www.ijecs.in
International Journal of Engineering and Computer Science
Volume 14 Issue 04 April 2025, Page No. 27030- 27039
ISSN: 2319-7242 DOI: 10.18535/ijecs/v14i04.5060

Letter Recognition from Noisy Images Using Deep Learning

Esraa Farouk Abou-Elmagd [1]

[1] Faculty of Commerce South Valley University Qena, Egypt

Abstract

In many fields, both machine learning and deep learning are being used. One of the most active and difficult areas of research is letter recognition. Recently, deep learning and letter recognition have drawn the attention of many researchers. Letter recognition is a problem that has been worked on in many languages. Letter recognition refers to the process of identifying and distinguishing letters based on their visual features. It involves analyzing patterns of responses and detecting specific components that form different letters. This perceptual activity plays a crucial role in recognizing individual letters and is often studied using rapid visual displays and visual search tasks. Additionally, letter recognition in context explores how letters are processed within words and contributes to our understanding of reading processes. In the process of recognition, pre-processing techniques increase image quality by reducing noise and correcting orientation, while convolutional neural networks extract letter features. However, the existing letter recognition system faces many challenges in extracting text from noisy and distorted images or complex layouts, and extraction is mostly limited to numbers and the English alphabet. So, several studies using deep learning models have been conducted to achieve better accuracy. The proposed technique achieves 96.2% accuracy in recognizing letters from input images.

Keywords: Deep Learning, Convolutional Neural Network (CNN), Letter Recognition, Noisy images.

1. Introduction

Pattern recognition is a major research field that comprises the process of identifying and recognizing different means of input, such as images. Letter, face, and speech recognition are common variants of pattern recognition. One of the pattern recognition steps involves the study of the object, identifying attributes (features), and extracting the difference in other objects to determine matching or mismatching. Letter recognition is an area of pattern recognition that has become the subject of research during the last few decades. The computer can identify and interpret letters automatically [1]. A computer or device can take letters from images as input and then interpret this as text.

Letter recognition is a process by which letters can be recognized in any language. It is achieved through important steps of feature extraction and classification [2]. The recognition system involves four major steps, especially data collection, pre-processing, feature extraction, and classification: -

1. Data Collection: For data acquisition many researchers use standard data sets available globally and some of research uses own data sets for processing.
2. Pre-processing: In pre-processing phase many mathematical and morphological operations are applied on input letter image for gay scale conversion, normalization, finalization, baseline detection, skew correction, slant detection, slant correction, noise removal etc.
3. Feature Extraction: The letter is taken for further processes, features. Then, the features extracted are compared with the training image features. If the matching

class is presented, then that image is recognized.

4. Classification: for classification process classifiers are used like support Vector Machine (SVM), Convolutional Neural Network (CNN) etc.

Letter recognition is difficult for many reasons. The reason is that various individuals have various styles of writing [3]. The best reason is that there are numerous letters, like capital letters, small letters, and special symbols. As a result, a huge dataset is needed to prepare a close-to-exact neural network model. The many issues in creating a letter recognition model are due to noise and distortion in the image, making it difficult to recognize the text part. Letter recognition is a pivotal issue in machine learning. Methodologies and techniques have been proposed, but it is still an unresolved issue. Using a neural network, it is done, but it has some problems with it, so one can use deep learning over it for better results.

Letter recognition was implemented using many machines learning and deep learning techniques [1]. It is really a challenging issue to develop a practical letter recognition system that can maintain high recognition accuracy.

Deep Learning (DL) is a new application of machine learning for learning representation of data. The DL algorithms have occupied a high position in the field of object recognition due to the great improvement in the performance they have provided [4]. There are several fields in which deep learning is being utilized. Letter recognition is an important area where deep neural networks are being used.

Significant advancements in the recognition process were achieved through the integration of a deep learning (DL) approach [5]. DL is used in detection, classification, and learning. DL architecture is based on automatic learning from the features without prior determination (extraction). DL uses many layers to get the final information from the raw data. Deep learning is a powerful feature extraction method applied to extract the features of letters [2]. Deep learning provides the task with a specific method, which inherits features from machine learning methods based on learning data representation.

Deep learning techniques are widely used to improve their efficiency without a percentage loss of accuracy or an increase in hardware cost in AI systems. The performance of DL comes from raw data after learning over a large amount and processing it for extracting features [6]. Recently, deep learning has drawn the attention of many researchers due to their capability of solving computer vision problems such as object detection, classification, and recognition.

Deep learning is one method that can be used to model letter recognition. Deep learning is the development of machine learning in which the machine automatically extracts features. CNN is one of the algorithms that is often used in deep learning [7]. In addition, CNN is an advanced neural network with significant applications in computer vision. Convolutional Neural Networks (CNNs) are a type of neural network that are applied in many fields and provide efficient solutions to many problems, where there is some translation invariance, like some applications of object recognition and speech recognition.

CNN is widely used for analysis, segment and visual image classification [1]. CNN is easier, faster to train, and has fewer parameters. After training, CNNs converge and generate acceptable results and accurate decisions. The performance of CNNs depends on processing large volumes of raw data to extract relevant features [3]. Although CNNs offer high accuracy, they require significant complexity.

2. Literature Review

Humans' ability to recognize and produce letters has been studied as a psychological phenomenon for nearly a century [8]. Early efforts sought to understand the rapid recognition of letters among machine-generated letter shapes. In such conditions, the shapes and sizes of the letters were standardized. However, human beings are also adroit in recognizing letters which are distorted variations of what typically appears in print [9]. Explain how the variants in the letters are accommodated in human perception that has yielded some theories about this human business, including the theories that support template matching and others that target feature analysis [10]. Feature detection theories propose that letter recognition includes the detection of special features [11]. In 1959, Selfridge proposed the Pandemonium Model of letter perception, contending that letters are recognized via their issue features. Visual complexity is proportional to the number of features [12], meaning that the time taken to recognize a letter is longer for more complex combinations of features. In addition to letter naming, another potential component of letter recognition is the ability to distinguish critical features of letters. Awareness of critical features may precede letter identification and letter-forming speed. Gibson et al. (1962) studied critical features of letters [13]. They found that difficulty in discrimination among features (e.g. horizontal lines, closed circles, etc.) varied for different transformations of letters (letters that are changed by orientation or addition of features). The study of critical features of letters has been applied not only to recognizing but also to producing letters. Pelli and colleagues identified seven distinctive features existing in letters [14] that were unique regarding complexity, overlap, height, width, area, and efficiency, which participants could readily detect. K. Swetha et al. (2021) studied the uses and main challenges of implementing a model [15] that converts images of letters into digital format. Ritik Dixit et al. (2021) studied and worked with different algorithms for letter recognition [16], including support vector machines (SVM), multi-layer perceptron (MLPs), and 17 convolutional neural networks (CNNs). They found that SVM had the highest accuracy on the training dataset, but CNN achieved the highest accuracy on the testing dataset. SVM also had the shortest execution time, while CNN had the longest execution time. Fathma Siddique et al. (2020) studied the accuracy of various hidden layers and the number of epochs of CNN for better accuracy [17] in building a model for letter recognition using the MNIST dataset. Ashish Shetty and Sanjeev Sharma (2023) have clearly explained the use of optical character recognition in different domains [18]. He tells the importance of convolutional neural networks on pattern recognition tasks. They mentioned the process involved in the letter recognition tasks. In reference [20], the paper explores enhancing offline letter recognition by combining "Hidden Markov Models" and "Artificial Neural Networks" in a hybrid model. [21] The paper explores the field of computer vision, focusing on letter recognition using neural networks, to efficiently transform handwritten text into digital format. [22], The paper focuses on letter recognition using artificial neural networks, detailing a process that includes pre-processing, data augmentation, convolutional and recurrent neural networks, and post-processing to improve accuracy, making handwritten text digitally accessible for various applications. [23], The paper focuses on letter recognition using "Multi-Layer Perceptrons (MLP)", and Deep Convolutional Networks (CNN) to compare their performance. [24] This paper extended the work of the previous paper and introduced the first successful application of CNNs to letter recognition. The authors achieved state-of-the-art results on several benchmark datasets. In [25], around 90000 images of more than 40 different classes of letters of the Devanagari script were segmented from the image. Used deep learning architecture for recognition and CNN for superior results to traditional shallow networks in many recognition tasks and focused on the use of the Idler and dataset increment approach to improve test accuracy. In

[26], they describe the same problem of letter recognition. They used a holistic approach to identify the handwritten words; each word taken is an individual entity, so a holistic approach is better and uses such methods as density features, long run features, and structural features for extraction in the handwritten document image. After that, they apply for classification by using Support Vector Machines (SVM). In [27], they propose a workflow and a machine learning model for recognizing letters on form images. It is based on CNN as a powerful feature extraction and Support Vector Machines (SVM) as a high-end classifier. Based on the experiment results using data, both for training and testing, the proposed method achieves an accuracy rate better than only the CNN method. The proposed method was also validated using ten-fold cross-validation, and it shows that the recognition rate for this proposed method is still able to be improved. Researchers face many challenges in the field of letter recognition, and as computational technology increases and resource limitations decrease [28], new methods need to emerge:

1. Letter Datasets: Some of the available datasets do not have many records. There is a need for a dataset with a large amount of data with different font sizes, styles, illuminations, users, and words.
2. Noise Manipulation: Images that contain unrequired data noise should be eliminated automatically or manually, or it may lead to misclassification. Some of the work depended on removing them manually. Especially for large data datasets, there is a need to find a simple and fast way to automatically remove them
3. Low-Quality Documents: Ancient papers and documents are a great challenge as they contain unrequired factors such as noise, and some of the characters are removed. It is an open area problem for researchers to work on.
4. Dataset Augmentation: Data augmentation can help to reduce overfitting and increase overall performance.
5. Overfitting manipulation: Overfitting is a challenging problem. Many solutions, such as data augmentation and regularization, can be used.

In 2018, research by [7] tried applying a deep neural network (DNN) for letter classification in the EMNIST Letters dataset. The stage begins with preprocessing the input image, which includes image thresholding, character thinning using morphological operations, slant correction, and image segmentation. After that, the images that have passed through the preprocessing process enter the feature extraction and classifier stages, using the DNN model. The DNN model makes use of a stacked autoencoder to train its many layers. It has three hidden layers: two hidden layers and one SoftMax layer on top of them, each with 300, 50, and 27 neurons. The results of this study achieved an accuracy of 88.8%.

In 2019, the study in [7] used the text cap model (capsule network). In this model, the input image is processed in three layers of folding layers, that is, capsule or drawing capsule layers, which are fully connected to the primary capsule layer. Dynamic routing with three routing iterations connects the primary capsule and character capsule. Based on the results of testing this model on the EMNIST Letters dataset, training with the full train dataset gave an accuracy of 95.36%, and training with 200 data samples for each class gave an accuracy of 92.79%.

Research by [7] tested the deep convolutional neural networks (DCNN) model, which has autonomous and continuous learning (ACL) capabilities so that it can automatically generate DCNN architecture for a specific vision task. [7] tested the deep convolutional neural networks (DCNN) model, which has independent and continuous learning (ACL) capabilities to generate a DCNN architecture automatically for a given vision task. The optimizer used is RMSprop. From the DCNN model formed, an accuracy of 95.58% is obtained in research using VGG-5 with a fully connected spinal cord [7]. In the experiment, the best model obtained using the EMNIST Letters dataset obtained an accuracy of 95.88%.

3. Background

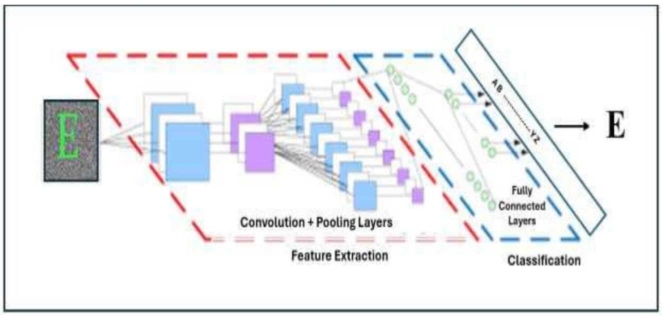
Image processing is a method to perform some operations on an image, to get an enhanced image or to extract some useful information from it [1]. It is a type of signal processing in which the input is an image, and the output may be an image or characteristics or features associated with that image.

3.1 The architecture of recognition process

1. Pre-processing: It is the initial stage in the recognition process, where the data is prepared for further algorithms [1]. It includes tasks like removing unnecessary distortion, fine-tuning images, slant correction, noise removal, and basic alignment. Pre-processed data is then passed to the next module to extract the features.
2. Feature extraction: After pre-processing, the feature extraction step involves extracting sets of features that will be used for classification [1]. Feature extraction is essential for deriving meaningful properties from the dataset, which will be used to classify letters or images.
3. Classification: It involves assigning labels or categories to images, enabling them to be categorized and trained accordingly.
4. Recognition: The final step is recognition, where preprocessed and feature-extracted images are passed through the trained model to identify and recognize the letters. The model is trained on a training dataset, and the recognized letters can be predicted based on its output.

3.2 Convolutional neural network

It can convert the input structure through each layer of the network to automatically extract the features of the images [29].



3.3 CNN Architecture

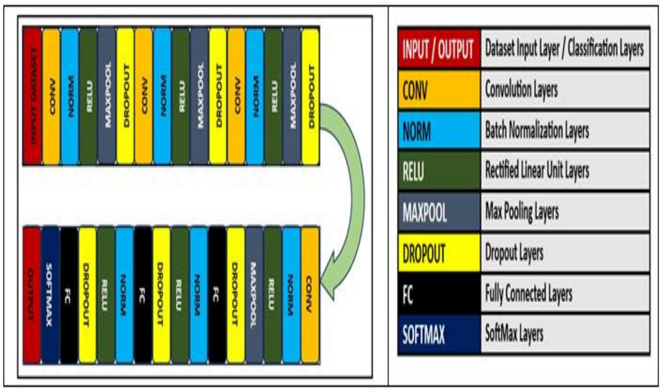
1. Input Layer: It is a buffer to hold the input and pass it on to the next layer [30].
2. Convolution layer: It performs the crucial task of feature extraction. It applies to the convolution operation on the input data using a kernel [28]. By sliding the kernel over the input and performing the sum of products at every location, multiple convolution operations are performed, resulting in different feature maps. The stride determines the size of the steps taken by the kernel as it moves.
3. Batch Normalization layer: It is added just before the nonlinearity and especially after the convolutional layers to limit its output away from the region of saturation using the mean and variance.
4. RELU Layer: It is the most used activation function in deep learning models. In this layer, the Rectified Linear Unit (ReLU) is an element-wise operation (applied per pixel) to introduce nonlinearity in a network. This operation converts each negative pixel into a feature map to zero and keeps each positive pixel. It can speed up the learning process [31]. The output of each convolution layer is passed through the ReLU activation function.
5. Pooling Layer: It reduces the spatial size of each feature map, which in turn reduces the computation in the network. Pooling also uses a sliding window that moves in stride across the feature map and transforms it into representative values. Min pooling, average pooling, and maxpooling are commonly utilized.

6. Dropout layer: It is also used in convolutional neural networks to reduce overfitting [30]. This layer “drops out” a random set of neurons in that layer by setting their activation to zero.
7. Fully connected layer: It connects every neuron in the layer to all the neurons in the previous layer. It learns the non-linear combination of the features and is used for classifying or predicting the output [32]. For classification problems, the fully connected layer is generally followed by a SoftMax layer; it produces the probability of each class for the given input. For regression tasks, a regression layer is added afterward to predict the output.
8. SoftMax Layer: It is frequently appended to the last layer of an image classification network, such as those in CNN. SoftMax’s input is the output of the fully connected layer immediately preceding it, and it outputs the final output of the entire neural network [1]. This output is a probability distribution of all the label class candidates. SoftMax activation function to output probabilities between 0 and 1 for each class, representing the confidence that a certain character belongs to a specific class.
9. Classification Layer: It computes the cross-entropy loss for classification and weighted classification tasks with mutually exclusive classes.

4. Proposed Method

Deep Learning (DL) is a nonlinear computational method for learning information from data and predicting complicated trends, regardless of the distributions to which errors are subjected or how complex the relationships hidden in the data are. It has been successfully applied to numerous areas such as speech recognition, human face recognition, letter recognition, handwriting recognition, crop yield prediction, crop type classification, weather forecasting, environmental monitoring, and image fusion. However, neural networks tend to get trapped in local extreme values during the training. Convolutional neural networks are part of deep learning. They deal with creating, optimizing, and implementing algorithms that enable automation of technology in such a way that it can operate perfectly without any human intervention. CNNs are well-suited for image classification tasks. CNN can extract handy high-label features that are hard for a human to understand. Convolutional Neural Networks (CNNs) learn to extract features in images and use these features to classify the images into various categories. A Convolutional Neural Network (CNN) is used to recognize the alphabet from an input image.

4.1 Proposed Method Architecture



A convolution layer consists of multiple filters that perform the convolution operation. Every image is considered a matrix of pixel values. By using batch normalization, each feature map will have a single mean and standard deviation used on all the features it contains. Once the feature maps are extracted, the next step is to move them to a ReLU layer. ReLU operates on each element individually, setting all negative pixel values to 0. This introduces non-linearity to the network, resulting in a rectified feature map as the output. The original image was scanned through multiple convolutions and ReLU layers to locate features. Then, the pooling operation selects the maximum element from the region of the feature map covered by the filter. The dropout layer is one of the techniques that reduces overfitting. The output from the dropout layer is fed by this fully connected layer. The feature vector generated from the fully connected layer is further utilized to classify images into different categories following the training process. Every input from this layer is connected to each activation unit in the next layer. Since all the parameters are occupied in a fully connected layer, it causes overfitting. SoftMax is an activation layer normally applied to the last layer of a network that acts as a classifier. The SoftMax function is used to map the non-normalized output of a network to a probability distribution. Finally, classification of the given input into distinct classes takes place at this layer.

5. Experiments and Results

5.1 Dataset

The dataset contains a total of 82 image categories. The images represent the English alphabet (A-Z), digits, and some symbols. The dataset can be used for various machine learning tasks, such as image classification and character recognition. It can be a useful resource for researchers, developers, and students working on projects related to computer vision, machine learning, and deep learning.

5.2 Deep Learning

Deep learning is used to extract relevant features that differentiate between 26 letters, 10 digits, and other symbols. The input to the deep learning step, for the first time, is the previous dataset letter recognition problem. This dataset is a group of labeled images that contain categories of the problem. This dataset is passed to the input layer of a convolutional neural network (CNN). The result of the last layer (the fully connected layer) is a set of labeled images with a probability for each image.

6. Performance of Technique

The Proposed CNN Network is experimented with using other networks.

6.1 Accuracy of Deep Learning Stage

According to the standard experimental design, all datasets were randomly split by assigning 70% of the images to training and 30% to validation

Furthermore, the accuracy measure is used, as it reviews the percentage of successfully recognized images with respect to the total dataset.

$$\textbf{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

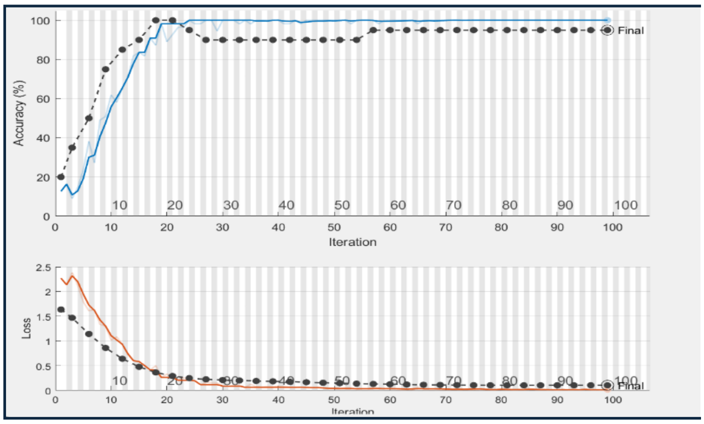
The following equation depicts the accuracy metric's formula. Where TP are the retrieved relevant matches, FP are the retrieved irrelevant matches, FN are the missed relevant matches and TN are the not retrieved relevant matches.

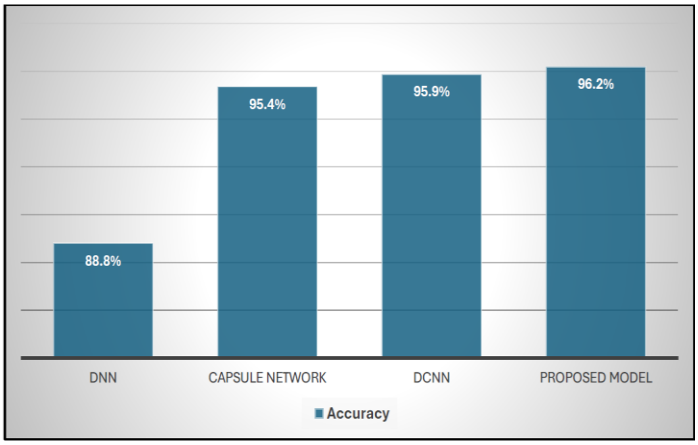
6.2 Biometric Metric for (CNNGA) Model

The performance of this research is evaluated by using the accuracy measure, as it reviews the percentage of successfully recognized images in the total dataset.

Method	Accuracy	FRR	FAR	EER
Proposed Model	**96.2%**	**4.9%**	**5.2%**	**5.5%**

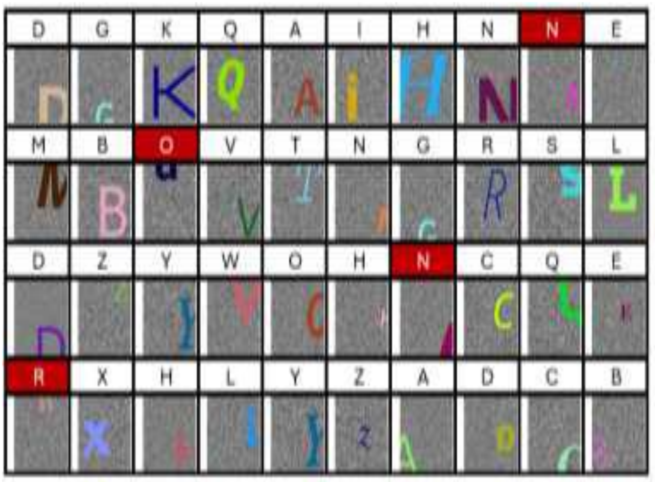
Another performance measure is biometric metrics that are used to evaluate the effectiveness of the proposed technique, as FRR refers to the false rejection ratio, which measures the percentage of valid inputs that are incorrectly rejected. FAR refers to the false acceptance ratio, which measures the percentage of invalid inputs that are incorrectly accepted. EER refers to an equal error rate, which is the rate at which both acceptance and rejection errors are equal.





Conclusion

An effective deep learning approach for letter recognition from input images was introduced in this paper. The presented CNN model achieved 96.2% accuracy. The challenge in this paper is using noisy and cropped input images. We add layers to the DL model to increase accuracy.



Future Work

Future work is on how to increase the OCR model's accuracy by gathering more diverse training datasets. Another work is allowing OCR systems to recognize letters in multiple languages.

References

1. Srivastava, Ankur, et al. "Handwritten character recognition using machine learning and deep learning." Emerging Trends in IoT and Computing Technologies. Routledge, 2023. 475-482.
2. Bhatt, Preeti P., and Isha Patel. "Optical character recognition using deep learning-a technical review." National Journal of System and Information Technology 11.1 (2018): 55.
3. Saqib, Nazmus, et al. "Convolutional-neural-network-based handwritten character recognition: an approach with massive multisource data." Algorithms 15.4 (2022): 129.
4. Younis, Khaled S. "Arabic hand-written character recognition based on deep convolutional neural networks." Jordanian Journal of Computers and Information Technology 3.3 (2017).
5. Balaha, Hossam Magdy, et al. "A new Arabic handwritten character recognition deep learning system (AHCR-DLS)." Neural Computing and Applications 33 (2021): 6325-6367.
6. Munappy, Aiswarya Raj, et al. "Data management for production quality deep learning models: Challenges and solutions." Journal of Systems and Software 191 (2022): 111359.
7. Liman, Muhamad Arief, Antonio Josef, and Gede Putra Kusuma. "Handwritten character recognition using deep learning algorithm with machine learning classifier." JOIV: International Journal on Informatics Visualization 8.1 (2024): 150-157.
8. Reutzel, Pamela, Kathleen AJ Mohr, and Cindy D. Jones. "Exploring the relationship between letter recognition and handwriting in early literacy

development." Journal of Early Childhood Literacy 19.3 (2019): 349-374.
9. Neisser, Ulric, and Paul Weene. "A note on human recognition of hand-printed characters." Information and Control 3.2 (1960): 191-196.
10. Naus, Mary J., and Robert J. Shillman. "Why a Y is not a V: a new look at the distinctive features of letters." Journal of Experimental Psychology: Human Perception and Performance 2.3 (1976): 394.
11. Neisser, Ulric, and Henry K. Beller. "Searching through word lists." British Journal of Psychology 56.4 (1965): 349-358.
12. Pelli, Denis G., Bart Farell, and Deborah C. Moore. "The remarkable inefficiency of word recognition." Nature 423.6941 (2003): 752-756.
13. Gibson, Eleanor P., et al. "A developmental study of the discrimination of letter-like forms." Journal of comparative and physiological psychology 55.6 (1962): 897.
14. Pelli, Denis G., et al. "Identifying letters." Vision Research 46.28 (2006): 4646-4674.
15. Sethi, Rohan, and Ila Kaushik. "Handwritten digit recognition using machine learning." 2020 IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT). IEEE, 2020.
16. Pashine, Samay, Ritik Dixit, and Rishika Kushwah. "Handwritten digit recognition using machine and deep learning algorithms." arXiv preprint arXiv:2106.12614 (2021).
17. Rai, Ankur, Siddharth Singh, and Sumit Kushwaha. "Optimized Handwritten Digit Recognition: A Convolutional Neural Network Approach." 2024 International Conference on Communication, Control, and Intelligent Systems (CCIS). IEEE, 2024.
18. Shetty, Ashish, and Sanjeev Sharma. "Ensemble deep learning model for optical character recognition." Multimedia Tools and Applications 83.4 (2024): 11411-11431.
19. Rakesh, S., et al. "Handwritten text recognition using deep learning techniques: A survey." MATEC Web of Conferences. Vol. 392. EDP Sciences, 2024.
20. Graves, Alex, et al. "A novel connectionist system for unconstrained handwriting recognition." IEEE transactions on pattern analysis and machine intelligence 31.5 (2008): 855-868.
21. Nikitha, A., J. Geetha, and D. S. JayaLakshmi. "Handwritten text recognition using deep learning." 2020 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT). IEEE, 2020.
22. Graves, Alex, and Jürgen Schmidhuber. "Offline handwriting recognition with multidimensional recurrent neural networks." Advances in neural information processing systems 21 (2008).
23. Liwicki, Marcus, et al. "A novel approach to on-line handwriting recognition based on bidirectional long short-term memory networks." Proceedings of the 9th International Conference on Document Analysis and Recognition, ICDAR 2007. 2007.
24. LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11 (1998): 2278-2324.
25. Arora, Sandhya, et al. "Devanagari Character Recognition: A Comprehensive Literature Review." IEEE Access (2024).
26. Madhvanath, Sriganesh, and Venu Govindaraju. "The role of holistic paradigms in handwritten word recognition." IEEE transactions on pattern analysis and machine intelligence 23.2 (2001): 149-164.

27. Fanany, Mohamad Ivan. "Handwriting recognition on form document using convolutional neural network and support vector machines (CNN-SVM)." 2017 5th international conference on information and communication technology (ICoIC7). IEEE, 2017.
28. Balaha, Hossam Magdy, et al. "A new Arabic handwritten character recognition deep learning system (AHCR-DLS)." Neural Computing and Applications 33 (2021): 6325-6367.
29. Younis, Khaled S. "Arabic hand-written character recognition based on deep convolutional neural networks." Jordanian Journal of Computers and Information Technology 3.3 (2017).
30. Bora, Mayur Bhargab, et al. "Handwritten character recognition from images using CNN-ECOC." Procedia Computer Science 167 (2020): 2403-2409.
31. Lahreche, Afafe, and Abdelhakim Cheriet. "Deep learning for Arabic letters recognition." (2019).

Author Profile

Esraa Farouk received the M.Sc. degree and PhD in Computer Science from Assiut University, Assiut, Egypt. She currently works as a professor at Faculty of Commerce, South Valley University, Qena, Egypt.

Email: esraa.mohamed11@com.svu.edu.eg