Handwritten Arabic Digit Recognition Using Convolutional Neural Network

Jawad H Alkhateeb
Computer Science Department, Taibah University, AlMadinah, Saudi Arabia

Abstract—In Computer vision systems, computer vision works by imitating humans in their vision way which is known as the human vision system (HVS). In HVS, humans use their eyes and brains in order to see and classify any object around them. Hence, computer vision systems imitate HSV by developing several algorithms for classifying images and objects. The main goal of this paper is to propose a model for identifying and classifying the Arabic handwritten digits with high accuracy. The concept of deep learning via the convolutional neural network (CNN) with the ADBase database is used to achieve the goal. The training is done by having a 3*3 and 5*5 filters. Basically, while the classification phase distinct learning rates are used to train the network. The obtained results are encouraging and promising.

Keywords—Deep Learning; CNN; ADBase database; ReLU.

1. Introduction

Mainly, some various applications and domains get benefits from the field of handwritten recognition (HWR). Those applications include, but not limited to, mail sorting, signature verification, cheques verification, office automation, and a large variety of banking, business as well as natural human-computer interaction. Basically, the HWR is categorized into two main tasks the online and offline based systems. In the online base systems, the computer traces the writing process. However, the offline based systems deal with the available digital images. This paper emphasizes on the offline line based systems.

The people who are using Arabic are estimated more than 250 million [1 -3]. The Arabic language is less investigated by researches comparing to other languages such as English, Latin, Japanese, and Chinses. Working in Arabic HWR is a big challenge. This challenge is due to the difficulty of the Arabic language which is different and distinguished from other languages. Basically, Arabic is written from right to left in a cursive way. The Arabic language has 28 main alphabet letters. Among those 28 letters, there are 16 letters have dots. The dots are categorized into three main types: one dot, two dots, or three dots. The main goal of the dots is to differentiate between similar letters that have the same shape. The following facts can be stated regarding the dots:

1) There are ten Arabic letters have one dot ( ) ب, ج, خ, ذ, ز, ض, ظ, غ, ف, ن.
2) Three Arabic letters have two dots (ت, ق, ي).
3) Two Arabic letters have three dots (ث, ش).

Removing any of the dots will lead to a misinterpretation of the letter. Thus, an efficient technique of pre-processing has to be applied to keep the meaning of the letter.

Since Arabic writing in cursive some letters are connected to the previous or/and following letters and some letters do not connect. So, the shape of the Arabic letter depends on its positions in writing. The letter might have four different shapes depending on the letter position. The four shapes of the Arabic letter are: isolated, connected from the left (Ending Form), connected from the right (beginning form), or connected from both sides (middle form) [4].

Removing any of the dots will lead to a misinterpretation of the letter. Thus, an efficient technique of pre-processing has to be applied to keep the meaning of the letter.

Since Arabic writing in cursive some letters are connected to the previous or/and following letters and some letters do not connect. So, the shape of the Arabic letter depends on its positions in writing. The letter might have four different shapes depending on the letter position. The four shapes of the Arabic letter are: isolated, connected from the left (Ending Form), connected from the right (beginning form), or connected from both sides (middle form) [4].

From the last three decades, the researches start working in Arabic text recognition. AlKhateeb [1] presented an offline Arabic handwritten digit recognition system using the Dynamic Bayesian Network (DBN). Features were extracted by using the sliding window technique. The system was evaluated on the ADBase and reported an average recognition rate of 95.26%.

AlKhateeb et al. [5] proposed a system to identify the handwritten Arabic digit by using the Dynamic Bayesian Network (DBN). Features were extracted by discrete cosine transform coefficients. Their system was evaluated on ADBase and reported an average recognition rate of 85.26%.

A handwritten Arabic digit recognition based on the convolutional neural network (CNN) was explored. This system was evaluated on a private dataset ending up with a recognition rate of 99.76% [6].

A deep learning novel model algorithm for Arabic numeral recognition proposed by Ashiquzzaman et al. [7]. This model reported an accuracy of 97.4%.

Arabic handwritten digits were investigated using the LeNet-5 convolutional neural network (CNN) by El-Sawy et al [8]. Their system was evaluated on the ADBase dataset with an accuracy of 88%.

Al-Hmouz et al. [9] proposed a digit numeral recognition system for Arabic, English, Devanagari, and Persian numbers. Structural features were extracted using the image geometrical primitives. The neural network and Naïve Bayes classifiers were used in evaluating the system end up with very good accuracy.

Khorsheed [10] addressed a new technique of handwritten Arabic script, where the segmentation into letters is not required at all. The technique found the skeleton of the word and decomposed the skeleton into an observation sequence, later, a single hidden Markov model (HMM) which used the structural features was used as a classifier.

Alma'adeed et al [11] addressed a recognition system for unconstrained Arabic handwritten words using HMM. A complete scheme system was addressed by Alma'adeed [12] for recognizing the unconstrained Arabic handwritten words using neural networks. T. Khan [13] proposed a deep learning model for recognizing the expiry date by getting an automatic notification by using smart expiry architecture. A. Ashiquzzaman and A. K. Tushar [14] proposed a novel algorithm using deep learning neural networks to improve and modify the recognition accuracy in recognizing Arabic numeral. Saad Bin Ahamd et al. [15] addressed the Arabic scene text recognition using Convolutional Neural Networks (ConvNets) classifier. Concerning the single occurrence of the Arabic character, five orientations were employed. Furthermore, during the classification phase, the training was formulated by keeping the size of the as 3 × 3, and 5 × 5 were kept with stride values 1 and 2. Katrina Sundus et al. [16] addressed the concept of the neural network of the feed-forward deep learning (DL) in recognizing the Arabic text. Layers were used in their model, in the first layer the frequency-inverse document (TF-IDF) where most of the frequent words were represented by vectors. The input of the second layer is the output of the first layer. Adam's optimizer was used to reduce the classification error. G. Saker et al. [17] addressed a recognition system to Arabic fonts. The convolution neural networks were used to find the name of the Arabic font. They used their dataset. Maidana et al. [18] used various convolutional neural network architectures for recognizing handwritten chinses characters. Several and multi distinct architectural fusion methods were employed.

The rest of the paper is organized as follows: Section 2 presents the ADBase database. Section 3 presents the background. Section 4 presents the proposed methodology. Results and discussion are presented in section 5. Finally, section 6 presents the conclusion.

2. ADBase Database

Ideally, developing a robust recognition system requires a huge database for training and testing the recognition system.

Frankly, dealing with real data from post offices with postal code, or banks for wither signature or cheques verification are considered as a confidential and cannot be accessed for doing research. In the literature, some work was carried out in Arabic handwritten, where they had small databases of their own. Also, some work was carried out using private databases that were unavailable to the public. Consequently, there was no benchmark for comparing the obtained results by researches. The ADBase database is available for free, is very important in this context as it has been used as a standard test set in such a context [13]. El-sherif and Abdleazeem [19] published an Arabic handwritten digit database (ADBase) which contains 70,000 digits written by 700 writers. Each writer wrote each Arabic digit (from 0 -9) ten times. To ensure having variation and different writing styles, the database was collected from various institutions: a high school, School of Medicine, College of Law, Colleges of Engineering, the Open University, and a governmental institution. Forms were scanned by using a 300 dpi scanner which was adjusted to produce a binary image. later the Arabic digits were automatically extracted, categorized, and bounded by bounding boxes. While producing the binary images, corrupted and noisy digit images were edited manually. Mainly, the database is divided into two main sets: the training set and the testing set. The training set contains 60,000 digits where each class has 6,000 images, while the testing set contains 10,000 digits where each class has 1000 images. The ADBase is available for free for researchers, and it can be accessed via http://datacenter.aucegypt.edu/shazeem/ Figure 1 shows samples of ADBase.

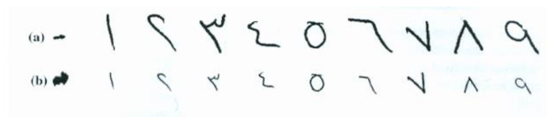

Fig. 1. Example of Arabic Digit

3. Background

Deep learning is considered as a subpart of machine learning. By using the backpropagation algorithm, deep learning can find out any complex structure in the form of massive data sets [20] [21]. Various tasks such as, but not limited to, image classification, image recognition, and other applications can be handled in deep learning. Deep learning can be achieved by applying the Convolutional Neural Networks (CNN / ConvNets) [22].

The CNNs are similar to Artificial Neural Networks (ANNs). Basically, the CNNs are considered as one of the deep learning classes. The CNNs are composed of both ANN and convolution. Firstly, the ANNs consist of artificial neurons that receive the set of inputs X = (x1, x2, x3, …….., xn). These inputs are connected to a function f(x) he artificial neuron receives one or more inputs. Furthermore, Each neuron has a set of weight vectors w = (w1, w2, w3, …….., wn). The neuron output is sent as input to another neuron of another layer with repeating the same computation (weighted sum of the input and transformation with activation function). The final network output (Y) is biased toward value as illustrated in equation 1.

Fig. 2 illustrates the model of artificial neurons in the ANNs [20] [23].

$$Y = \mathrm{f}\sum_{i=1}^{N\infty}(x_i w_i + \mathrm{b}) \qquad (1)$$

Secondly, the second part of the CNNs is the convolution operation. Usually, in the convolution operation, a simple filter is applied to the input image. The filter is known as wither mask, kernel, or window. Basically, the convolution filter passes over all the input image pixels at a given time where the dot product is taken of both convolution filter and the input image pixels resulting in one final value output. Fig 3. Illustrates the convolution filter operation [22].

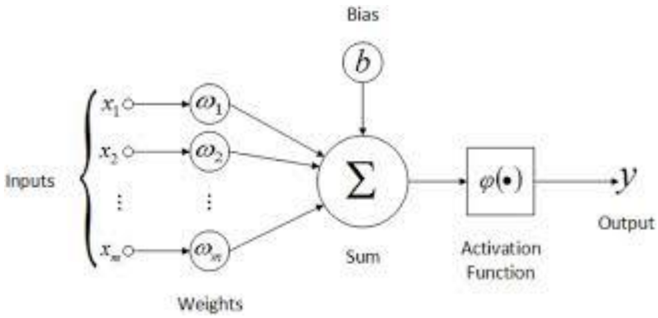

Fig. 2. The ANNs Neuron Model.

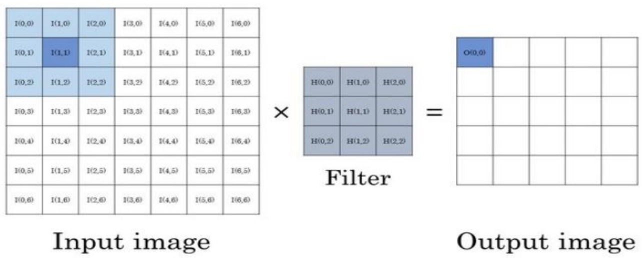

Fig.3. The convolution filter operation [26]

A. Convolutional Neural Networks (CNN)

The CNNs are similar to the multi-layer network (MLP). Recognizing a pattern from pixel images is the main goal of the CNNs. The CNNs are considered as deep learning algorithms [24] [25]. Generally, the CNNs are combined of three main layers. These three layers are the convolution layers, the pooling layers, and the fully connected layers [26].

B. Convolution Layer

In the CNN, the convolution layer is considered as the first layer used for extracting features from the input image. The convolution is applied to the input image using a convolution filter resulting in a feature map. Striding is taking in care in this layer. Striding means the number of pixel shifts across the input image at each step [20] [26].

C. Pooling Layer

The pooling layer is used for spatial size reduction. It is used to reduce computation and the number of parameters in the network. Mainly, the pooling layer operates on each feature map resulting in the convolution layer independently. In pooling, there are various pooling algorithm approaches: the average pooling, the stochastic pooling, and the max pooling. Max-pooling is the most common algorithm approach used in the pooling layer [26].

D. Fully Connected Layer

The fully connected layer is the last layer on CNN. In this layer, the last output in the pooling layer is considered the input of the fully connected layer. Actually, the behavior of this layer is similar to the ANNs, in which each neuron of the previous layer is connected to the next present layer. While comparing the first layer of the CNNs with the last layer, it is noted that the number of parameters in the fully connected layer is higher than the number of parameters in the convolution layer. In general, the fully connected layer is connected to the output layer which is known as a classifier [26].

E. The Activation function

While using the CNNs, various activation functions have been employed with various architectures of CNNs. Among those nonlinear activations, functions are LReLU, PReLU, and ReLU. Basically, using these functions increase the performance of the system in the training by speeding it up. In this paper, the ReLU (Rectified Linear Unit) is applied. Fig. 4 summarizes the architecture of CNN.

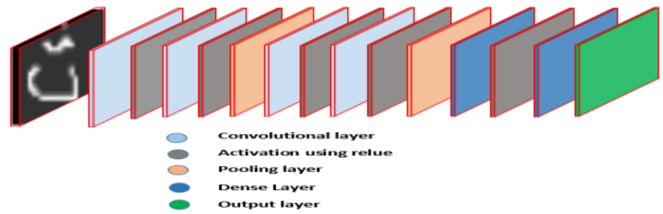

Fig.4. The convolution Neural Network architecture [26].

4. Proposed Methodology

Training and recognition are the main parts of the proposed model. Constructing, designing the architecture of the network, and training the network with the input image data is the main phase in the training part. However, recognizing the Arabic digits by using the training model is the main phase in the recognition part.

The ADBase database is used in the proposed model which contains 70000 images of Arabic digits written by 100 writers. Originally, The Arabic digit images were cropped and saved in gray format. During the development phase. In this paper, the images were converted to binary images and resized to 28×28 This is the only pre-processing task that was applied in this paper.

The proposed systems consist of six layers: three convolutional layers, two max-pooling layers, and one fully connected layer.

Layer 1 is the first convolutional layer with the ReLU. This is the first layer of the CNN architecture. Basically, this layer accepts and gets the input image of size $n \times n = 28 \times 28$, while the filter (f) size is set to 5×5. The stride (s) is 1, padding (p) is 0, and the number of filters is 32. A feature map is obtained at this level by applying the convolution operation. Using equation 2, after convolution operation is applied, the resulting feature map size is 24×24. It is noted that the ReLU activation function is done on each feature map.

$$fm = \left(\frac{n+2p-f}{s}\right) + 1 \qquad (2)$$

$$fm = \left(\frac{28+0-5}{1}\right) + 1 = 24$$

Layer 2 is the first max-pooling layer which gets its input from layer 1 as the size of 24×24. The pooling size is set to 2×2, the stride is 2, padding is 0. Mainly, max pooling is performed in each feature map independently. After max pooling is operation is applied, the resulting feature map size is 12×12. It is noted that this layer has no activation function.

$$fm = \left(\frac{24+0-2}{2}\right) + 1 = 12$$

Layer 3 is the second convolutional layer with the ReLU, where it gets its input from the previous layer (layer 2) as the size of 12×12. The filter size is set to 5×5. The stride is 1, padding is 0, and the number of filters is 32. After the convolution operation is applied, the resulting feature map size is 8×8. It is that the ReLU activation function is done on each feature map.

$$fm = \left(\frac{12+0-5}{1}\right) + 1 = 8$$

Layer 4 is the second max-pooling layer which gets its input from layer 3 as a size of 8×8. The pooling size is set to 2×2, the stride is 2, padding is 0. Mainly, max pooling is performed in each feature map independently. After max pooling is operation is applied, the resulting feature map size is 4×4. It is noted that this layer has no activation function.

$$fm = \left(\frac{8+0-2}{2}\right) + 1 = 4$$

Layer 5 is the third convolutional layer without ReLU, where it gets its input from the previous layer (layer 4) as the size of 4×4. The filter size is set to 4×4. The stride is 1, padding is 0, and the number of filters is 64. After the convolution operation is applied, the resulting feature map size is 1×1. Basically, the output of this layer is a one-dimensional vector of size 64.

$$fm = \left(\frac{4+0-4}{1}\right) + 1 = 1$$

Layer 6 is the fully connected layer which takes its input from the previous layer (layer 5). The input of this fully connected layer is the one-dimensional vector of size 64 with the ReLU activation function. The output of this layer is a one-dimensional vector of size 256.

Layer 7 is the output layer of the network. This layer computes the score of the classes resulting in a vector of size 10. It is a softmax classifier with ten classes of the ADBase database. The ten output classes are either of {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}.

5. Results and Discussion

As discussed earlier, the ADBase dataset has been applied in evaluating many several systems. To enable the consistent of the performance testing, the ADBase dataset is used in evaluating and testing the proposed system. The proposed system was implemented by using the MATLAB platform on a CPU with 4 GB RAM.

The ADBase database is used to test the proposed system. Training and testing of the proposed system have been done on ten classes of the ADBase database. The training was performed on 60000 images where each class has been trained using 6000 images. The trained network is tested and evaluated on 10000 images where each class has been evaluated on 1000 images. Mainly, the ADBase is stored into two main folders: training and testing respectively. Each of the raining and the testing folder contains 10 folders which represent the ten classes of the Arabic digit. Each class is read and evaluated separately. As mentioned earlier, each image of the ADBase was converted to binary images and resized to 28×28. Table 1 summarizes the proposed system accuracy in terms of testing for each class. It is noted from table 1 that the proposed system gives an average of 98.95% of validation accuracy.

Table 1. The Proposed System Accuracy (R)

	0	1	2	3	4	5	6	7	8	9
R %	98.8	98.5	98.6	98.5	99.7	98.6	99.3	99.1	99.2	99.2

Similar to all other recognition systems, the proposed system has an error rate. The proposed system has a 1% level of error rate due to the variation of writing among writers. The comparison result of the proposed system and other systems is illustrated in table 2.

Table. 2. Performance of recognition rate from different systems using the ADBase dataset

System	Recognition Rate %	Used Classifier
Alkhateeb [1]	95.26	DBN
Alkhateeb et al. [5]	85.26	DBN
Ahamed et al. [6]	99.76	CNN
El-Sawy et al [8]	88	CNN
Proposed System	98.95	CNN

Table 2 shows the work in Ahamed et al. [6] performs the best among all the systems. However, it reported in the work in the work in Ahamed et al. [6] that they had used a developed dataset from the ADBase. Since the proposed system using the entire dataset, the proposed system overcomes the work in Ahamed et al. [6]. Fig 6. Summarizes the confusion matrix of the proposed system.

	0	1	2	3	4	5	6	7	8	9
0	988	0	0	0	0	12	0	0	0	0
1	0	985	0	0	0	0	0	13	0	2
2	0	0	986	14	0	0	0	0	0	0
3	0	0	15	985	0	0	0	0	0	0
4	0	0	3	0	997	0	0	0	0	0
5	14	0	0	0	0	986	0	0	0	0
6	1	0	0	0	0	0	993	0	0	6
7	0	4	0	0	0	5	0	991	0	0
8	6	0	0	0	0	2	0	0	992	0
9	5	0	0	0	0	3	0	0	0	992

Fig 5: The Proposed System Confusion matrix

6. Handwriting Importance and Conclusion

The mechanism of communicating in the modern recent technology has been changed. The writing mechanism is considered an important way of communication. There is a rapid increase in using computers for writing. The handwriting skill is an important factor in many domains such as, education, health care, employment, business, and in everyday life.

Thus handwriting recognition plays an important role in education, health care, and many other fields. Nowadays, there is big shift towards various ways of electronic communication. In addition, modern computers use the handwriting recognition technique in interacting with users. Therefore, handwriting and handwriting recognition still plays an important role for communication

A deep learning approach for handwritten offline Arabic digit recognition has been proposed in this paper. In this paper, the recognition of Arabic digits from raw images concept of the deep learning tool has been utilized. The use of the proposed CNN architecture has eliminated the need for any segmentation from the captured images. The performance of the proposed system has been evaluated on the ADBase resulting in a good validation accuracy of 94.3%. This accuracy can be improved enhancing and tuning the hyper- parameters such as padding and stride. Most of the errors were due to the variation of the writing of similar digits. Fig 6. Summarizes the writing variation of digit zero.

Fig.6: Variation writing of digit zero

References

[1] Jawad H Alkhateeb, "Offline Handwritten Arabic Digit Recognition Using Dynamic Bayesian Network", in proceeding of 1st International Conference on Computing and Information Technology (ICCIT 2012), PP. 176 – 180, 2012.

[2] V L. M. Lorigo and V. Govindaraju, "Offline Arabic handwriting recognition: a survey", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 28, pp. 712-724, 20

[3] M.S. Khorsheed," Off-Line Arabic Character Recognition – A Review", Pattern Analysis & Applications, vol.5, pp. 31-45, 2002.

[4] Sabri Mahmoud, " Recognition of Writer-independent off-line handwritten Arabic (Indian) numerals using hidden Markov models", Signal Processing, vol 88, issue 4, pp. 844-857, 2008.

[5] Jawad H Alkhateeb , and Marwan Alseid, “DBN - based learning for Arabic handwritten digit recognition using DCT features”, in proceedings of the 6th International Conference on Computer Science and Information Technology (CSIT), pp. 222–226, 2014.

[6] Patrick Ahamed et al., “Handwritten Arabic numerals recognition using convolutional neural network”, Journal of Ambient Intelligence and Humanized Computing, March 2020.

[7] A. Ashiquzzaman and A. K. Tushar, "Handwritten Arabic numeral recognition using deep learning neural networks," 2017 IEEE International Conference on Imaging, Vision & Pattern Recognition (icIVPR), pp. 1-4,2017.

[8] El-Sawy A., EL-Bakry H., Loey M. (2017) CNN for Handwritten Arabic Digits Recognition Based on LeNet-5. In: Hassanien A., Shaalan K., Gaber T., Azar A., Tolba M. (eds) Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016. AISI 2016. Advances in Intelligent Systems and Computing, vol 533. Springer.

[9] Ahmad Al-Hmouz et al., "Enhanced Numeral Recognition for Handwritten Multi-language Numerals Using Fuzzy Set-Based Decision Mechanism," International Journal of Machine Learning and Computing, Vol. 10, No. 1, pp. 100 – 107, 2020.

[10] M. Khorsheed, “Recognising handwritten Arabic manuscripts using a single hidden Markov model”, Pattern Recognition Letters, vol. 24, pp. 2235-2242, 2003.

[11] S. Alma’adeed, C. Higgens, and D. Elliman, “Off-line recognition of handwritten Arabic words using multiple hidden Markov models”, Knowledge-Based Systems, vol. 17, pp. 75-79, 2004.

[12] S. Alma’adeed, “Recognition of Off-Line Handwritten Arabic Words Using Neural Network”, proc. Of the Geometric Modeling and Imaging – New Trends, 2006.

[13] Tareq Khan, "Expiry Date Digits Recognition using Deep Learning," 2019 IEEE National Aerospace and Electronics Conference (NAECON), pp. 302-304, 2019.

[14] Akm Ashiquzzaman and Abdul Kawsar Tushar, "Handwritten Arabic numeral recognition using deep learning neural networks," 2017 IEEE International Conference on Imaging, Vision & Pattern Recognition (icIVPR), pp. 1-4, 2017.

[15] Saad bin Ahmed, Saeeda Naz, Muhammad I. Razzak and Rubiyah Yousaf, "Deep learning based isolated Arabic scene character recognition," 2017 1st International Workshop on Arabic Script Analysis and Recognition (ASAR), pp. 46-51, 2017.

[16] Katrina Sundus, Fatima Al-Haj, and Bassam Hammo, "A Deep Learning Approach for Arabic Text Classification," 2019 2nd International Conference on New Trends in Computing Sciences (ICTCS), pp. 1-7, 2019.

[17] George Sakr, Ammar Mhanna, and Rony Demerjian, "Convolution Neural Networks for Arabic Font Recognition," 2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), pp. 128-133, 2019.

[18] R. G. Maidana, J. Monteiro dos Santos, R. L. Granada, A. d. M. Amory and R. C. Barros, "Deep Neural Networks for Handwritten Chinese Character Recognition," 2017 Brazilian Conference on Intelligent Systems (BRACIS), pp. 192-197, 2017.

[19] El-Sherif, E. & Abdleazeem, S. “A Two-Stage System for Arabic Handwritten Digit Recognition Tested on a New Large Database”. International Conference on Artificial Intelligence and Pattern Recognition, AIPR-07., 2007.

[20] Shubham Sanjay Mor et al., “HANDWRITTEN TEXT RECOGNITION: with Deep Learning and Android”, International Journal of Engineering and Advanced Technology (IJEAT), vol. 8, Issue 3S, pp 819-825, 2019.

[21] W. Xu and M. Zhang, "Theory of Generative Deep Learning II: Probe Landscape of Empirical Error via Norm Based Capacity Control," 2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS), pp. 470-474, 2018.

[22] Murtada K. Elbashir and Mohamed E Mustafa. “Convolutional Neural Network Model for Arabic Handwritten Characters Recognition”. International Journal of Advanced Research in Computer and Communication Engineering, Vol. 7, Issue 11, pp. 1-5, 2018.

[23] Jawad H Alkhateeb, "Off-Line Arabic Handwritten Isolated Character Recognition", International Journal of Engineering Science and Technology (IJEST), vol 7, No 7, pp 251 -257, 2015.

[24] Md. Anwar Hossain and Md. Mohon Ali , “Recognition of Handwritten Digit using Convolutional Neural Network (CNN)”. Global Journal of Computer Science and Technology: D Neural & Artificial Intelligence Volume 19 Issue 2 Version 1.0 2019.

[25] Batuhan Balci , Dan Saadati , and Dan Shiferaw , "Handwritten Text Recognition using Deep Learning". CS231n: Convolutional Neural Networks for Visual Recognition, Stanford University, Course Project Report, spring, 2017.

[26] Kavitha B. R. And Sriathi C, " Benchmarking on offline Handwritten Tamil Character Recognition using convolutional neural networks", Journal of King Saud University - Computer and Information Sciences, in press 2019. https://doi.org/10.1016/j.jksuci.2019.06.004.