NTU Journal of Engineering and Technology (2024) 3 (4) : 42-51
DOI: https://doi.org/10.56286/ntujet.v3i4.1086

IRAQI Academic Scientific Journals

P-ISSN: 2788-9971 E-ISSN: 2788-998X
NTU Journal of Engineering and Technology
Available online at: https://journals.ntu.edu.iq/index.php/NTU-JET/index

High-Performance Character Recognition System Utilizing Deep Convolutional Neural Networks

Duaa Mowafak Hameed [1,2], Raid Rafi Omar Al-Nima [1]

1. Technical Engineering College / Mosul, Northern Technical University, Mosul, Iraq, duaa.mowafak@ntu.edu.iq , 2. Directorate of Ninawa Health, Ministry of Health, Iraq

Article Informations

Received: 30-07- 2024,
Revised: 15-09-2024,
Accepted: 13-10-2024,
Published online: 20-12-2023

Corresponding author:
Name: Duaa Mowafak Hameed
Affiliation : Technical Engineering College / Mosul
Email:
duaa.mowafak@ntu.edu.iq

Key Words:
Blind,
Letter Recognition System,
Deep Learning,
Visual Impairment.

A B S T R A C T

Accessible reading is still a major challenge for those with visual impairments in our digitally-driven world, particularly for those who are born blind. The creation of a Letter Recognition System (LRS) for the blind is the novel solution to this problem that this research suggests. With the help of this device, blind people can access printed letters according to its reading solution. In order to obtain images of printed letters, this study presents practical hardware that uses a webcam and a manual printing machine. The paper can be moved to acquire the written letters on it. Hence, a large collection of data containing letters known as Printed English Letters-version 2 (PEL2) dataset was gathered for the printed English letters (A–Z). Following acquisition, the input images undergo preparation, segmentation, and resizing. After that, a Deep Convolutional Neural Network (DCNN) is used to recognize letters from them. Ultimately, it is recommended that the identified letter be transformed into letter-to-speech audio to enhance the system's efficacy in assisting blind or visually impaired individuals with reading. In this work, a very high accuracy of 99.70% for letter identification has been calculated and attained.

Introduction

Supporting the visually impaired should be taken into account in technical and scientific studies. Individuals who are blind or visually impaired struggle to read texts or even recognize their surroundings. Nearly 285 million people worldwide, according to the World Health Organization (WHO), have visual impairments that make it difficult for them to access written or visual resources on a basic level [1]. Experts define "visual impairment" as the total or partial loss of vision that cannot be reversed by remedial procedures such as wearing glasses, contact lenses, or even surgery [2].

A person who loses his vision usually also becomes less independent and needs help with everyday duties. Illiterate or blind people are unable to do visual tasks like reading. Thus, to assist with text reading, the use of a braille reader or digital voice synthesizer is necessary [3]. Therefore, creating an intelligent system to aid those with vision impairments in reading seems necessity.

In the past few years, deep learning (DL) techniques have demonstrated considerable promise in a number of fields, such as speech recognition, image and textual categorization, face and facial expression identification, semantic-based video searches, and many more[4][5]. Therefore, it makes sense to implement it for this crucial role of assisting those who are blind or visually impaired with their reading.

The suggestion made here is for an intelligent LRS that is capable of efficiently reading, interpreting, and identifying written letters. A system like that would need a lot of experience with pattern recognition and image processing. It is advised for those who are blind or visually challenged. The contributions of this paper can be illustrated as follows:

- Establishing a new dataset of printed letters called the Printed English Letters-version 2 (PEL2) dataset by launching a useful hardware.
- The input images are acquired, pre-processed and prepared.
- A Deep Convolutional Neural Network (DCNN) model is proposed in this study for recognizing the printed letters. Then, the recognized letters can be read loudly in order to aid visual impairment people.

The remaining sections of this paper are organized as follows: literature review is presented in Section 2, theories of the suggested LRS are described in Section 3, experimental results are discussed in Section 4 and finally the conclusion of this study is given in Section 5.

Literature Review

Previous studies concentrated on the letter recognition in different aspects, this highlights its importance and significance. Examples of such studies can be reviewed as follows:

In 2016, Ardiansyah developed an image-to-speech application for blind people. It was for reading aloud passages from a book, magazine, paper, … etc. Raspberry Pi board, light sensor, Tesseract Optical Character Recognition (TOCR) library, Open source Computer Vision (OpenCV) library and Text-To-Speech (TTS) synthesizer library were integrated to help blind people in being able to read. It was explained that the suitably higher light condition could lead to better optical recognition process results. The using of image thresholding process, brightness amendment and contrast adjustment improved the optical recognition result with the median improvement of about 10.9 % [6].

In the same year, Suhasini *et al*. described a prototype system to read printed text on hand-held objects for assisting blind persons in their day to day lives. Motion based method was proposed to detect the object of interest. A novel framework was proposed to extract text strings with various colors, sizes and variable orientations from scene photos with complex and crowded backgrounds. To determine if the text was presented in the captured image or not, text region localization was used. The task of identifying the text in an image was performed for the utilized OCR. Text stroke direction and edge pixel distribution were analyzed by a neural network, which recognized every letter in the text and read the text that was extracted from the picture. Ultimately, after detecting the text and processing the data to produce audible signals, the user heard the text codes as voice [7]. In the same year, Thilakarathna and Wanniarachchi attempted to develop a recognition strategy for printed English characters (A to Z) in order to help visually impaired people to overcome their disability in reading. Images in determined formats were used as inputs. Subsequently, the noise presented in any acquired image was filtered. Next, an appropriate threshold was applied to turn the colored image into a binary image. After that, linked regions were identified using segmentation and individual letters were identified using bounding boxes. Ultimately, a statistical technique that made use of a previously saved sample database was utilized for recognition. The employed database included image-based compilation of English alphabet letters. The size of the database and the word length had a major impact on how quickly words are recognized. Apart from the recognition, a complete system of image acquisition and a text to voice conversion strategy were developed to be more useful in helping visually impaired persons to understand a writing context [8].

In 2019, Mathur *et al.* proposed an OCR artificial intelligence reading system. It used a smart phone camera to take pictures for typed, handwritten or printed texts. Either scanning the user-provided document or converting it into machine-encoded text was considered. After scanning an image, it was transformed into digital text where the system could recognize. Then, from the translated text, speech output was produced (text-to-speech strategy). English alphabets from A-Z or a-z could accurately be recognized. Numerical integer values from 0 to 9 could also be identified. All outputs could be delivered orally [9].

In 2020, Kurlekar *et al*. introduced a portable text-to-speech system that could convert an image of text into sound with a performance level that was satisfactory, a readability tolerance of less than 2% and processing times were typically less than three minutes for an image of paper size (A4). This portable system can be used by anyone on their own, without the need of internet connection. Word recognition on a localized text region was done using an OCR, which was then converted into an audio output to help blind users. A camera streamed the information into a Raspberry Pi board. Graphical User Interface (GUI) displayed the streaming data. After positioning an object in front of the camera for text reading, a captured button could be clicked to project the image onto the Raspberry Pi. The image was transformed into data using the Tesseract library and the data that were extracted from the image were displayed on a status bar. Text-to-speech synthesis was used to enunciate the acquired data over the headphones [10].

In 2022, OUALI *et al*. approached a system to correctly read traffic signs with Arabic text in a natural scene using the Augmented Reality (AR) technology. The AR is an interactive experience that enhances the real world with generated perceptual information. Smartphone camera was used to shoot, store and process a determined image. Then, the text in the image could be recognized, detected, and demonstrated in a highly clear and loud way. Reading aloud and recognizing textual inputs as Two-Dimensional (2D) visuals were the ultimate goal [11].

In 2023, Al-Nima and Al-Nima provided a study for recognizing handwritten Arabic alphabets. Images of separated Arabic letters from [12] were employed. A deep learning network model was suggested, it consisted of input layer, hidden layers and output layer. The input layer was capable of receiving a digital image of each handwritten Arabic alphabet separately. The hidden layers included consecutive convolutional and Rectified Linear Unit (ReLU) layers. The output layer was for configuring the character class, it was consisted of 28 neurons, where each neuron corresponded to a single Arabic letter. The recognition accuracy were firstly reached 82.47%. Upon analyzing the inputs and outputs, it was found that similar and relative characters had an impact on the recognition process. In addition, the characters that were unclearly or wrongly demonstrated also influenced the recognition. So, after excluding affected characters in training the proposed deep learning model, the recognition accuracy was significantly improved to 94.98% [13].

The literature study indicates that, to date, no prior research has employed the same technique for acquired letter images, thereby enabling blind individuals to become self-sufficient without external assistance.

Suggested Letter Recognition System (LRS)

As previously stated, this work suggests an LRS. There are three primary phases to it. Installing technology capable of accurately capturing, analyzing, recognizing, and interpreting printed letters is taken into consideration in the first step. The PEL dataset is established in the second stage, which involves applying appropriate pre-processing and preparation for the acquired images. Developing a DCNN model that can effectively recognize PEL2 images is the main goal of the third stage. After that, noises can be created using the identified letters. The entire process of obtaining PEL2 images (or reading them) and creating appropriate sounds for them is associated with the Tex-to-Speech (TTS) policy. People who are blind or visually impaired can benefit from the recommended LRS when reading.

In the first stage, appropriate hardware equipment are considered. Figure 1 shows all hardware equipment for the suggested LRS. They can be listed as follows:

(1) Typewriter machine.
(2) Webcam.
(3) Connection cable.
(4) Personal computer.
(5) Sound amplifier.

Figure 1: All hardware equipment for the suggested LRS

Every piece of equipment has a purpose. The first hardware item is the typewriter machine. Its main purposes are to let a piece of paper with a printed letter on it from entering and to move it letter by letter or step by step to the appropriate spots. Located atop the typewriter, the webcam is oriented at a PEL2. Its job is to take images of the PEL2, which pertains to the letters A through Z. The camera and a personal computer are then connected via a connection cable using a Universal Serial Bus (USB) connector. The computer performs the tasks of character recognition and collection. Lastly, the sound amplifier can be used to read aloud the identified PEL2.

It is worth highlighting that the suggested LRS includes acquiring or collecting input (raw) images, applying pre-processing, employing a DL for all phases of training, validation and testing, and obtaining the outputs.

Raw images

To begin with, prepared papers are controlled and inserted using a typewriter. As previously noted, a webcam facing a written letter is fixed atop the typewriter. The typewriter moves the paper that is input, and the camera may capture any focusse PEL2. The webcam is a Kisonli model. A total of 2600 images are taken. We obtain 100 samples for every letter. There are printed letters from A to Z involved. Each raw image that is captured must have dimensions of 640 x 480 x 3 pixels in the Joint Photographic Expert Group (JPG) format. Pre-processing processes are necessary for the raw data to be ready, as expected.

Pre-processing

Consequent pre-processing operations are applied. These are the: binarization, segmentation and resizing. They can be illustrated as follows:

1. **Binarization:** A useful operation in a lot of image processing systems is binarization. Following binarization, pixel values are reduced to just two values. These are, with 1 standing for the color white and 0 for the color black [14]. Binarization aims to reduce superfluous information in an image while preserving its valuable information [15]. Here, the Otsu thresholding is employed, it is a dynamic thresholding. By selecting a threshold that reduces the interclass variance of the thresholded black and white pixels, Otsu [Otsu 1979] creates a global threshold, assuming the existence of two classes: background and foreground [16], [17]. Let $I(x, y, z)$ is the original image. It is firstly converted to the grayscale $I(x, y)$. Then, it can be binarized, translated to a binary image.

2. **Segmentation**

As mentioned, the camera in the suggested LRS is already located facing a PEL2 in a paper. So, the captured and binarized image is easily segmented for each letter using fixed coordination. However, some images are resulted with combinations between the focused letters and their neighbors. This has been addressed here according to the following algorithm:

Step 1: Verify the image specs to make sure a black text is surrounded by white borders.

Step 2: If the letter is not truly located according to the specifications in Step 1, this means that there is a combined neighbor letter affects the white boundary.

Step 3: Begin removing the barrier where the combined neighbor letters are located.

Step 4: Repeat Step 3 until all affecting black is removed. This means that only the white boundary is left and the image specifications are set according to Step 1.

Step 5: Save and consider the resulted image.

Figure 2 demonstrates how the provided segmentation algorithm can address the problem of combined letters in an image.

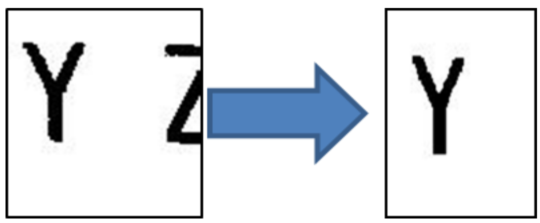


Figure 2: Demonstration of how the provided segmentation algorithm can address the problem of combined letters in an image

Segmentation of letter is a crucial step in letter recognition because it can affect the accuracy of a system [18]. Therefore, any LRS can be considered as robust if the data of letter are correctly segmented. Any segmented image here is considered as $S(x, y)$.

3. Resizing

The next stage of DL requires that each letter image be normalized to a set size. Consequently, all segmented pictures are normalized to a 100×100 pixels size in order to utilize them in the DL model for letter recognition. Any resized image can be represented as R(x,y).

DL model

In this study, a DCNN is proposed, it is inspired by the Fully Convolutional Neural Network (FCNN). Basically, a CNN can be considered as deep if it consists of no less than 12 layers [19], [20]. So, our network is called the DCNN as it consists of many layers and involves multiple convolutions and Rectified Linear Unit (ReLU) layers. Specifically, it composes of input layer, six convolution layers each followed by a ReLU layer, pooling layer, Fully Connected (FC) layer, softmax layer and classification layer. Full architecture of the proposed DCNN is given in figure 3.

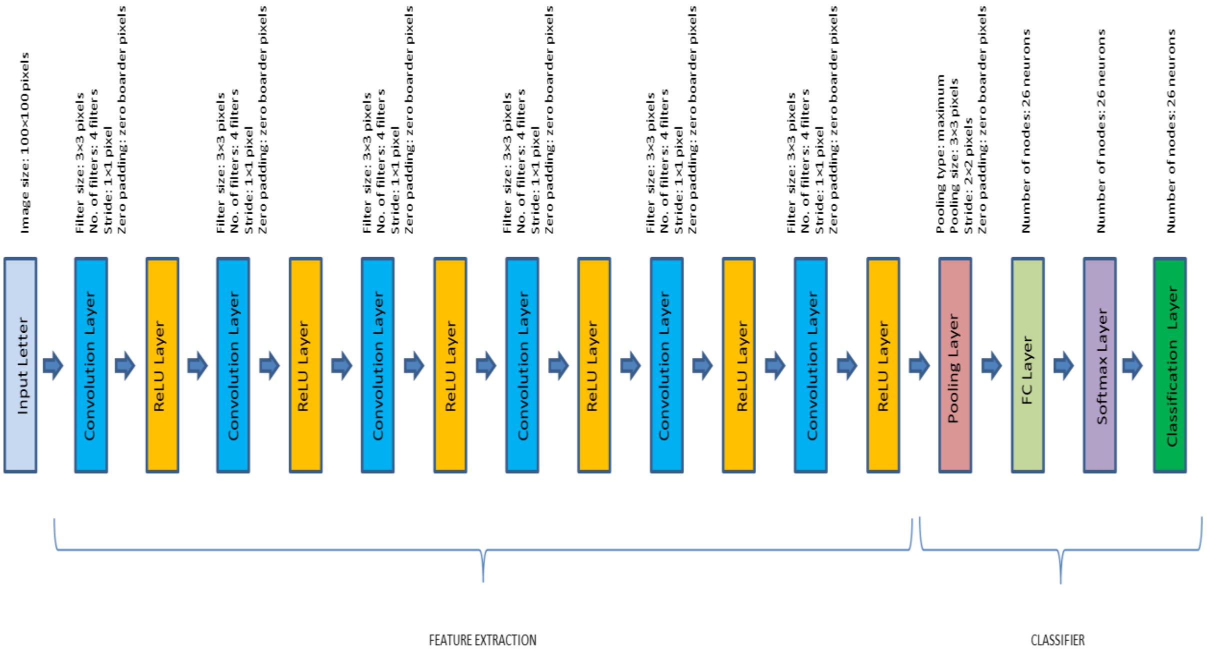


Figure 3: Full architecture of the proposed DCNN model

This figure displays the two primary DCNN components, the classifier and feature extraction, each of which has several layers. The R(x,y) letter image is actually fed into it as input. As a result, it uses multiple layers of convolutions, ReLUs, and pooling to extract the features. Subsequently, the features are analyzed in the classifier of FC, softmax, and classification layers.

Convolutional layers are crucial for extracting various information from images. Convolution of an image channel with a kernel (set of weights) is how this is accomplished. The convolution function can be represented by the following equation:

$$Z(x,y) = \sum_{m=1}^{M} \sum_{n=1}^{N} R(m,n) \times W(x-m, y-n) \quad (1)$$

where: $Z(x,y)$ represents a calculated convolution, M and N respectively represents the width and height of the Two-Dimensional (2D) input image, and $W(x-m, y-n)$ represents a kernel value [21].

It is worth mentioning that the $Z(x,y)$ is updated here for a certain convolution layer as there are multiple convolution layers in the proposed DCNN. So, it is here can be considered as $Z_\alpha(x,y)$, where: α represents the index of a certain convolution layer (α=1,2,3,…,6).

The ReLU is a well-known transfer function in the DL that is also utilized in many layers. Any previous channel's (feature map) negative values are eliminated while the positive values are preserved via the ReLU transfer function. Thus, it provides a non-linear computation [22]. Moreover, the ReLU transfer function can be represented by the following equation:

$$Re_\beta(x,y) = max(0, Z_\alpha(x,y)) \quad (2)$$

where: $Re_\beta(x,y)$ represents a computed ReLU, max represents to the maximum operation [23] , [24]and β represents the index of a ReLU layer that follows a convolution layer (β=1,2,3,…,6).

Subsequently, the received channel sizes are shrunk by the pooling layer, producing feature maps with small sizes. There are two possible types: average and maximal [25]. The maximum pooling type is considered here in the proposed DCNN. The general equation in the maximum pooling layer can be expressed as follows:

$$P(i,j) = max\,(\mathbf{V}) \quad (3)$$

where: $P(i,j)$ represents a computed pooling and $\mathbf{V}$ is a small matrix part (or a window) in a previous channel [20], [26].

Next, the DC layer is completed. It is made up of several neurons, each of which is connected to every other neuron in the layer above it. Here, the number of neurons may match the number of necessary classes (C). The essential equation in the FC layer can be illustrated as follows:

$$FC(c) = \sum_{l=1}^{p} (FW(c,l) \times P(l)) + B(c) \quad (4)$$

where: $FC(c)$ represents an FC value, c represents a class number in the FC layer (c=1,2,3,…, C), p represents the number of neurons in the previous layer, $FW(c,l)$ represents a connection weight value between the pooling and FC layers, and $B(c)$ represents a bias value [27].

The softmax layer provides the related probabilities of classes for a certain input, such probabilities sum up to one. This layer basically uses the softmax activation function. The general equation in the softmax layer can be provided as follows:

$$S(c) = \frac{exp(FC(c))}{\sum_{lo=1}^{C} exp(FC(lo))} \quad (5)$$

where $S(c)$ represents a calculated softmax value and exp is the exponential [28].

The classification layer is used at the end to achieve the recognition or classification decision. The winner-takes-all rule is employed in this layer. Its general equation can be expressed as follows:

$$D(c) = \begin{cases} 1 & \text{if } S(c) \text{ has the maximum value} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where $D(c)$ represents the output decision value [29].

Performance metrics:

The confusion matrix is the initial step in evaluating machine learning performance in categorization through a number of metrics. The confusion matrix provides information on the successful and unsuccessful classification of negative samples in addition to the accurate and erroneous classification of positive samples [30]. A two-class confusion matrix includes the parameters of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN), where they are firstly calculated and then the classifier performances can be analyzed in more details [31]. TP is the number of instances that are correctly predicted as positives, TN is the number of instances which are correctly predicted as negatives, FP is the number of instances that are incorrectly predicted as positives and FN is the number of instances which are incorrectly predicted as negatives [32]. Other metrics can be derived from these parameters: Accuracy, precision, sensitivity, Specificity and F1_score.

Results and Discussion:

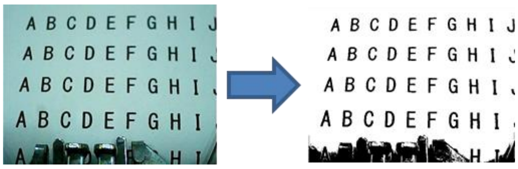
Established dataset:

First of all, a new dataset is established, it is termed the PEL2 dataset. It comprises of a very big number of images which is 2600 images. For each letter, 100 samples are acquired. It has the separated English letters from A to Z. The images of dataset are acquired using a webcam of type Kisonli High-Definition (HD) digital camera. This webcam is located on a typewriter. A printed paper is inserted into the typewriter and it is also be controlled by the typewriter. The camera can acquire a raw image for many PELs including the focused PEL. Each captured image has a resolution of 640×480×3 pixels and the format of all images is of type JPG, as mentioned. Figure 4 shows different samples of raw images for the established PEL dataset.

Figure 4: Different samples of raw images for the established PEL dataset

Pre-processing results:

Pre-processing steps are implemented following the creation of the PEL dataset. Binarization is the initial operation, in which the values of the pixels are transformed into one of two values: One for the color white and zero for the color black. Figure 5 shows an image example before and after binarization.



(a) (b)

Figure 5: (a) Example of acquired image before binarization. (b) binarized image.

An illustration of the binarization impact on an acquired colored image is shown in this figure. By lowering the image's tiny noise, binarization reduces unnecessary detail while preserving the image's important details. Additionally, it can deal with the issue of uneven illumination in an image.

Subsequently, the second operation of segmentation is applied. Here, each binarized image is segmented to obtain a certain letter by using fixed coordination.

Too many images are segmented depending on fixed coordination as the camera is well positioned facing a focused letter. However, there are some images included combinations between the focused letters and parts of other letters. These parts may be existed at the right side or at left side of the image. This problem has been sorted out by applying the provided segmentation algorithm that has been previously explained as shown in figure 6.

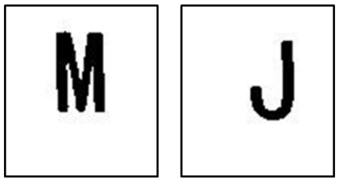


Figure 6: Different examples of segmented images

As it can be noticed from this figure, the provided segmentation algorithm can efficiently address the problem of combined letters in an image. Henceforth, all segmented images are normalized to a fixed size, as this is essential for the next step of DL. Each segmented image is resized to 100×100 pixels as given in figure 7.

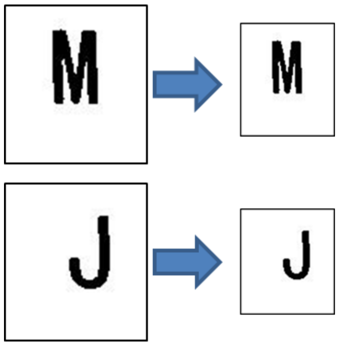


Figure 7: Examples of images before and after resizing, where the images at the left are segmented images and the images at the right are their resizing

DCNN results:

Firstly, the proposed DCNN parameters are determined after many experiments. They are also revealed in figure 3. That is, the DCNN parameters are set up as follows:

Input layer which accepts a binary image of size 100×100 pixels, filter size for each convolution layer is 3×3 pixels, number of filters for each convolution layer is 4 filters, stride for each convolution layer is 1×1 pixel, padding for each convolution layer is of type zero padding, pooling layer of type maximum, filter size for the pooling layer is 3×3 pixels, stride for the pooling layer is 2×2 pixels, padding for the pooling layer is of type zero padding and number of nodes for each of the last three layers is 26 neurons.

Moreover, overall images of the PEL dataset are partitioned to 50% for the training phase, 25% for the validation phase and 25% for the testing phase.

The DCNN network has been trained using the following parameters: Stochastic Gradient Descent with Momentum (SGDM) optimizer, 0.9 momentum value, 0.0001 initial learning rate, 128 mini-batch size and 500 maximum number epochs. Figure 8 shows the accuracy and loss (error) curves for the DCNN training and validation phases.

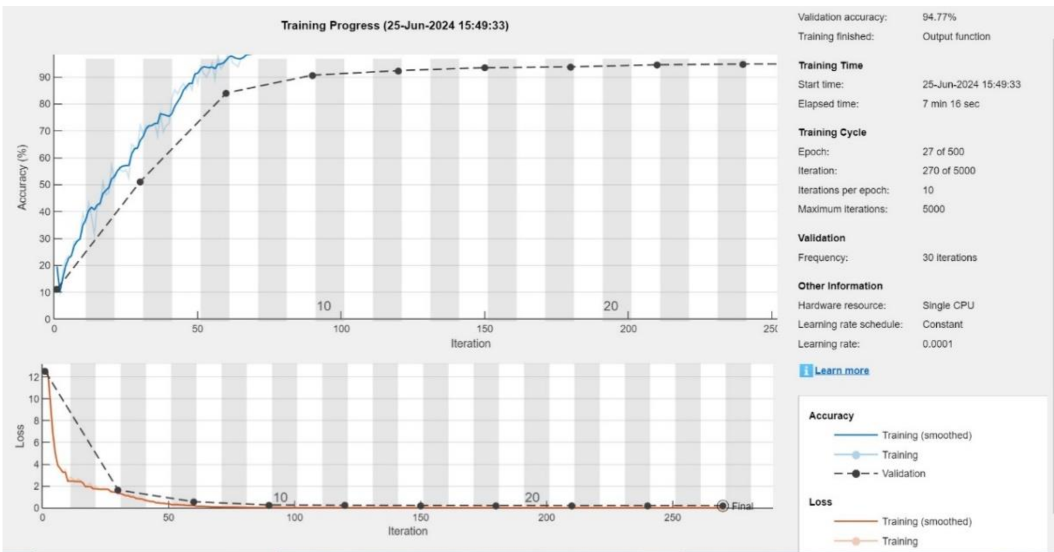


Figure 8: The accuracy and loss (error) curves for the DCNN training and validation phases

It is evident that as training progresses, the training and validation accuracies have successfully grown and their losses have gradually decreased. At the conclusion, the training loss has dropped to its lowest value of 0 and the training accuracy has reached its highest value of 100%. The validation loss has decreased to a very low value of about 0, while the validation accuracy has achieved a high value of 94.46%. The suggested DCNN model appears to have been successfully trained.

Consequently, the model is tested for evaluating its performances on data that have never been seen before. The general confusion matrix of testing the DCNN for by using the PEL dataset give results of the number of TP cases is 625 and the number of TN cases is 16225. Furthermore, the number of FP cases equals to the number of FN cases where each equal to 25. As mentioned, from these essential parameters the values of other metrics can be reached. Table 1 shows such values.

Table 1: Values of various measurement metrics for the DCNN performances by using the PEL dataset

Measurement Metric	Accuracy	Precision	Sensitivity	Specificity	F1_score
Value	99.70%	96.15%	96.15%	99.84%	96.14%

The testing results revealed that the accuracy is benchmarked to a very high ratio value of 99.70%. Other statistical results of precision, sensitivity, specificity and f1-score achieve the ratios of 96.15%, 96.15%, 99.84% and 96.14%, respectively. Obviously, all of these results are excellent, which refer to successful testing performances for the proposed DCNN.

Conclusion:

In this project, we presented a new database of letters termed the PEL2. An extensive collection of English letter images (from A to Z) can be found in this database. Additionally, a DCNN model was put forth. It was used to identify the alphabetic letters, which could be useful for blind persons. Thus, the PEL2 and the DCNN had two significances in this investigation. Moreover, a unique piece of hardware was recommended. It included a sound amplifier, a personal computer, a webcam, a connection cable, and a typewriter.

The PEL2 dataset contains a total of 2600 images. A total of 100 samples were obtained for every letter. Three pre-processing steps were applied to the acquired images: binarization, segmentation, and resizing. The known PEL2 dataset was used to assess the suggested DCNN model. Extensive experiments were explored for reaching best DCNN parameters. The main results were impressive and superior. That is, values of statistical metrics were recorded as 99.70%, 96.15%, 96.15%, 99.84% and 96.14% for the accuracy, precision, sensitivity, specificity and f1_score, respectively.

References

[1] R. Rajan and M. Devasena, “A Novel Reading Technique for Visually Impaired Person Using Enhanced Optical Character Recognition Method,” *Neuro Quantology*, vol. 20, no. 9, 2022.

[2] V. Kondhalkar, A. Shaikh, Y. Chaudhari, O. Bhandar, and G. Katkar, “An Intelligent Assistive System for Visually Impaired People,” *International Research Journal of Engineering and Technology*, 2021.

[3] Rajkumar N, A. G, and B. N, “Portable Camera-Based Product Label Reading For Blind People,” *International Journal of Engineering Trends and Technology*, vol. 10, no. 11, 2014.

[4] S. Abbas *et al.*, “Convolutional neural network based intelligent handwritten document recognition,” *Computers, Materials and Continua*, vol. 70, no. 3, pp. 4563–4581, 2022, doi: 10.32604/cmc.2022.021102.

[5] M. M. Ghazi and H. K. Ekenel, “A Comprehensive Analysis of Deep Learning Based Representation for Face Recognition,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 102–109, 2016, doi: 10.1109/CVPRW.2016.20.

[6] R. A. Ardiansyah, “Design of an electronic narrator on assistant robot for blind people,” *MATEC Web of Conferences*, vol. 42, pp. 1–5, 2016, doi: 10.1051/matecconf20164203013.

[7] P. S. Suhasini, D. Youbarani, C. Vijayasarathi, and R. Vani, “A Robust Embedded Based String Recognition for Visually Impaired People,” *International Conference on Computer Applications 2016 [ICCA 2016]*, pp. 94–97, 2016.

[8] G. M. N. Thilakarathna and W. K. I. L. Wanniarachchi, “English Character Recognition of an Image and Voicing System,” *Proceedings of APIIT Business, Law & Technology Conference*, pp. 136–140, 2017.

[9] A. Mathur, A. Pathare, P. Sharma, and S. Oak, “AI based Reading System for Blind using OCR,” *Proceedings of the 3rd International Conference on Electronics and Communication and Aerospace Technology, ICECA 2019*, pp. 39–42, 2019, doi: 10.1109/ICECA.2019.8822226.

[10] S. Kurlekar, O. A. Deshpande, A. V. Kamble, A. A. Omanna, and D. B. Patil, “Reading Device for Blind People using Python, OCR and GTTS,” *International journal of Science and*

Engineering Applications, vol. 9, no. 4, pp. 049–052, 2020, doi: 10.7753/ijsea0904.1003.

[11] I. Ouali, M. Ben Halima, and A. Wali, "Augmented Reality for Scene Text Recognition, Visualization and Reading to Assist Visually Impaired People," *Procedia Comput Sci*, vol. 207, no. Kes, pp. 158–167, 2022, doi: 10.1016/j.procs.2022.09.048.

[12] "Arabic Handwritten Characters Dataset."

[13] Raid Rafi Omar Al-Nima and Maqbool Ali Basheer Al-Nima, "Arabic Handwriting Recognition by Using Deep Learning," *Kashkool for publication and distribution*, 2023.

[14] P. Puneet and N. Garg, "Binarization Techniques used for Grey Scale Images," *International Journal of Computer Application*, vol. 71, no. 1, pp. 8–11, 2013, doi: 10.5120/12320-8533.

[15] A. Choudhary, S. Ahlawat, and R. Rishi, "A Neural Approach to Cursive Handwritten Character Recognition Using Features Extracted from Binarization Technique," in *Studies in Fuzziness and Soft Computing*, vol. 319, 2015. doi: 10.1007/978-3-319-12883-2.

[16] P. Stathis, E. Kavallieratou, and N. Papamarkos, "An evaluation technique for binarization algorithms," *Journal of Universal Computer Science*, vol. 14, no. 18, pp. 3011–3030, 2008.

[17] N. OTSU, "A threshold selection method from gray-level histograms," *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND SYBERNETICS*, vol. 9, no. 1, pp. 855–860, 1979, [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4310076

[18] M. R and S. K, "Morphological Operations and Projection Profiles based Segmentation of Handwritten Kannada Document," *International Journal of applied Information System*, vol. 4, no. 5, pp. 13–19, 2012, doi: 10.5120/ijais12-450704.

[19] L. Zhu and P. Spachos, "Annotation Efficiency in Multimodal Emotion Recognition with Deep Learning," *IEEE Conference on Global Communications*, 2022.

[20] I. Firdausi, I. Mukhlash, A. Gama, and N. Hidayat, "Sentiment analysis of customer response of telecommunication operator in Twitter using DCNN-SVM Algorithm," *Journal of Physics: Conference Series*, 2019.

[21] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," *Proceedings of 2017 International Conference on Engineering and Technology*, vol. 2018, pp. 1–6, 2017, doi: 10.1109/ICEngTechnol.2017.8308186.

[22] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Proceeding of the 25th International Conferences on Neural Information Processing Systems*, vol. 60, 2017.

[23] A. Al-Obaidi, R. Al-Nima, and T. Han, "Interpreting Arabic Sign Alphabet by Using the Deep Learning," *AIP Conference Proceeding*, vol. 2862, no. 1, 2023, doi: 10.1063/5.0171428.

[24] S. O. Ali, R. R. O. Al-Nima, and E. A. Mohammed, "Individual Recognition with Deep Earprint Learning," *International Conference on Communication and Information Technology*, pp. 304–309, 2021, doi: 10.1109/ICICT52195.2021.9568410.

[25] I. Firdausi, I. Mukhlash, A. Gama, and N. Hidayat, "Sentiment analysis of customer response of telecommunication operator in Twitter using DCNN-SVM Algorithm," *Journal of Physics: Conferences Series*, vol. 1490, no. 1, 2020, doi: 10.1088/1742-6596/1490/1/012071.

[26] S. O. Ali, R. R. O. Al-Nima, and E. A. Mohammed, "Individual Recognition with Deep Earprint Learning," *International Conference on Communication and Information Technology, ICICT 2021*, pp. 304–309, 2021, doi: 10.1109/ICICT52195.2021.9568410.

[27] H. Wu and J. Zhao, "Deep convolutional neural network model based chemical process fault diagnosis," *Computers and Chemical Engineering*, vol. 115, pp. 185–197, 2018, doi: 10.1016/j.compchemeng.2018.04.009.

[28] S. Abbas *et al.*, "Convolutional neural network based intelligent handwritten document recognition," *Computers, Materials and Continua*, vol. 70, no. 3, pp. 4563–4581, 2022, doi: 10.32604/cmc.2022.021102.

[29] J. Wu, "Introduction to Convolutional Neural Networks," *National Key Lab for Novel Software Technology*, 2017.

[30] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing and Management* , vol. 45, no. 4, pp. 427–437, 2009, doi: 10.1016/j.ipm.2009.03.002.

[31] S. Martínez, D. Gila, A. Beyaz, J. Ortega, and J. García, "A computer vision approach based on endocarp features for the identification of olive cultivars," *Comput Electron Agric*, vol. 154, no. May, pp. 341–346, 2018, doi: 10.1016/j.compag.2018.09.017.

[32] N. Saqib, K. Haque, V. Yanambaka, and A. Abdelgawad, "Convolutional-Neural-Network-Based Handwritten Character Recognition: An Approach with Massive Multisource Data," *Algorithms*, vol. 15, no. 4, Apr. 2022, doi: 10.3390/a15040129.