Journal of Computer Science

Research Article

Improving The Eastern Arabic Hand Written Digits Recognition Using Deep Learning

[1,2]Hamdy Morsy, [1]Hashim M. Hasanien and [1,2]Mostafa M. Abutaleb

[1]Unit of Scientific Research, Applied College, Qassim University, Buraydah, Saudi Arabia
[2]Department of Electronics and Communications Engineering, Helwan University, Cairo, Egypt

Article history
Received: 05-02-2025
Revised: 03-04-2025
Accepted: 08-04-2025

Corresponding Author:
Hamdy Morsy
Unit of Scientific Research, Applied College, Qassim University, Buraydah, Saudi Arabia
Email: h.morsy@qu.edu.sa

Abstract: Convolutional neural networks CNNs or simply deep learning was designed for the challenging of object recognition such as digits, characters and even objects that can be classified. The deep convolutional networks are designed to be trained through using a large dataset of the classes. The degree of complexity of recognizing handwritten digits are different from person to person and even from language to language. The Eastern Arabic handwritten digits recognition system (EAH), which has ten classes, is one of the most challenging object recognitions due to its complexity and similarities of classes. There are many techniques worked on proposed systems for recognizing EAH but with limited accuracy and output cost function. In this paper, a proposed new technique with multi hidden layers is utilized for Eastern Arabic handwritten recognition. This technique proved to have higher output accuracy and performance than other existing systems.

Keywords: Deep Learning, Convolutional Networks, Machine Learning, Object Recognition

Introduction

Techniques for deep learning have become more crucial in the design of pattern recognition systems. In fact, it might be claimed that the recent success of pattern recognition applications is directly related to the accessibility of learning techniques. The important takeaway of deep learning is that by depending more on automated learning and less on manually created heuristics, stronger pattern recognition systems can be produced. The most crucial layers in a neural network are the input and output layers, where there are a certain number of neurons or nodes in each layer and a certain number of neurons in the output layer that correspond to the number of classes in this pattern (Ramzan *et al.*, 2018).

Deep convolutional networks have hidden layers in between input layer and output layer which add some complexity to the system. Simply put, to achieve the highest output classifications, the parameters in the system have to be maximized so that a system can be trained and tested for the input object. These variant parameters are changing such that maximizing the output classification accuracy (O'Shea and Nash, 2015). The minimal dataset categorization comprises of handwritten digits with ten categorized objects. One of the most widespread numeral systems is the eastern Arabic digit system, which utilizes digits from 0 = (٠) to 9 = (٩). The digit (٠) is very similar to English full stop and the digit (٩) is similar to English digit 9. The eastern Arabic digits are {٠,١,٢,٣,٤,٥,٦,٧,٨,٩} which are changed with handwritten typing from male to female and from young to old people (Aljarrah *et al.*, 2021). To correspond to distinct writing systems, there are around 30 different numeric systems. Some of these writing systems have various forms, such as the three-form Chinese numeral system.

The Arabic language has two varieties: the western Arabic numeral system, which is primarily used by western nations, and the eastern Arabic numeral system, which is used by the majority of Arabic nations. There are more than 453 million people speak Arabic language with many more countries using the eastern Arabic digits (Kualo, 2025). This brought up the need for handwritten digits recognition specially for electronic applications.

The system processes datasets using an input layer, an output layer, and one or more hidden layers. Each layer has number of nodes which represents the number of pixels at the input and number of classes at the output. For example, for 28x28 image size, there is 784 pixels equivalent to 784 nodes and at the output there is ten classes for the number of digits to be identified (Islam *et al.*, 2017). In each layer, the nodes are connected to previous and subsequent nodes of the layers. The nodes of input layer are connected to every node in the next hidden layer. Each connection has a weight (w) and a bias (b) that is added to the sum of all incoming connections. For the weights and biases, initial random values are employed as provided by the system (Dumoulin and Visin, 2018; Abiodun *et al.*, 2019). Each node in a layer will be the sum of the previous nodes multiplied by the respective weights and a bias. At the starting of the system training, there are initial values for all weights and biases assigned to the input layer nodes and biases. These values feed forward from the input layer to output layer. Then, the output will be tested against the correct class of the output and the error is measured. Afterwards, the output is back propagated to the input to optimize the weights and biases of the system. Minimizing the output error or the cost function will require a derivative function for each node in the system. An activation function is used at each node to limit the fluctuation of the values for each node and in addition to optimize the weights and biases (Richard & Xia, 2018).

Literature Review

The structure of deep learning convolutional neural network (DNN) consists of three main categories the input layer, one or mor hidden layer and the output layer. The main idea of DNN is to train the system for some datasets with each input object such an image contains some features. The function of the system is to extract these features so that the system can classify any object with the same features. The value at each node is the values of the previous nodes multiplied by weights for each node and a bias. Figure (1) shows the input to a node which equals to $x = wa + b$. The node formula is then enhanced using activation function which is a nonlinear multi derivative function. Equations (1-3) show the activation functions $sigmoid\,(x) = \sigma\,(x)\,, \tanh\,(x)\, andReLU\,(x)$ functions respectively, and x is the value of the node before applying the activation function:

$$\sigma\,(x) = \frac{1}{1+e^{-x}} \tag{2}$$

$$tanh\,(x) = \frac{\left(e^{x}-e^{-x}\right)}{\left(e^{x}+e^{-x}\right)} \tag{2}$$

$$ReLU\,(x) = \max\,(0.1x, x) \tag{3}$$

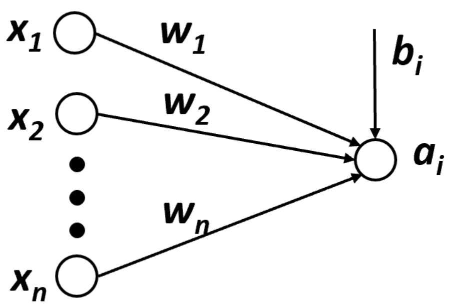

Fig. 1: Simple node connections

The derivative of sigmoid function for example is $\sigma(x)(1 - \sigma(x))$ where $\sigma\,(x)$ is the sigmoid function. The result function is another derivative function which can help in reaching the optimum values for weights and biases. The same results can be obtained from finding the derivative of tanh, ReLU or even leaky ReLU (Vialatte *et al.*, 2017; Wiatowski and Bolcskei, 2018). The value at each node after utilizing the activation function is local receptive field in addition to the bias which has an effective factor for optimizing the output accuracies as shown in Fig. (2). Lecun is a research who with his colleagues have introduced the convolutional neural networks concept (LeNet-5 algorithm) in Lecun *et al.* (1998). His system structure has 5 layers: two convolutional layers with filter size 5x5 with stride equals to 1 and 3 fully connected layers. The first convolutional layer has 6 feature maps or channels the second convolutional layer has 16 channels. For each convolutional layer, there is a pooling layer which minimize the node size to half and the pooling layer is 2x2 filter with stride equals to 2. The 2x2 pooling layer may be calculated as the average pooling or the maximum pooling (Albawi *et al.*, 2017; Calix, 2020; Czum, 2020).

The eastern Arabic handwritten digits recognition EAH is method utilizing deep neural networks DNN to train datasets to identify digits. The eastern Arabic digits are different from other numeral systems in such a way that the digit may have different patterns for various individuals. Figure (3) shows the EAH digits with some error where the upper right corner is the right digit and the lower right corner is the wrong digit (Morsy, 2020a-b; 2022). With handwritten digits recognitions, some digits look similar such as ٣ and ٦ ,٢ and ٩, and ١ and ٦. There are many handwritten digits datasets on the web and others are prepared manually.

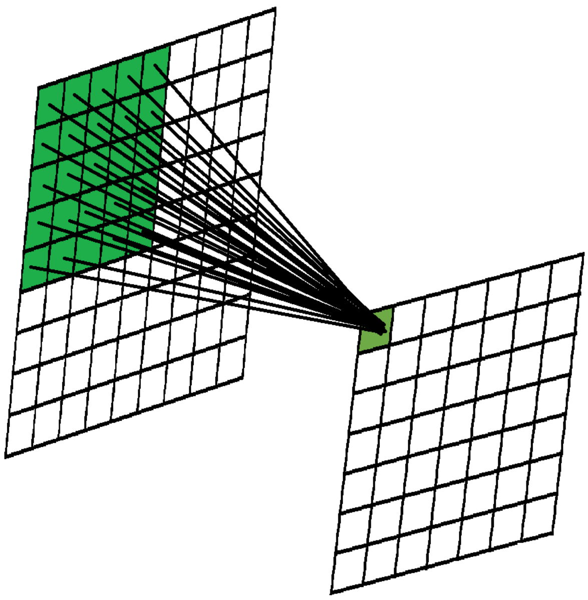

Fig. 2: Local receptive field in a neural network

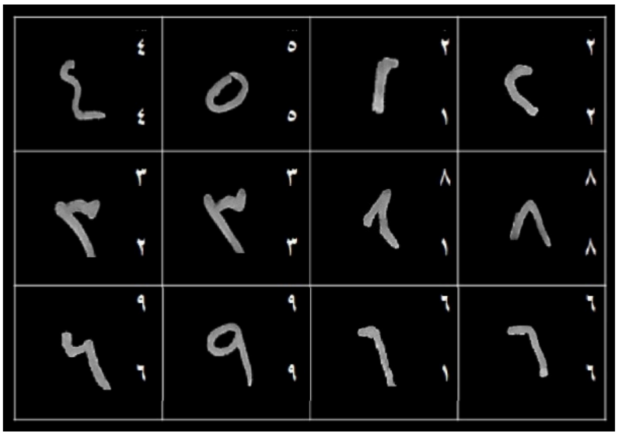

Fig. 3: Example of output EAH digits

The standard input image is 28x28 which provides 784 nodes for the input to the next hidden layer. The nodes for the next hidden layer is utilizing an activation function with filters for most neural network systems is 5x5 and a stride equals to one. The second hidden layer in most neural network techniques is a pooling layer with average or maximum pooling function and a filter of 2x2 with a stride equals to 2.

El Sawy and his colleagues proposed a handwritten digits recognition based on LeNet-5 (EL-Sawy *et al.*, 2017). Their system was based on 8 layers (as they defined): input layer, output layer, two convolutional layers, two sub-sampling layers for automatic feature extraction and two fully connected layers as multi-layer perceptron hidden layers for nonlinear classification. The layers symbols are input layer, C1, S2, C3, S4, C5, F6 and output layer with the same kernel size, strides and pooling layers as in LeNet-5. An input image of size is 32x32 with one channel is utilized. They performed their calculations on MAD Base dataset which has 70,000 images written by 700 different people (Abdleazeem and El-Sherif, 2008). The result for this method is 1% for training and 12% for testing.

Mahmoud's system was based on Gabor filter extraction features and private database of 21120 images for 70% for training and 30% for testing (Mahmoud, 2008). His results were 0.15% error for training and 2.16% error for testing. Takruri *et al.* used Fuzzy C-Means (FCM) based classifier for the lower level and Support Vector Machine (SVM) for the second level. They utilized a public database of 3510 images with 60% for training and 40% for testing. With this dataset which is considered to be small, they obtained a result of 12% error for testing (Takruri *et al.*, 2014). AlKhateeb *et al.* applied AD Base dataset which utilized 60,000 training images and 10,000 testing images (Alkhateeb and Alseid, 2014). They utilized Discrete Cosine Transform (DCT) coefficient approach for feature extraction. Their results were 14.74% error for testing datasets. Hafiz and Bhat employ K-Nearest Neighbor and Support Vector machine approaches for digits recognition (Hafiz and Bhat, 2015). Their technique achieved a testing error of 5%.

We have selected MAD Base as the dataset for comparison over other types of datasets. Since the MAD Base is a modified version of AD Base which resized and normalized to match the format of MNIST dataset. The MAD Base is specific for Arabic Handwritten digits but AHCD is for Arabic characters and also has larger datasets (70,000 for MAD Base and 16800 characters for AHCD). MAD Base's alignment with MNIST's format allows it to serve as a standardized benchmark for Arabic digit recognition. It is also chosen over HODA dataset, because the second is designed for Persian digits. MAD Base's dataset reflects Arabic writing more accurately (Ali and Abdulrazzaq, 2024).

In the previous techniques, the dataset employed was different in size from technique to another which affects the performance of the system and accuracy and may result in overfitting problems in some of them. The testing error rates is also considered to be high and can cause significant problems when dealing with digits recognition in transactions.

We adopt here in this study LeNet-5 algorithm, since it is naturally fit to our requirements for 32x32 grayscale digit images and it is efficiently captures local features such as edges and curves in digits with simple algorithm design. Modern CNN such as ResNet or EfficientNet are designed for deeper feature extraction in high resolution, multi-channel (RGB) images (e.g., ImageNet). LeNet-5 is lightweight compared to ResNet (millions of parameters), MobileNet (optimized but still heavier), or EfficientNet (scaled for efficiency but resource intensive). Transformers (ViTs) and CNN-RNN hybrids require significantly more memory and processing power due to attention mechanisms or recurrent layers, which are unnecessary for static, small scale digit classification (Zhang, 2024).

In the next section, we will introduce our new technique which is based on a modified version of LeNet-5 algorithm.

Materials and Methods

The method in this technique depends mainly on a modified version of LeNet-5 algorithm. The input image is 32x32 grayscale image. The dataset is MAD Base dataset with 60,000 images for training and 10,000 for testing which consists of handwritten digits from 700 individuals of different ages and genders. The image is adjusted for 28x28 size and the digits are centred in the image. We applied some of the most common activation functions such as sigmoid, tanh, ReLU, and Leaky ReLU activation functions and also we employed both maximum pooling and average pooling. The learning rate is 0.0001, the batch size is 64 and the number of epochs is 10. Different modified version of LeNet-5 is employed to find the optimum system structure for Eastern Arabic digits recognition. This includes the original system with different activation functions and different pooling values. In addition to removing testing the performance of the system with and without removing one of the last two fully connected layers. We kept the number of features in each layer fixed to have a complete overview and comparison between our suggested system and the state-of-the-art LeNet-5 system structure.

Results

We utilized python code to implement the LeNet-5 algorithm with ability to modify the hyper parameters. The hyper parameters applied in this system are: learning rate is 0.0001, batch size is 64, number of epochs is 10. The system used here is input layer is 784 neurons, C1 is a convolutional layer with 6 feature maps and size of 28x28, S2 is a sub-sampling layer with 6 feature maps and size of 14x14, C3 is a convolutional layer with 16 feature maps and size of 10x10, S4 is a sub-sampling layer with 16 feature maps and size of 5x5, C5 is a convolutional layer with 120 neurons, F6 is a fully connected layer with 84 neurons and the last layer is 10 classes. As shown in Table (1), the system has employed different activation functions with different pooling functions. The best performance is 98.96% which obtained with Leaky ReLU activation function and average pooling function.

Table 1: Activations functions and Pooling functions

	ReLU	Sigmoid	Tanh	Leaky ReLU
Average Pooling	98.88%	97.45%	98.91%	98.97%
Maximum Pooling	98.84%	97.14%	98.16%	98.96%

Figure (4) shows the output performance of the LeNet-5 algorithm and the two modified versions for using 120 neurons and 84 neurons layers used as convolutional layers with maximum pooling at the layers S2 and S4. From the figure, it can be observed that the system with 84 neurons has better performance than the other two systems. As shown in Figure (5) the output performance of LeNet-5 algorithm and the system with 120 neurons layer only and the system with 84 neurons layer only which is used as convolutional layer. The system with 84 neurons convolutional layer proved to have better performance compared to other systems. Table (2) shows a comparison between systems with employing maximum pooling and average pooling with respect to testing error. As can be observed, the minimum testing error is obtained with 84 neurons convolutional layer with average pooling using as sub-sampling.

Table 2: The output testing error for the three systems with maximum and average pooling

	LeNet-5 Algorithm	With 120 neurons layer (F6 removed)	With 84 neurons layer (F6 removed and C5 replaced with 84 neurons)
Maximum Pooling	1.084%	1.076%	1.059%
Average Pooling	1.045%	1.028%	0.999%

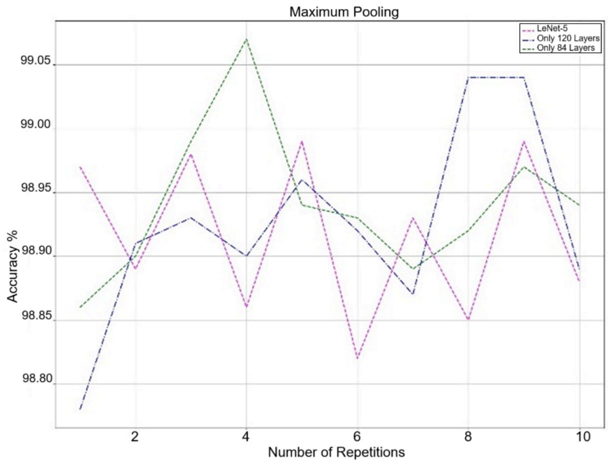

Fig. 4: Performance of systems LeNet-5 algorithm, system with 120 neurons layer only, system with 84 neurons layer only with maximum pooling

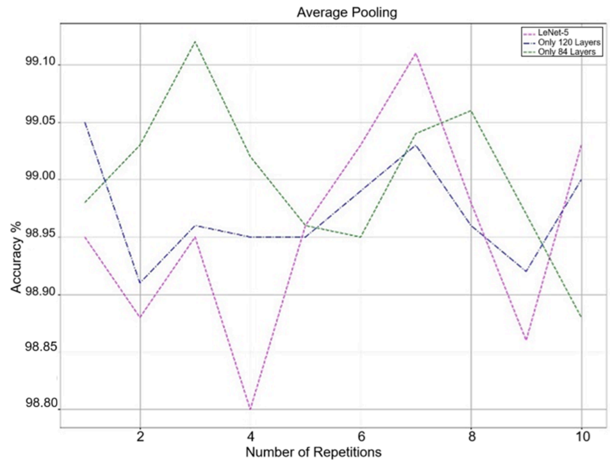

Fig. 5: Performance of systems: LeNet-5 algorithm, system with 120 neurons Layer only, system with 84 neurons Layer only with average pooling

Discussion

The modified system with properties: input image of size 32x32 and on channel (gray scale), C1 is a convolutional layer of 6 feature maps and size of 28x28, S2 is a sub-sampling layer with 6 feature maps and size of 14x14, C3 is a convolutional layer with 16 feature maps and size of 10x10, S4 is a sub-sampling layer with 16 feature maps and size of 5x5, C5 is a convolutional layer with one feature map and size of 84 neurons as shown in Figure (6). The hyper parameters of this system are: batch size is 64, learning rate is 0.0001 and number of epochs is 10. The system with given structure proved to have higher performance and lower testing error rate compared to other systems with different structures. The changing from LeNet-5 algorithm to other simpler versions proved to have better performance and faster than the original system. The activation functions employed in this study to find the maximum output performance and minimum testing error is leaky ReLU activation function.

DOI: 10.3844/jcssp.2025.1293.1298

Conclusion and Future Work

The deep convolutional neural networks become the state of the art for object recognition with utilizing machines for doing the hard work. The automated system is handling and manipulating the feature maps and minimizing the processing time of pattern recognition. The DNN networks with some network modification can prove to have higher performance and lower error rates compared to other variation of the system. The system structure may differ from classification to another depending on the type of data to be classified. The system suggested in this study with simple structure has been fit with high accuracy for eastern Arabic handwritten recognition EAH. The need for future work is necessary to handle some other types of handwritten digits recognition for input methods such as touch screen and electronic pens also handling inputs with added noise can be a trigger for future work.

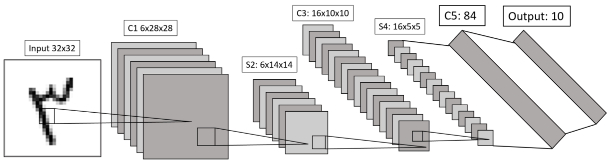

Fig. 6: The complete deep convolutional neural network

Acknowledgment

The Researchers would like to thank the Deanship of Graduate Studies and Scientific Research at Qassim University for financial support (QU-APC-2025).

Funding Information

This research received no external funding.

Author's Contributions

All authors made a significant contribution to the preparation, development and publication of this manuscript.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

Conflicts of Interest

The author has no conflicts of interest to declare.

References

Abdleazeem, S., & El-Sherif, E. (2008). Arabic Handwritten Digit Recognition. *International Journal of Document Analysis and Recognition (IJDAR)*, *11*(3), 127-141. https://doi.org/10.1007/s10032-008-0073-5

Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Umar, A. M., Linus, O. U., Arshad, H., Kazaure, A. A., Gana, U., & Kiru, M. U. (2019). Comprehensive Review of Artificial Neural Network Applications to Pattern Recognition. *IEEE Access*, *7*, 158820-158846. https://doi.org/10.1109/access.2019.2945545

Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a Convolutional Neural Network. *2017 International Conference on Engineering and Technology (ICET)*, 1-6. https://doi.org/10.1109/icengtechnol.2017.8308186

Ali, S. Hasen, & Abdulrazzaq, M. B. (2024). KurdSet Handwritten Digits Recognition Based on Different Convolutional Neural Networks Models. *TEM Journal*, *13*(1), 221-233. https://doi.org/10.18421/tem131-23

AlJarrah, M. N., Zyout, M. M., & Duwairi, R. (2021). Arabic Handwritten Characters Recognition Using Convolutional Neural Network. *2021 12th International Conference on Information and Communication Systems (ICICS)*, 182-188. https://doi.org/10.1109/icics52457.2021.9464596

AlKhateeb, J. H., & Alseid, M. (2014). DBN - Based Learning for Arabic Handwritten Digit Recognition Using Dct Features. *2014 6th International Conference on Computer Science and Information Technology (CSIT)*, 222-226. https://doi.org/10.1109/csit.2014.6806004

Calix, Ricardo A. (2020). *Deep Learning Algorithms: Transformers, Gans, Encoders, Cnns, Rnns and More*.

Czum, J. M. (2020). Dive Into Deep Learning. *Journal of the American College of Radiology*, *17*(5), 637-638. https://doi.org/10.1016/j.jacr.2020.02.005

Dumoulin, V., & Visin, F. (2018). A Guide to Convolution Arithmetic for Deep Learning. *ArXiv Preprint ArXiv:1603.07285*, *2*. https://doi.org/10.48550/arXiv.1603.07285

El-Sawy, A., EL-Bakry, H., & Loey, M. (2017). CNN for Handwritten Arabic Digits Recognition Based on LeNet-5. *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016*, *533*, 566-575. https://doi.org/10.1007/978-3-319-48308-5_54

Hafiz, A. M., & Bhat, G. M. (2015). Boosting OCR for Some Important Mutations. *2015 Second International Conference on Advances in Computing and Communication Engineering*, 128-132. https://doi.org/10.1109/icacce.2015.120

Islam, K. T., Mujtaba, G., Raj, R. G., & Nweke, H. F. (2017). Handwritten Digits Recognition with Artificial Neural Network. *2017 International Conference on Engineering Technology and Technopreneurship (ICE2T)*, 1-4. https://doi.org/10.1109/ice2t.2017.8215993

Kualo. (2025). Numerals in Various Writing Systems. *Omniglot*. https://omniglot.com/language/numerals.htm

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, *86*(11), 2278-2324. https://doi.org/10.1109/5.726791

Mahmoud, S. A. (2008). Arabic (Indian) Handwritten Digits Recognition Using Gabor-Based Features. *2008 International Conference on Innovations in Information Technology*, 683-687. https://doi.org/10.1109/innovations.2008.4781779

Morsy, H. A. (2020a). Developing a New CNN Technique for Arabic Handwritten Digits Recognition. *International Journal of Recent Technology and Engineering (IJRTE)*, *9*(3), 520-524. https://doi.org/10.35940/ijrte.c4588.099320

Morsy, H. A. (2020b). Optimization of Arabic Handwritten Digits Recognition Using CNN. *International Journal of Scientific and Engineering Research*, *11*(11), 372-376.

Morsy, H. A. (2022). Performance Analyses of the Eastern Arabic Hand Written Digits Recognition Using Deep Learning. *American Journal of Science, Engineering and Technology*, *7*(3), 57-61. https://doi.org/10.11648/j.ajset.20220703.11

O'Shea, K., & Nash, R. (2015). An Introduction to Convolutional Neural Networks. *ArXiv Preprint ArXiv:1511.08458*, *2*, 1-11. https://doi.org/10.48550/arXiv.1511.08458

Ramzan, M., Ullah, H., Ilyas, M., Akhtar, W., Mehmood, S., Mahmood, A., & Zamir, A. (2018). A Survey on Using Neural Network Based Algorithms for Hand Written Digit Recognition. *International Journal of Advanced Computer Science and Applications*, *9*(9), 519-528. https://doi.org/10.14569/ijacsa.2018.090965

Richard, E. N., & Xia, J. (2018). Neural Networks and Deep Learning. *Artificial Intelligence*.

Takruri, M., Al-Hmouz, R., & Al-Hmouz, A. (2014). A Three-Level Classifier: Fuzzy C Means, Support Vector Machine and Unique Pixels for Arabic Handwritten Digits. *2014 World Symposium on Computer Applications & Research (WSCAR)*, 1-5. https://doi.org/10.1109/wscar.2014.6916798

Vialatte, J.-C., Gripon, V., & Coppin, G. (2017). Learning Local Receptive Fields and Their Weight Sharing Scheme on Graphs. *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 623-627. https://doi.org/10.1109/globalsip.2017.8309034

Wiatowski, T., & Bolcskei, H. (2018). A Mathematical Theory of Deep Convolutional Neural Networks for Feature Extraction. *IEEE Transactions on Information Theory*, *64*(3), 1845-1866. https://doi.org/10.1109/tit.2017.2776228

Zhang, X. (2024). Handwritten Digit Recognition Based on Deep Learning Techniques. *Applied and Computational Engineering*, *36*(1), 7-14. https://doi.org/10.54254/2755-2721/36/20230416