Combination of Machine Learning Algorithms and Resnet50 for Arabic Handwritten Classification

Raidah S. Khudeyer, Noor M. Al-Moosawi
College of Computer Science & Information Technology, University of Basrah, Iraq
E-mail: raidah.khudayer@uobasrah.edu.iq, almoosawinoor2@gmail.com

Keywords: CNN, Deep Learning, Arabic Handwritten, Random Forest, Support Vector Machine

Received: September 6, 2022

The recognition or classification of Arabic handwritten characters is extremely crucial in many applications and, at the same time, one of the biggest challenges that machine learning faces. The emergence of deep learning, particularly Convolutional Neural Networks (CNN), is considered a suitable technique to face these challenges. In this research paper, an investigation model is proposed to make recognition for Arabic handwriting utilizing one of CNN architectures: ResNet50 architecture, after replacing the last layer with one of two types of machine learning algorithms, Support Vector Machine (SVM) and Random Forest (RF), to reduce training time and increase overall accuracy. Our experimental work was performed on three data sets: Arabic Handwritten Character Dataset (AHCD), Alexa Isolated Alphabet Dataset (AIA9K), and Hijja Dataset. Experimental results show that combining ResNet50 with random forest produces more accurate and consistent results than the ResNet50 model produces by itself.
Finally, the comparison with the other methods across all data sets demonstrates the robustness and effectiveness of the combination of random forest with the ResNet50 approach. Where the modified ResNet50 architecture has achieved a rate of 92.37%, 98.39%, and 91.64%, while the combination architecture has achieved 95%, 99%, and 92.4% for AIA9K, AHCD, and Hijja datasets, respectively.

Povzetek: Avtorji so razvili novo metodo kot nadgradnjo Resnet50 z dodatkom zadnjega nivoja v obliki SVM in RF. Na več domenah je dosegla boljše rezultate kot osnovni Resnet50.

1 Introduction

In recent years, methods that rely on deep learning, particularly Convolutional Neural Networks (CNNs), have excelled in several areas, such as the classification of images, object detection, facial recognition, fingerprint analysis, computer-aided diagnosis, and facial expressions. CNN is the most current method for extracting highly discriminative features that make our task more reliable [1].

Character recognition technologies provide users with an automatic mechanism for recognizing the text on an image. These technologies are used in many verification applications, e.g., verifying official documents and bank cheques.

Handwriting recognition is considered a more challenging task in computer vision because handwriting varies in the sizes and styles of a writer's handwriting characters [2]. However, the issue is that the majority of these experiments are conducted in English, as it is the most widely spoken language in the world [3]. Recognizing handwritten Arabic characters is a complicated process compared to English because of the nature of Arabic words and due to the Arabic script's unique characteristics, e.g., its cursive nature, diacritics, and diagonal strokes, developing an Arabic character recognition system is still challenging [4]. Deep learning techniques have become more prevalent in the field of Arabic recognition over the last few years.

Many Machine Learning (ML) algorithms have been successfully used for Arabic Handwritten Recognition and in recent years, deep learning techniques have become more prevent in this field. Where Arabic Handwritten Recognition technologies can be significantly improved by introducing Deep Learning (DL) architectures, especially Convolutional Neural Networks (CNNs) [2], [5], [6].

The survey paper of Alrobah1 et al. [7] presents a comprehensive review of all the works reported in Arabic Handwritten recognition that use deep learning approaches.

The combination of one or more systems is a popular way of improving accuracy in different tasks, where the new system performs the same task to exploit the unique advantage of each system and reduce some of the random errors [8]. In this work, we combine two commonly used Machine Learning (ML) algorithms: RF and SVM algorithm with CNN which are used for the classification task.

In SVM, in which you can visualize unprocessed data as points in an n-dimensional space (where n is the number of features you have). each feature's value is then connected to a specific coordinate, making it easy to classify the data. The data can be divided into groups and plotted on a graph using lines known as classifiers.

Random Forest (RF) is a collective of decision trees. To classify a new object based on its attributes, each tree is classified, and the tree "votes" for that class. the classification with the most votes is chosen by the forest.

2 Machine learning (ML)

ML is a computer science discipline, a branch of AI that includes all the tools and techniques to enable computers to learn from data and to allow computers to function without being explicitly programmed [9], [10].

Deep Learning (DL) is a broader ML method subfield that has increased interest lately [11]. It is a term that refers to structures with multiple hidden layers that aid in the learning of features at various abstraction levels. Numerous DL algorithms have the primary advantage of transforming input data (e.g., images) to outputs (e.g., binary value, True or False, or class label in multiclass) while capturing increasingly higher-level features [11], [12].

Unlike traditional ML methods, which require the model's designer to select and encode features in advance, DL enables the model to learn relevant features automatically. In DL, a network learns how important a feature is to the output by applying weights to its connections [13]. Simply feed the raw image, and the network will identify patterns within inputs. This is critical, as feature engineering may frequently be the most time-consuming aspect of ML practice [10]. Figure 2.1 compares hierarchical representation learning the features in DL layer by layer to ML approaches.

On the other hand, due to advancements in GPU technology, faster network connectivity, and improved software infrastructure for distributed computing, multiple layers can now work in concert, resulting in increased accuracy, complexity, and impact for DL models [13], [12].

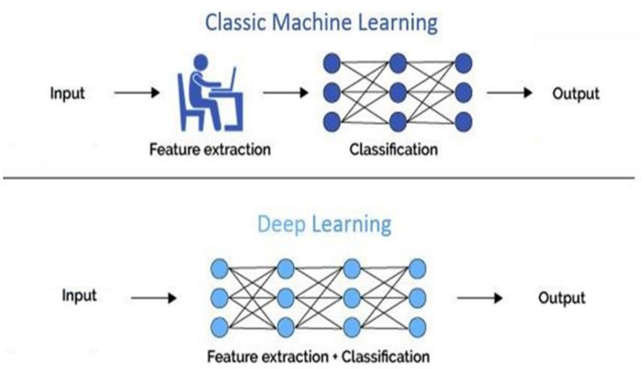

Figure 2.1: ML using handcrafted feature extraction vs. features learn automatically in DL.

2.1 Deep learning architectures

Today, several DL architectures are built based on ANNs as vector graphic models with a specific architecture consisting of the neuron's building blocks. Generally, these DL architectures involve three essential structures: recurrent, fully connected, and convolutional neural networks, where each network is suitable for a specific task [10]. CNN is considered the power core of computer vision and has widespread use and success in several applications [14], [15]. We have chosen CNN as the baseline supervised DL model for the classification of Arabic handwritten because it exploits local links in the images.

2.2 Convolutional neural network

CNN was first developed successfully by LeCun et al. for handwritten image recognition [16]. CNN, like MLP, has several neurons arranged in layers, but they're arranged and connected differently, it relies on filters with convolution operation applied on input data, with the fact that the weights are shared to form filters.

Each layer in a CNN learns increasingly complex kernels. The initial layers are taught fundamental features, and the middle layers acquire filters for detecting the components of objects. The final layers have a higher level of representation to recognize complete objects in various shapes and sizes [10], [12].

In the standard CNN architecture, a multi-stage stack with linear and non-linear operations is composed of feature extraction and classification. The feature extraction stage contains main layers stacked in sequences such as the convolutional layer (Conv layer) with an activation function and a pooling layer, while the classification stage has several fully connected layers.

One of the popular CNN models, ResNet (Residual Network), was first introduced in 2015 [17]. ResNet includes residual connections that enable the ConvNet to learn not only from the activation function's output but also from the previous layer's input (Figure 2.2). In this way, information is transferred from one layer to another with the easily gradient flow during backpropagation due to the additional operations that distribute the gradients.

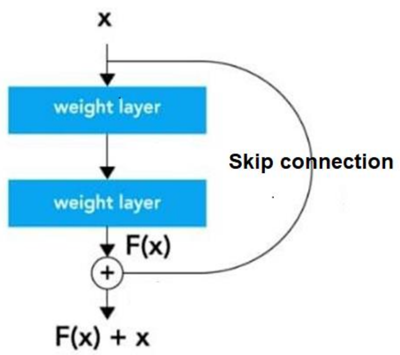

Figure 2.2: A learning block of residual learning.

2.3 ResNet50

There are many types of ResNet, different in the number of layers deep, such as ResNet50. ResNet50 is ResNet run on the same concept but has 50 layers deep as in Figure 2.3.

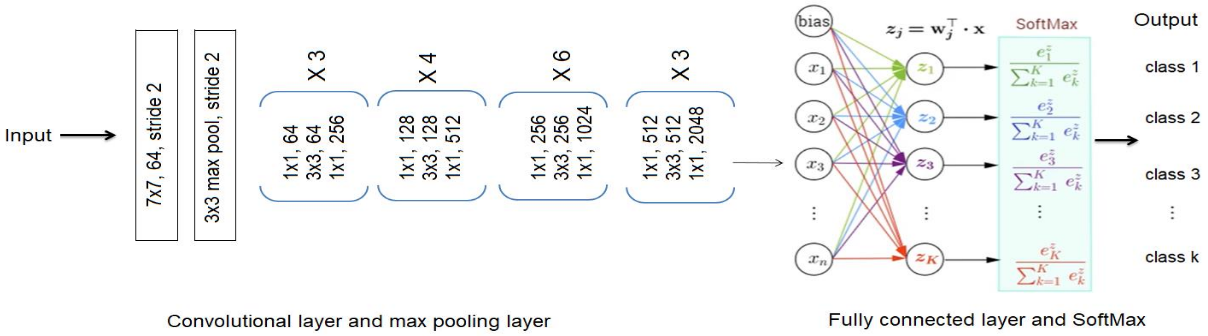

Figure 2.3: ResNet50 architecture.

3 Modified ResNet50

Due to the fact that the ResNet50 model was pre-trained on ImageNet[1] data, which contains a thousand categories, some modifications were made and fine-tuned to fit our task and reduce trainable parameters while increasing accuracy. We improved the traditional ResNet50 model structure with two different approaches for research and access to the best and most efficient model for the classification of Handwritten Arabic letters, as illustrated in Figure 3.1. The proposed work in two approaches can be summarized as follows:

Table 3.1: Summary of parameters and techniques used for training.

Hyperparameters	Setting
Input Size	32 x 32
Batch size	64
learning Rate	0.001
Epochs	100
Early stopping	After 10 epochs Val-loss does not improve
Optimizer	Adam
Loss function	Sparse_Categorical_Crossentropy

a- First approach:

The convolutional base in the original ResNet50 (Conv and pooling layers) was used as a core part of feature extraction while the FC layer (dense layer) was removed and replaced with two dense layers. The first dense layer with 256 nodes was used to capture the features extracted from the convolution base after flattening by the flattening function. The second dense layer was set to be equal to the number of categories in Arabic characters (28 nodes) with used SoftMax as an activation function for the final classification task.

During the training phase, the ImageNet weights were used as a starting point for weights and biases in ResNet50 kernels to improve the performance as much as possible and then retrained to 100 epochs. Adam was used as an optimization algorithm with a 0.001 learning rate.

The trained model with optimal parameters was saved and then used during the testing phase for classifying new letters images. Optimal regularization parameters for training modified ResNet50 are presented in Table 3.1.

b- Second approach:

The full connection portion in the traditional CNN consumes a huge computing resource and increases the risk of an overfitting problem, so the calculation effectiveness of this model is inefficient due to defects in the fully connected layer when processing images. To solve this problem, the proposed work in the second approach involved creating ResNet_SVM and ResNet_RF models by replacing the final dense layer with non-linear Support Vector Machine (SVM) and Random Forest (RF) as classifiers, which effectively reduce learnable parameters and total training time.

Firstly, each handwritten Character image is directly fed foreword into the previously trained model from the first approach as input after freezing the convolutional base with the trained weights. Secondly, the representative feature maps are output from the last convolutional layer converted to one-dimensional feature vectors with 2048 features. Finally, the dimensionality of the derived feature vectors was reduced to 256 in the first dense layer and then utilized to train SVM and RF as classifiers. Optimal regularization parameters for SVM and RF are present in Table 3.2.

[1] https://image-net.org/challenges/LSVRC/2015/

Table 3.2: SVM and RF parameters setting.

ResNet_RF		ResNet_SVM	
Parameter	**Value**	**Parameter**	**Value**
n_estimators	500	Kernel	‘RBF’
max_depth	80	C	10
		γ	0.001

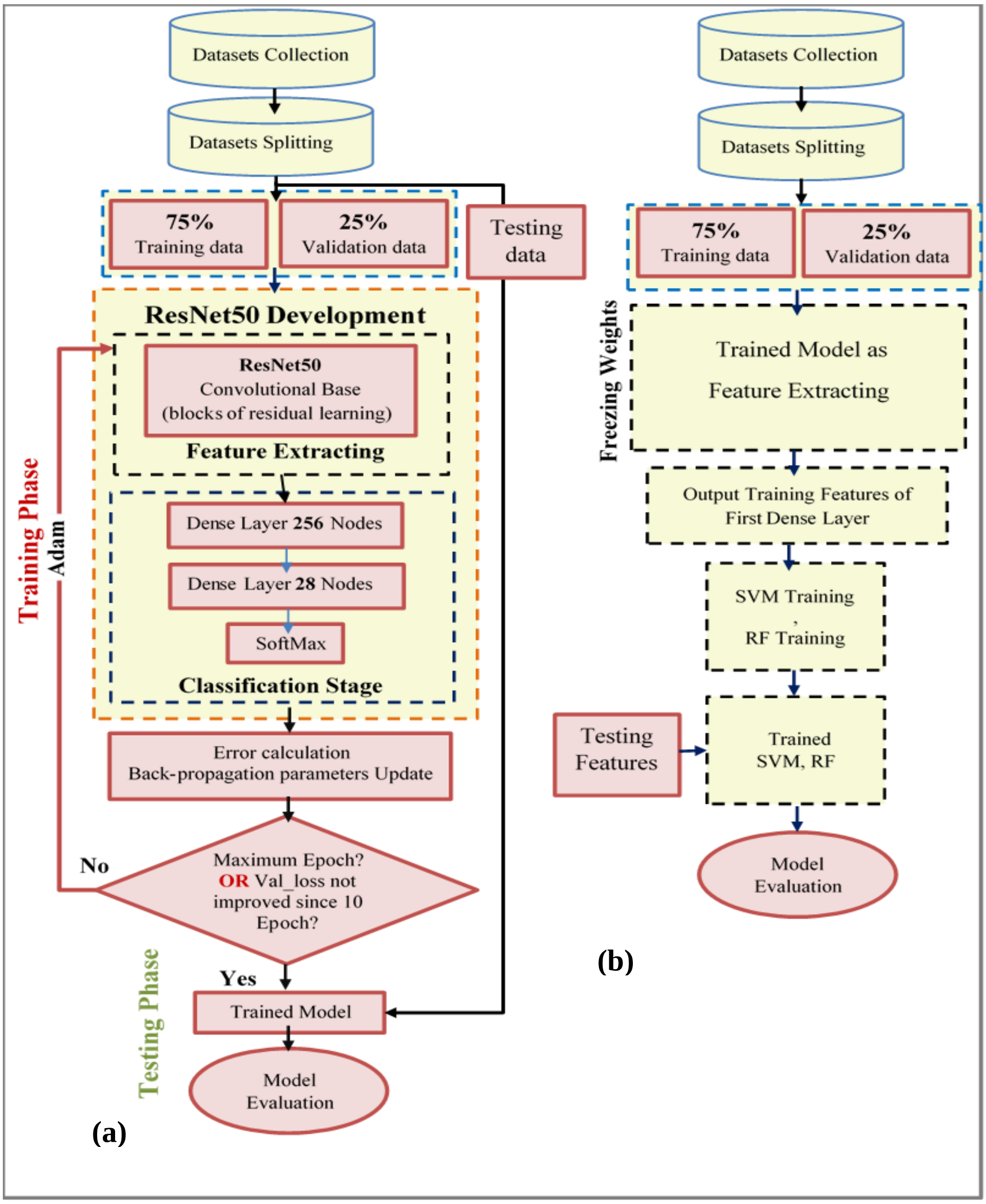

Figure 3.1: An illustration of our proposed framework in two approaches,
(a) the first proposed approach, (b) the second approach.

4 Data sets

Below is the description of three different publicly available datasets of Arabic handwritten, that were used to evaluate this work.

4.1 Arabic handwritten character dataset (AHCD)

This dataset is written by 60 participants with an age range from 19 to 40 years and most participants are right-handed. Each participant wrote the letters (from "Alef" to "Yeh") ten times to create the 16,800 characters that make up this dataset [18].

4.2 AlexU isolated alphabet (AIA9K) dataset

The AlexU Isolated Alphabet (AIA9K) dataset, created in 2014 by Torki et al. [19], is a collection of handwritten Arabic letters. This dataset introduces a compacted 9K novel dataset of 28 classes that represent the isolated 32x32 pixel Arabic handwritten alphabet. This dataset was compiled from 107 B.Sc.-educated volunteer writers who were between the ages of 18 and 25 or M.Sc. students. The writers, who included 45 men and 62 women, wrote each Arabic letter three times. 8,737 letters make up the entire collection of characters that are valid.

4.3 Hijja dataset

Altwaijry et al. [20] created their own dataset, Hijja, of handwritten Arabic letters in 2020. Hijja is a free public dataset of single letters assembled by Arabic-speaking schoolchildren between the ages of 7 and 12. In Riyadh, Saudi Arabia, data were collected from January to April 2019. 47,434 characters total, written by 591 participants in various forms, are represented by it.

5 Results

Our models were implemented in python3 through the Kaggle framework which provides free GPU for 30 hours a week. For a fair comparison, all training datasets were partitioned into two parts training and validation with the same ratio of **75:25**. The training set was used for learning and the validation set checked the model performance after each training phase to fine-tune hyper parameters and stop training early. The evaluation of each model's performance on the testing dataset (unseen data) is given in Table 5.1. In this work, all models used regularization methods to prevent overfitting and optimized methods to control the parameters of the models to achieve the best accuracies.

Table 5.1: Models performance on the testing set.

Model	Dataset	Size	Acc. %
ResNet50	AIA9K	8737	**92.37**
	AHCD	16,800	**98.39**
	Hijja	47,434	**91.64**
ResNet_SVM	AIA9K	8737	**93.92**
	AHCD	16,800	**97.4**
	Hijja	47,434	**91.83**
ResNet_ RF	AIA9K	8737	**95**
	AHCD	16,800	**99**
	Hijja	47,434	**92.4**

Table 5.2 shows the comparison between our works and previous works that used the same dataset, which appears our work approximately gives more accuracy.

Table 5.2: Models performance on the testing set compared with another researcher's work.

Researcher (s)	Year	Model	Dataset	Acc. %
Torki et al.	2014	SVM-RBF	AIA9K	94.28
El-Sawy et al.	2017	CNN	AHCD	94.9
Younis	2018	CNN	AIA9K AHCD	94.8 97.6
De Sousa	2018	CNN	AHCD	98.42
Najadat et al.	2019	CNN	AHCD	97.2
Almansari et al.	2019	CNN	AHCD	95.27
		MLP		72.08
Alyahya et al.	2020	CNN	AHCD	98.30
Altwaijry et al.	2020	CNN	AHCD	97
			Hijja	88
Balaha et al.	2020	CNN	AHCD	97.3
			AIA9k	98.4
Our Work	**2022**	**ResNet 50**	AIA9K	**92.37**
			AHCD	**98.39**
			Hijja	**91.64**
		ResNet SVM	AIA9K	**93.92**
			AHCD	**97.4**
			Hijja	**91.83**
		ResNet RF	AIA9K	**95**
			AHCD	**99**
			Hijja	**92.4**

6 Discussion

ResNet model with SVM and with Random Forest approximately give the best accuracies compared with only the ResNet model. Svm and RF are classifier algorithms thatwork on the features given by the CNN algorithm making the process of classification faster and more accurate and don't need more experiments to build the best full connection layer.

7 Conclusion

This article demonstrates the effectiveness and viability of the suggested work for categorizing Arabic handwritten based on the modified ResNet50 models. The proposed models used ResNet50 as a fully automated feature extraction and replaced a fully connected layer in the original ResNet50 model with one of the machine learning algorithms such as the CNN-SVM or CNN-RF models. Due to Resnet's high feature extraction ability, SVM and RF training became more effective and accurate, improving letter classification accuracy. The modified models had a lower overall trainable parameter count and training time, but they still had the highest accuracy.

8 References

[1] N. M. Al-Moosawi and R. S. Khudeyer, "ResNet-34/DR: A Residual Convolutional Neural Network for the Diagnosis of Diabetic Retinopathy," Inform., vol. 45, no. 7, pp. 115–124, 2021.
https://doi.org/10.31449/inf.v45i7.3774

[2] N. Altwaijry and I. Al-Turaiki, "Arabic handwriting recognition system using convolutional neural network," Neural Comput. Appl., vol. 33, no. 7, pp. 2249–2261, Apr. 2021.
https://doi.org/10.1007/s00521-020-05070-8

[3] H. T. Weldegebriel, H. Liu, A. U. Haq, E. Bugingo, and D. Zhang, "A New Hybrid Convolutional Neural Network and eXtreme Gradient Boosting Classifier for Recognizing Handwritten Ethiopian Characters," IEEE Access, vol. 8, pp. 17804–17818, 2020.
https://doi.org/10.1109/access.2019.2960161

[4] S. Albahli, M. Nawaz, A. Javed, and A. Irtaza, "An improved faster-RCNN model for handwritten character recognition," Arab. J. Sci. Eng., vol. 46, no. 9, pp. 8509–8523, 2021.
https://doi.org/10.1007/s13369-021-05471-4

[5] C. Boufenar, A. Kerboua, and M. Batouche, "Investigation on deep learning for off-line handwritten Arabic character recognition," Cogn. Syst. Res., vol. 50, pp. 180–195, Aug. 2018.
https://doi.org/10.1016/j.cogsys.2017.11.002

[6] I. Palatnik de Sousa, "Convolutional ensembles for Arabic Handwritten Character and Digit Recognition," PeerJ Comput. Sci., vol. 4, p. e167, Oct. 2018.
https://doi.org/10.7717/peerj-cs.167

[7] N. Alrobah and S. Albahli, "Arabic Handwritten Recognition Using Deep Learning: A Survey," Arab. J. Sci. Eng., no. January, 2022.
https://doi.org/10.1007/s13369-021-06363-3

[8] M. Alabbas, R. S. Khudeyer, and S. Jaf, "Improved Arabic characters recognition by combining multiple machine learning classifiers," in Proceedings of the 2016 International Conference on Asian Language Processing, IALP 2016, 2017, pp. 262–265.
https://doi.org/10.1109/ialp.2016.7875982

[9] N. Silaparasetty, "Programming with Tensorflow," in Machine Learning Concepts with Python and the Jupyter Notebook Environment, K. B. Prakash and G. R. Kanagachidambaresan, Eds. Berkeley, CA: Apress, 2020, pp. 173–189.
https://doi.org/10.1007/978-1-4842-5967-2_9

[10] M. A. Wani, F. A. Bhat, S. Afzal, and A. I. Khan, Advances in Deep Learning, vol. 57, no. June. 2019.

[11] M. A. Ranzato, Deep Learning for Vision Systems, no. October. 2013.

[12] X. Ma et al., "Understanding adversarial attacks on deep learning based medical image analysis systems," Pattern Recognit., vol. 110, 2021.
https://doi.org/10.1016/j.patcog.2020.107332

[13] E. Alpaydin, "Neural Networks and Deep Learning," Mach. Learn., 2021, doi: 10.7551/mitpress/13811.003.0007.

[14] A. Kumar, S. Sarkar, and C. Pradhan, "Malaria Disease Detection Using CNN Technique with SGD, RMSprop and ADAM Optimizers," no. January, pp. 211–230, 2020.
https://doi.org/10.1007/978-3-030-33966-1_11

[15] A. Haque, I. Sutradhar, M. Rahman, M. Hasan, and M. Sarker, "Convolutional Nets for Diabetic Retinopathy Screening in Bangladeshi Patients," Jul. 2021, [Online]. Available: http://arxiv.org/abs/2108.04358.

[16] H. S. R. Rajula, G. Verlato, M. Manchia, N. Antonucci, and V. Fanos, "Comparison of conventional statistical methods with machine learning in medicine: Diagnosis, drug development, and treatment," Med., vol. 56, no. 9, pp. 1–10, 2020.
https://doi.org/10.3390/medicina56090455

[17] S. Wu, S. Zhong, and Y. Liu, "Deep residual learning for image Recognition," Multimed. Tools Appl., pp. 1–17, 2017, doi: 10.1007/s11042-017-4440-4.

[18] M. N. Aljarrah, M. M. Zyout, and R. Duwairi, "Arabic Handwritten Characters Recognition Using Convolutional Neural Network," in 2021 12th International Conference on Information and Communication Systems, ICICS 2021, 2021, pp. 182–188.
https://doi.org/10.1109/icics52457.2021.9464596

[19] M. Torki, M. E. Hussein, A. Elsallamy, M. Fayyaz, and S. Yaser, "Window-Based Descriptors for Arabic Handwritten Alphabet Recognition: A Comparative Study on a Novel Dataset," arxiv.org, 2014, Accessed: Aug. 05, 2022. [Online]. Available: https://arxiv.org/abs/1411.3519.

[20] Khaled younis and abdullah khateeb, "Arabic Hand-Written Character Recognition Based on Deep Convolutional Neural Networks," Jordanian J. Comput. Inf. Technol., vol. 3, no. 3, p. 186, 2017.
https://doi.org/10.5455/jjcit.71-1498142206