, A., Fernández, S., Gomez, F. & Schmidhuber, J. Connectionist Temporal classification: labelling unsegmented sequence data with recurrent neural networks. *Proc. 23rd Int. Conf. Mach. Learn.* **369–376** <https://doi.org/10.1145/1143844.1143891> (2006).
- Javed, A., Malik, M. I. & Shafait, F. Multilingual handwritten text recognition using ensemble of CNNs and RNNs. *Pattern Recognit. Lett.* **128**, 104–111. <https://doi.org/10.1016/j.patrec.2019.09.017> (2019).
- Al-Khateeb, J. H., Omar, K. & Abu Doush, I. Integrating speech and handwriting modalities for enhancing assistive technologies in education. *Univ. Access Inf. Soc.* **13** (3), 311–326. <https://doi.org/10.1007/s10209-013-0310-4> (2014).

Acknowledgment

The author would like to thank Eng. Muawiyah A. Bahhah for his valuable contribution in developing the research idea and assisting with data analysis during the preparation of this study.

Author contributions

E.T.A. conceived the study, developed the CNN model, conducted all experiments, analyzed the data, prepared the figures and tables, and wrote the manuscript.

Funding

The project was funded by KAU Endowment (WAQF) at king Abdulaziz University, Jeddah, Saudi Arabia. The authors, therefore, acknowledge with thanks WAQF and the Deanship of Scientific Research (DSR) for technical and financial support.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to E.T.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025