Arabic Letter Classification Using Convolutional Neural Networks for Learning to Write Quran

Mohammad Farid Naufal[1], Muhammad Zain Fawwaz Nuruddin Siswantoro[2], and Andre[3]

[1,3]Department of Informatics Engineering, Faculty of Engineering, University of Surabaya, Indonesia
[2]Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology, Institut Teknologi Sepuluh Nopember, Indonesia
[1]faridnaufal@staff.ubaya.ac.id, [2]6025222009@mhs.its.ac.id, [3]andre@staff.ubaya.ac.id

Abstract. Learning to write the Arabic language, particularly the Arabic letters used in the Quran, is essential for individuals who aim to understand and recite the holy book accurately. In this research, we propose a classification method utilizing Convolutional Neural Networks (CNNs) with MobileNet architecture to automatically identify and classify Arabic letters. The CNN model is trained on a large dataset of labeled Arabic letter images, encompassing various styles and variations commonly found in the Quranic script. The dataset is carefully curated and annotated, incorporating a wide range of Arabic letters with different diacritics and ligatures. The significance of this research lies in its potential to support educational initiatives aimed at teaching Arabic and Quranic studies. The proposed CNN-based Arabic letter classification system can serve as an interactive learning tool, assisting individuals in recognizing and memorizing Arabic letters, thereby facilitating the process of writing the Quran. Additionally, the system can be integrated into mobile applications, making it accessible to a broader audience. The experimental results demonstrate the effectiveness and efficiency of the proposed CNN model for Arabic letter classification, validating its potential to contribute to the field of Arabic language learning. The trained CNN demonstrates remarkable performance in accurately classifying Arabic letters, achieving high accuracy rates of 94% for classifying Arabic letters and 98.43% for classifying harakat.

Keywords: Arabic letter classification, Convolutional Neural Networks, Quranic script, Arabic language learning, educational technology.

1 Introduction

Approximately 420 million individuals across over 20 countries globally communicate in Arabic. The Arabic script comprises 28 characters, as mentioned in [1]. Notable characteristics of Arabic writing involve its cursive style, interconnected characters,

M. Hartono et al. (eds.), *Proceedings of the 4th International Conference on Informatics, Technology and Engineering 2023 (InCITE 2023)*, Atlantis Highlights in Engineering 21,
https://doi.org/10.2991/978-94-6463-288-0_47

right-to-left directionality, and the presence of marks or dots that can alter the meaning of words [2]. Learning to write the Quran, particularly the Arabic letters used in the Quran, is of utmost importance for individuals seeking to gain a deep understanding of the holy book and to recite it with precision and accuracy [3]. However, acquiring proficiency in recognizing and differentiating Arabic letters can be a challenging task [4], especially for non-native speakers or individuals new to the Arabic script [5]. To address this challenge and facilitate the learning process, there is a growing need for advanced technological solutions that can assist learners in their journey of mastering Arabic letter recognition.

The development of an Arabic letter classifier has several key benefits in an educational context [6]. Firstly, it provides learners with immediate feedback and validation, allowing them to assess their progress and identify areas for improvement. Secondly, it serves as a reliable resource for teachers and instructors, assisting them in designing effective instructional strategies tailored to the needs of individual learners. Moreover, the classifier can enhance the accessibility and inclusivity of Arabic language learning, accommodating diverse learning styles and catering to learners with different levels of prior knowledge.

In recent years, machine learning and computer vision techniques have made significant advancements in various domains, including pattern recognition and image classification [7]. One such technique, Convolutional Neural Networks (CNNs) [8], has proven to be highly effective in solving complex image classification problems [9]. By leveraging the power of deep learning, CNNs can automatically learn and extract relevant features from raw images, enabling accurate classification of objects and patterns. Almekhlafi et al. [10] addresses Arabic automatic speech recognition (ASR) and classification of Arabic alphabet phonemes. A new dataset called AAPD is created with sound recordings from 1420 individuals. Experimental results highlight the effectiveness of the Mel-frequency Cepstral Coefficient (MFCC) feature extraction method, achieving a high accuracy of 95.68% with the proposed VGG-based model. Fakhet et al. [11] introduces a deep learning model for recognizing handwritten Arabic characters using Convolutional Neural Networks (CNNs). The model achieves a recognition rate of 98% when tested on the Handwritten Arabic Characters Database (HACDB) dataset. This approach addresses the challenges posed by the cursive nature of Arabic writing and the variations in character forms within words. Shams et al. [12] present a method for identifying handwritten Arabic characters and letters through the integration of deep convolutional neural networks (DCNN) and support vector machines (SVM). Their approach primarily involves assessing the likeness between input templates and pre-existing templates, resulting in remarkable accuracy in classification. Experimental outcomes underscore the efficacy of this algorithm, showcasing an impressive correct classification rate (CRR) of 95.07%, accompanied by an error classification rate (ECR) of 4.93%. In a similar vein, Almisreb et al. [13] conducted an evaluation of deep transfer learning models. Their study encompassed the assessment of seven distinct models, including AlexNet, GoogleNet, ResNet18, ResNet50, ResNet101, VGG16, and VGG19, aimed at discerning native from non-native Arabic handwriting images. Both original and augmented datasets were employed for training and validation purposes.

The findings revealed that the GoogleNet model exhibited the highest accuracy, achieving 93.2% accuracy for normal data and an even higher 95.5% for augmented data, effectively distinguishing native handwriting. Another notable contribution by Miloud and co-authors [14] introduces an innovative approach to Arabic Handwritten Recognition (AHR) systems. The method involves the utilization of optimized VGG-16 and ResNet-50 Deep Convolutional Neural Networks (CNN) architectures for extracting features and subsequent classification. The assessment was executed using the HACDB database, with the optimized ResNet-50 architecture showcasing exceptional performance, achieving an impressive 100% success rate. The comparative analysis further accentuated the efficiency and robustness of this proposed technique.

Previous studies have made significant contributions to Arabic letter recognition using deep learning techniques. However, these studies have not focused on building an educational app that provides immediate feedback. To address this gap, our research implements a CNN with MobileNet architecture, which offers lightweight computation suitable for mobile devices, ensuring real-time feedback in an educational context. The classifier will be trained on a comprehensive dataset of Arabic letter images, covering various fonts, styles, and diacritics commonly encountered in the Quranic script. Through rigorous experimentation and evaluation, we aim to demonstrate the effectiveness and efficiency of the proposed classifier model, highlighting its potential to revolutionize Arabic language education and Quranic studies. In conclusion, the development of an Arabic letter classifier using CNNs represents a significant advancement in educational technology, offering a transformative approach to learning the Arabic language and mastering the recognition of Quranic letters.

2 Methodology

The methodology subsection of this research involves dataset collection, training, testing, and performance calculation. The dataset consists of Arabic letter images, which are divided into training and testing sets. The models are trained on the training set and evaluated using the testing set. This systematic approach ensures reliable classification results and facilitates the selection of the most effective models for Arabic letter classification.

2.1 Dataset Collection

The first step in our research methodology was to collect a comprehensive dataset of Arabic letters along with their corresponding harakat (diacritical marks). The dataset consisted of 29 Arabic letters, namely Alif, Ba, Ta, Tha, Jim, Ha, Kha, Dal, Dhal, Ra, Zay, Sin, Shin, Sad, Dad, Ta, Za, Ain, Ghain, Fa, Qaf, Kaf, Lam, Mim, Nun, Ha, Waw, and Ya. Each letter was annotated with its respective harakat, which included Fathah, Damma, and Kasra. Table 1 and Table 2 show the harakat, Arabic letter, and their spelling.

Table 1. Harakat

Harakat	Spelling	Training	Testing
◌َ	A	105	132
◌ُ	U	104	184
◌ِ	I	104	134

Table 2. Arabic Letters

Letter	Spelling	Training	Testing	Letter	Spelling	Training	Testing
ا	Alif	32	177	ط	Ṭa	29	131
ب	Ba’	30	158	ظ	Ẓa	29	116
ت	Ta’	29	131	ع	‘Ain	29	101
ث	Tsa	30	108	غ	Ghain	29	111
ج	Jim	29	140	ف	Fa	29	104
ح	Kha	29	101	ق	Qaf	29	109
خ	Kha	29	109	ك	Kaf	29	100
د	Dal	29	108	ل	Lam	29	119
ذ	Żal	29	114	م	Mim	29	100
ر	Ra	29	112	ن	Nun	29	100
ز	Zai	29	107	و	Waw	29	107
س	Sin	29	103	ه	Ha’	29	124
ش	Syin	29	102	ي	Ya’	29	107
ص	Ṣad	29	100				
ض	Ḍad	29	116				

2.2 Training

In this study, the dataset was partitioned into two distinct subsets: a training set and a testing set. The chosen division ratio for this purpose was 80% allocated for training and 20% designated for testing. The training set, encompassing 80% of the dataset, was employed to facilitate the training of the CNN model. Throughout the training process, the model acquired an understanding of patterns and correlations within the input images, enabling the optimization of its parameters for enhanced classification performance. Conversely, the testing set, constituting the remaining 20% of the dataset, was employed to assess the trained model's proficiency when presented with unfamiliar samples. This division ensured a comprehensive evaluation of the model's capability to generalize and accurately classify new and unseen images. By reserving a distinct testing set, the research aimed to provide an impartial assessment of the model's precision, thereby guaranteeing its efficacy in classifying Arabic letters beyond the realm of the training data.

In the training phase of this research, the MobileNet architecture was utilized for feature extraction. MobileNet is a popular CNN architecture known for its efficiency and lightweight design, making it suitable for resource-constrained devices like mobile phones [15]. Fig. 1 shows the architecture of MobileNet.

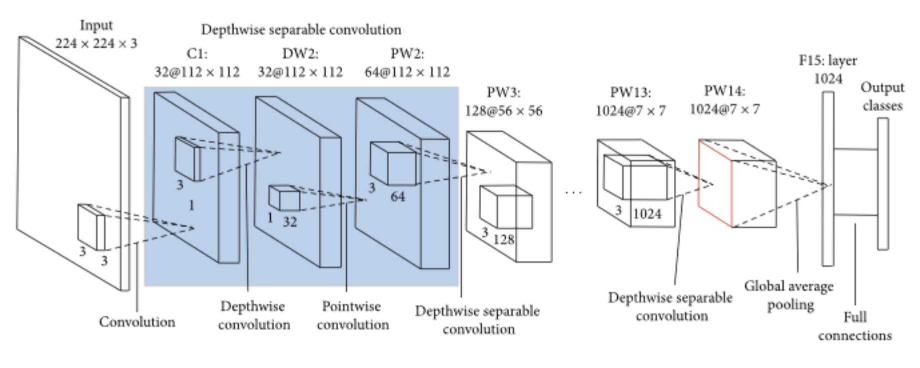

Fig. 1. MobileNet Architecture [15]

By employing MobileNet, the research aimed to extract discriminative features from the input images of Arabic letters, which are crucial for accurate classification. After the feature extraction stage, fully connected layers were incorporated into the network for classification purposes. These layers are responsible for learning complex patterns and relationships in the extracted features. This research also conducted experiments involving different types of pooling. Pooling layers are commonly inserted between convolutional layers to reduce the spatial dimensions of the feature maps while retaining important information. The pooling operation can be performed using various techniques, such as global max pooling or global average pooling. By evaluating different pooling methods, the researchers aimed to determine which approach enhances the classification accuracy for Arabic letters. MobileNet architecture was primarily utilized for feature extraction rather than the entire network architecture. This means that only the convolutional layers of MobileNet were used to capture the distinctive features, while the fully connected layers were modified and varied for optimal classification performance. The experiments with different pooling techniques provided insights into the impact of pooling on the accuracy of the classification model.

Table 3 provides detailed information about the variations in the fully connected layers used in the research. This includes the number of neurons in each layer and pooling type used in MobileNet. By exploring different configurations, this research aimed to find the optimal arrangement that yields the best classification performance.

Through these comprehensive experiments and variations, the research aimed to develop a robust CNN model for accurately classifying Arabic letters. By leveraging the MobileNet architecture for efficient feature extraction and exploring different fully connected layer configurations and pooling techniques, the study sought to improve the overall classification performance and gain insights into the most effective approaches for Arabic letter recognition.

Table 3. Model Variations

Model	Id	Pooling Type	Number of neuron on each layer	*Drop out*
Arabic Letter Classification	1	*Max*	512, 512, 512	0.1
	2	*Max*	512, 256, 512	0.1
	3	*Max*	512, 512, 512	-
	4	*Max*	1024, 512, 512	0.2
	5	*Average*	2048, 1024, 512	0.2
	6	*Average*	2048, 2048, 1024	0.2
	7	*Max*	512, 256, 512	-
	8	Max	1024, 512, 512	0.1
	9	Average	1024, 1024, 512	0.1
Harakat Classification Model	10	Max	512, 512, 512	-
	11	Average	1024, 512, 512	0.1
	12	Max	512, 512, 512	0.1

2.3 Testing

During the testing phase of the study, the model that underwent training was subjected to evaluation using a distinct testing dataset. The model's efficacy was gauged through a range of metrics, which encompassed the utilization of the confusion matrix. This matrix serves as a structured representation summarizing the outcomes of classification, detailing true positives, true negatives, false positives, and false negatives. This comprehensive framework furnishes crucial insights into the model's effectiveness by elucidating accuracy levels and potential misclassifications. Through an in-depth analysis of the confusion matrix, researchers were able to compute the accuracy metric, affording a holistic comprehension of the model's proficiency in the classification of Arabic letters and their respective harakat.

2.4 Performance Calculation

In the performance calculation subsection, various evaluation metrics were employed to assess the effectiveness of the proposed model in classifying Arabic letters. The metric used in this research is accuracy. Accuracy was calculated as the ratio of correctly classified samples to the total number of samples. It provided an overall measure of the model's classification performance. In addition, a confusion matrix was generated to provide a more detailed analysis of the model's classification results. The confusion matrix displayed the number of correctly classified instances for each class as well as the instances that were misclassified. It allowed for a deeper understanding of the model's performance on individual Arabic letter classes and helped identify areas for improvement. Based on the performance calculations, the proposed model demonstrated promising results, achieving high accuracy. The confusion matrix further revealed that the model successfully classified most Arabic letters accurately, with some variations in performance across specific classes. These performance calculations and analysis provided valuable insights into the model's efficacy in classifying Arabic letters. The performances highlighted the strengths and limitations of the proposed approach and served as a basis for further optimization and refinement of the model. Furthermore, the research aimed to select the model variation that achieved the highest accuracy for inference in a mobile app. After evaluating the performance of different model variations, the one with the highest accuracy was identified. This selection process ensured that the mobile app would utilize the most accurate and reliable model variation for predicting and classifying Arabic letters in real-time scenarios.

3 Results and Discussion

Table 4 shows the accuracy for each model variations. Among the different model variations evaluated in the research, model Id 9 stood out as the best model for classifying Arabic letters with 100% of accuracy. This model employed a specific architecture with distinct characteristics that contributed to its exceptional performance. Model Id 6 utilized a combination of global average pooling and fully connected layers. The global average pooling layer helped in reducing the spatial dimensions of the feature maps, resulting in a more compact representation. The three fully connected layers in this model contained a substantial number of neurons, with 2048 neurons in the first and second layers, and 1024 neurons in the third layer. This configuration allowed the model to capture intricate patterns and dependencies within the Arabic letter images, enabling it to learn highly discriminative features. The complexity and the high number of neurons in each layer of model number six provided a greater capacity for representation and abstraction. This allowed the model to learn intricate details and extract more nuanced information from the input data, leading to improved classification accuracy.

All model variations designed for classifying the harakat achieved an identical accuracy performance with 98.43% of accuracy. This high accuracy rate suggests that the different model variations were equally effective in accurately identifying and classifying the harakat. The reason behind this consistent accuracy performance across all model variations could be attributed to the nature of the dataset and the distinct features of the harakat. Since the harakat have well-defined and distinguishable visual patterns, it becomes relatively easier for the models to recognize and classify them correctly. Furthermore, the dataset used for training and testing the models might have contained sufficient and representative samples of the harakat. This ensures that the models were exposed to a diverse range of harakat instances, enabling them to learn and generalize well. Moreover, the models might have successfully captured the essential features and patterns associated with the harakat during the training process. This allows them to make accurate predictions even on unseen test data. The consistent accuracy performance across all model variations for classifying the harakat indicates that the harakat possess unique visual characteristics that are distinguishable and recognizable by the models. Additionally, it suggests that the chosen CNN architecture, along with the specific parameter settings, was effective in extracting and leveraging these distinctive features for accurate classification. These findings demonstrate the robustness and reliability of the developed model variations for classifying the harakat, indicating their suitability for applications involving Arabic text recognition and analysis. However, future work should involve expanding the dataset.

Table 4. Accuracy for each model variations

Model	Id	Accuracy
Arabic Letter Classification	1	90.40 %
	2	88.84 %
	3	92.00 %
	4	93.88 %
	5	89.24%
	6	93.44 %
	7	89.82 %
	8	92.60 %
	9	94 %
Harakat Classification	10	98.43 %
	11	98.43 %
	12	98.43 %

The extensive capacity of the model, along with its ability to capture intricate patterns, enabled it to discern subtle variations and distinguish between different Arabic letters more effectively. As a result, model number six achieved the highest accuracy among the evaluated variations, making it the most suitable choice for accurately classifying Arabic letters. It is important to note that the selection of this model was based on its ability to handle the complexity and intricacies of the Arabic letter dataset. The high number of neurons in each layer allowed for a richer representation of the input data, ultimately leading to superior classification performance.

Validation using the pre-test and post-test method was also conducted in three stages. In the first stage, participants used an application with a test-only feature for the pre-test. The application measured the participants' ability to recognize sounds and write them down. The scores of the participants were recorded and averaged. In the second stage, participants used the full version of the application to learn letters and harakat. This stage took place after the participants successfully completed the pre-test. In the third stage, participants reattempted the test feature available in the full version of the application for the post-test. The scores of the participants were recorded and averaged.

The validation process involved 36 participants, including 18 children aged 6-15 years and 18 users aged 20-30 years. Prior to using the full version of the application, the participants underwent a pre-test, where their average score was recorded as 57.5. After using the application for learning letters and harakat, the participants underwent a post-test, where their average score improved to 78.05. This indicates a significant increase in the participants' ability to recognize and write the sounds accurately after using the application. Fig. 2 shows the UI of the application developed in this research.

c d

Fig. 2. Player starts to write Arabic letter without harakat (a) and with harakat (b). Writing check result (c) and (d)

In the evaluation of the written Arabic letters, the classification results from the CNN model served as the basis. The checking process took place once the player finished writing on the canvas within the application. The written letter was then passed through the CNN model, and the softmax activation function provided output values indicating the probabilities of the input letter belonging to different classes. To determine the accuracy of the written letter, a threshold of 98% was set. If the output value for a particular letter exceeded this threshold, it was classified as true, indicating that the letter was correctly written. This checking process aimed to ensure that the letter recognition was performed accurately and reliably, allowing the application to provide immediate feedback on the correctness of the written letters.

4 Conclusion

In conclusion, this research successfully demonstrates the effectiveness of CNN models for classifying Arabic letters, with fathah, kasrah, and dhammah harakat. The trained models achieved accuracy rates of 94%, validating their suitability for educational purposes such as Quranic writing. However, future work should involve expanding the dataset to include other harakat, such as fathahtain, kasrahtain, and dhamahtain, to enhance the model's capability in recognizing a broader range of Arabic letter variations. By incorporating more diverse harakat, the model's performance can be further improved, enabling comprehensive Arabic letter classification and advancing the field of Arabic handwriting recognition.

References

1. Guellil, H., Saâdane, F., Azouaou, B., Gueni, D., Nouvel.: Arabic natural language processing: An overview. J. King Saud Univ. - Comput. Inf. Sci. 33(5), 497–507 (2021).
2. Lawgali, A.: A Survey on Arabic Character Recognition. Int. J. Signal Process. Image Process. Pattern Recognit 8(2), 401–426 (2015).
3. Papadopoulus, M., Ibrahim, Z.: Teaching the Arabic Alphabet to Kindergarteners - Writing Activities on Paper and Surface Computers. In: Proc. 6th Int. Conf. Comput. Support. Educ., (2014).
4. Academia,https://www.academia.edu/40769916/The_Challenges_of_learning_Modern_Standard_Arabic_as_a_Foreign_Language_in_China", Last Accessed 2023/06/15
5. Zahrani, M.: The Teaching and Learning of Arabic as a Second Language: Challenges and a Proposed Methodological Approach. ACTFL Pap 1(190), 1–70 (2018).
6. Alheraki, M., Al-Mahtam, R., Al-Khalifa, H.: Handwritten Arabic Character Recognition for Children Writing Using Convolutional Neural Network and Stroke Identification. Human-Centric Intell. Syst 3(2), 147–159 (2023).
7. Wiley, V. Lucas, T.: Computer Vision and Image Processing: A Paper Review. Int. J. Artif. Intell. 2(1), 29–36 (2018).
8. Yamashita, R., Nishio, M., Do, R., Togashi, K: Convolutional neural networks: an overview

and application in radiology. Insights Imaging, 9(4), 611–629 (2018).
9. Rawat, W. Wang, Z.: Deep convolutional neural networks for image classification: A comprehensive review. Neural Comput 29(9) 2352–2449 (2017).
10. Almekhlafi, E., AL-Makhlafi, M., Zhang, E., Wang, J., Peng, J.: A classification benchmark for Arabic alphabet phonemes with diacritics in deep neural networks. Comput. Speech Lang 71, 1-10 (2022).
11. Fakhet, W., El Khediri, S., Zidi, S.: Guided classification for Arabic Characters handwritten Recognition. Proc. IEEE/ACS Int. Conf. Comput. Syst. Appl. AICCSA, pp. 1-6. Abu Dhabi (2022).
12. Shams, M., Elsonbaty, A., Elsawy, Z.: Arabic Handwritten Character Recognition based on Convolution Neural Networks and Support Vector Machine. IJACSA Int. J. Adv. Comput. Sci. Appl., 11(8) (2020).
13. Almisreb, A., Tahir, M., Turaev, S., Saleh, A., Al Junid, M.: Arabic Handwriting Classification using Deep Transfer Learning Techniques. Pertanika J. Sci. Technol 30(1) 641–654 (2022).
14. Miloud, K., Lakhdar, M., Ilyas, R.: Arabic Handwriting Recognition System Based on Genetic Algorithm and Deep CNN Architectures. 2021 Int. Conf. Decis. Aid Sci. Appl. DASA 2021, 583–586, Shakeer (2021).
15. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, C.: MobileNetV2: Inverted Residuals and Linear Bottlenecks. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 4510–4520, Salt Lake City (2018).