Arabic Handwritten Recognition Using Hybrid CNN, HMM and an Intelligent Network Based on Dentate Gyrus of the Brain

Nazal Modhej
Department of Computer Engineering, Karaj Branch, Islamic Azad University, Karaj, Iran

Mohammad Teshnehlab*
Department of Electrical Engineering of K. N. Toosi University of Technology, Tehran, Iran

Azam Bastanfard
Department of Computer Engineering, Karaj Branch, Islamic Azad University, Karaj, Iran

Somayeh Raiesdana
Department of Biomedical Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran

Received: 28 August 2022 – Revised: 11 December 2022 - Accepted: 20 January 2023

Abstract—Handwritten character recognition has occupied a substantial area due to its applications in several fields and is used widely in the modern world. Handwritten Arabic recognition is a major challenge because of the high similarity in its characters and its various writing styles. Deep learning algorithms have recently shown high performance in this area. The problem is that a deep learning algorithm requires large datasets for training. To overcome this problem, an efficient architecture is presented in this study, which comprises Hidden Markovian Model for character modeling, Convolutional Neural Network for feature extraction, and an intelligent network for recognition. The proposed network is modeled based on the dentate gyrus of the hippocampus of the brain. This part of the brain is responsible for identifying highly overlapping samples. The handwritten Arabic alphabet is characterized by this high overlap. Modeling the functionality of the dentate gyrus can improve the accuracy of the handwritten Arabic characters. Experiments are done using IFN/ENIT, CMATERdb3.3.1 and, MADBase datasets. The proposed approach outperformed recently published works using the same dataset. Although in all the experiments, a character error rate (CER) of less than 1.63 was obtained, the CMATERdb3.3.1 dataset resulted in a CER of 0.35. Compared with the convolutional neural network, the proposed network showed superiority in recognizing patterns with high noise.

Keywords: Handwritten Arabic recognition, Convolutional Neural Network, Hidden Markovian Model, Dentate Gyrus; Overlap.

Article type: Research Article

© The Author(s). Publisher: ICT Research Institute

* Corresponding Author
I. INTRODUCTION

In the field of character recognition, little progress has been made for the Arabic language [1]. Handwritten Arabic character recognition has still remained a challenge [2]. The alphabet of Arabic consists of 28 characters, and each letter has several shapes based on its place of occurrence in a word, which can be at the beginning, in the middle, or at the end. However, some letters are isolated in a word. In the Arabic alphabet, 15 letters have one or more dots above or below the characters. Some Arabic characters are similar in type, being only different from each other in the number and position of dots. Table I presents some of the variable shapes for Arabic characters and the written types of dots. The last row presents different shapes for the same characters based on their position in Arabic words.

The automatic recognition of Arabic handwriting requires building a hybrid recognition system: computing different features and applying various classifiers to achieve improved performance [3].

Most methods have dealt with handwritten Arabic characters and have reduced the extracted features [1]. The original features must be transformed into another domain [4]. However, these methods are not reliable for narrowing down the most efficient features [4]. In the study by [5], PCA was applied to choose the most informative features, and the highest classifying accuracy was reported to be 87%. A feature selection method based on PSO used in [6] reduces the entire feature set by choosing the best-produced features. Several classifiers have been applied to the produced features. [7] used a feature selection method based on a swarm intelligence algorithm named the bat algorithm. A 50% feature reduction was successfully performed based on their proposed approach, with a classification performance of 84%. The recognition of handwritten Arabic characters is still a relatively unaddressed research problem [1]. Traditional methods have effectively been replaced by deep learning methods and have yielded significantly better results. Deep learning is one of the fastest-growing areas in machine learning and promises to reshape the future of artificial intelligence [2]. However, deep learning requires large datasets for training because of the huge number of parameters that have to be tuned by a learning algorithm [2]. In this study, an intelligent network based on the functionality of the Dentate Gyrus (DG) of the brain is proposed. Pattern separation is a memory processing that separates patterns that have highly overlapping information [8] and is supported by the DG of the brain [9]. Therefore, modeling the functionality of pattern separation in this part of the brain can help recognize handwritten Arabic words with high overlap. Convolutional Neural Network (CNN) was used as a feature extractor in this study. Character modeling was performed using the Hidden Markovian Model (HMM), and in the end, recognition of handwritten Arabic was performed using the proposed DG’s network. The full proposed architecture was tested using three well-known datasets: IFN/ENIT, CMATERdb3.3.1, and MADBase datasets. The proposed network was compared with CNN-based methods. The results show the superiority of the proposed network over other CNN-based methods.

The remainder of this paper is organized as follows: Section 2 summarizes previous related work. The full description of the proposed network is provided in Section 3. The experimental results are presented in Section 4. Section 5 is dedicated to the discussion. Finally, Section 6 contains the conclusion.

II. RELATED WORK

Studies that have investigated the recognition of handwritten Arabic can generally be divided into traditional methods, deep learning-based methods, and hybrid methods. These three categories are explained in the following subsections.

A. Traditional Recognition of Handwritten Arabic

Traditional methods are limited in acting directly on a raw form [2]. The artificial neural network (ANN) was used to train and test the CENPRMI dataset [10]. They devised a new method for recognizing handwritten Arabic based on novel preprocessing operations and different features and reported 88% accuracy of test recognition. A combination of features extracted from curvelet transform and spatial domains was trained with the ANN classifier [11]. An accuracy rate of 90.3% was achieved in this research. The ANN was also used for training in the study by [12]. The mentioned study used statistical features and geometric moment features independent of the character's font and size. The results showed that using only statistical features was less accurate than using systems of hybrid features. Although the execution time was high, an average recognition rate of 97% was achieved. The structure of the LeNet neural network could recognize the main and secondary components [13].

TABLE I. CHARACTER VARIATIONS IN HANDWRITTEN ARABIC

<table>
<thead>
<tr>
<th>Challenging</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dots exit or not</td>
<td>ش غ ض ه ز ع ض ه ز ع ض ه ز ع ض ه ز ع</td>
</tr>
<tr>
<td>Number and</td>
<td>ج ي ح ت ط ن ب ب ب ب ب ب ب ب ب ب ب ب ب</td>
</tr>
<tr>
<td>position of dots</td>
<td></td>
</tr>
<tr>
<td>Different shapes</td>
<td>خ ﻕ ﺑ ﺑ ﻕ ﺑ ﺑ ﺑ ﺑ ﺑ ﺑ ﺑ ﺑ ﺑ ﺑ ﺑ ﺑ</td>
</tr>
</tbody>
</table>

Traditional methods are limited in acting directly on a raw form [2]. The artificial neural network (ANN) was used to train and test the CENPRMI dataset [10]. They devised a new method for recognizing handwritten Arabic based on novel preprocessing operations and different features and reported 88% accuracy of test recognition. A combination of features extracted from curvelet transform and spatial domains was trained with the ANN classifier [11]. An accuracy rate of 90.3% was achieved in this research. The ANN was also used for training in the study by [12]. The mentioned study used statistical features and geometric moment features independent of the character's font and size. The results showed that using only statistical features was less accurate than using systems of hybrid features. Although the execution time was high, an average recognition rate of 97% was achieved. The structure of the LeNet neural network could recognize the main and secondary components [13].
network configuration had more than three layers, which increased the computational potential and complexity of the network. The mean square error average reported in [13] was on the test set. The recurrent neural network (RNN) is one of the popular networks that was integrated efficiently with bidirectional Long Short-Term Memory (BLSTM) [14]. This combination was proposed for the overfitting of avoidance. In [15], a multi-dimensional RNN (MDRNN) was combined with Connectionist Temporal Classification (CTC). A combination of multiple residual networks (ResNet) was applied for the recognition of Arabic words and yielded superior results. This combination outperformed the single residual network [16]. A genetic algorithm was used in [17] study to classify handwritten Arabic characters. Structural features were extracted to distinguish the shape of each character. The system achieved an accuracy of approximately 87%. Various classification methods, such as Quadratic Discriminant Analysis (QDA), Diagonal QDA (DQDA), Linear Discriminant Analysis (LDA), Diagonal LDA (DLDA), and KNN (K-Nearest Neighbours) with an average accuracy of 87% were used in a study by [4]. Also, methods based on Hidden Markovian Models (HMM) for handwritten Arabic recognition were presented by [18, 19]. Furthermore, different sets of HMM were applied for the classification of different Arabic word segments [20]. The HMM outperforms BLSTM when applied to the distribution of concavity features due to character modeling [18].

B. Deep learning-based methods

Deep learning methods have attracted much attention [21] and have brought great success in diverse application domains [22]. The primary objective of deep learning is to learn high-level and deep characteristics that are most potentially useful for classification or target detection [23]. Deep learning techniques demonstrate better recognition performance than the other crafted feature detectors in computer vision and building recognition systems [3].

Different types of deep networks, like multi-column deep neural networks (MCDNN) [24] and deep belief networks (DBN) [25], with satisfactory results, have recently been proposed. [26] investigated Dropout and DropConnect as regularization techniques for training for DBN architecture. Besides, the most popular deep neural networks are known as Convolutional Neural Networks (CNN). CNN has been widely used in handwritten recognition. CNN has shown promising results compared to other relevant Arabic OCR by reducing the dimension of images [27]. The architecture proposed by [25] consisted of two RBMs, each with 1000 hidden neurons. An accuracy of 97.90% was reported for the experimental study by using the HACDB database. CNN was also applied for feature extraction and classification as well [28]. This widespread deep network has found its way into recognizing handwritten texts [29].

From scratch, a DCNN model was recently proposed in [30] study to recognize isolated handwritten Arabic characters. A handwritten Arabic Character Dataset (AHCD) was proposed in later research, which achieved significant improvements with 94.9% of satisfying classification. In [31], a new method was presented for recognizing the handwritten Farsi/Arabic digits by fusing the recognition results of several CNNs with a gradient descent training algorithm. The experiments were conducted on an extended IFHCDB test database. The results revealed a very high accuracy classifier outperforming most of the previous systems.

C. Hybrid methods

A combination of a special type of RNN and Deep Bidirectional LSTM (DBLSTM) was proposed in [32]. Another approach by [33] proved that the combination of a Convolutional DBN (CDBN) and a Support Vector Machine (SVM) outperforms the DBN-based approaches.

Other approaches [34] integrated CNN and the SVM. Applying SVM on features extracted by CNN achieved a 92.95% recognition rate [34]. CNN was also combined with HMM in other approaches by [35,36]. Using the CNN model in the field of on-line handwritten Arabic recognition was discussed in [2]. A novel architecture was proposed by [2], which was composed of three convolutional and max-pooling layers followed by two fully connected layers. The obtained results were very promising, and an accuracy of 97.32% was achieved.

III. THE PROPOSED METHOD

The proposed approach comprises four stages: preprocessing, character modeling, feature extraction, and recognition. The full architecture of the proposed approach is given in Fig. 1.

Figure 1. The full proposed architecture.
The output of the first two-step of the proposed architecture is given to CNN as a feature extractor. The idea behind the proposed architecture is that instead of passing the output features to perceptron, CNN’s extracted features are given to the proposed network that is modeled based on the computation function of the DG of the brain. Full descriptions of the four stages of the proposed approach, as depicted in Fig. 1, are given in the following subsections.

A. Preprocessing

The main objective of preprocessing is to improve the character images by removing the non-informative pixels. The preprocessing stage and binarization in the proposed approach are performed based on [37]. Some noise removal, such as dilation and median filtering based on [38], were applied.

B. Character Modeling

Segmentation of the training and testing images as preprocessing is critical for consistent post-analysis. This stage leads to better final accuracy of the recognition. Any Arabic word image can be modeled by the concatenation of the sequence of characters that are arranged horizontally. Each Arabic word can be segmented into two units: characters or graphemes. In this study, HMM’s character modeling was used [18], in which each character model has a right-left topology.

Character models are built to consider the different shapes of characters according to their position in words: at the beginning, in the middle, or at the end of the word. Fig. 2 shows the Arabic model of the word “رواد”.

Generally, 167 characters/spaces are built that include 26 basic characters of the Arabic alphabet, the characters with additional marks such as chadda (‘), and space inter pseudo words.

C. Feature Extraction Based on CNN

Feature extraction is one of the essential parts of handwritten recognition, which affects the recognition accuracy. The convolutional neural network is an essential type of deep neural network capable of performing an appropriate representation of the inputs. Deep neural networks do not require predefined features. Rather, they can learn specific features during the training process. This process has made them a good option for a feature extractor [18]. The entire architecture of CNN can be divided into two broad sections. Feature learning section and classification section, as shown in Fig. 3.

As shown in Fig. 3, CNN consists of the convolutional layer’s plies and pooling layer pairs. The convolutional layer, as its name implies, converts the images using the convolution operations. The pooling layer integrates the neighboring pixels into a single pixel; therefore, the pooling layer reduces the dimensions of the image.

In this study, the convolutional neural network is used as a feature extractor. Then, the feature extracted from CNN is passed into the proposed network, which is modeled based on the DG of the hippocampus, as illustrated in Fig. 4.

Various types of architectures have been presented for the convolutional neural networks. These different CNN architectures are different in the number and types of layers. The most popular CNN architectures are AlexNet architecture [39], ResNeXT [40], DenseNet [41], and Residual Networks (ResNets) [42]. In this study, CNN was used for feature extraction of handwritten Arabic based on AlexNet architecture. The employed architecture is depicted in Fig. 5.
Figure 3. Architecture of the convolutional neural network

Figure 4. The proposed architecture comprises HMM, CNN, and an intelligent network based on the brain's Dentate Gyrus.

Figure 5. AlexNet architecture of CNN

Figure 6. Structure of the proposed network based on Dentate Gyrus of hippocampus

In this study, a single convolution layer with 20 (9×9) filters was used. The output of the convolutional filter was passed through a ReLU function.

D. Character recognition based on DG

Hippocampus has a critical role in memory [43] and spatiotemporal information [44]. This region of the brain can represent the temporal order of events.
[43], Dentate Gyrus is one of the main parts of the hippocampus and has a principal role in pattern separation [45]. Pattern separation is a process that transforms similar inputs into one without similarity and reports non-overlapping [45,46]. Regarding the DG’s role in pattern separation, many studies have attempted to represent computational theories for the DG [47-50]. According to the basic computational model of the DG of the brain [51], a network for Arabic character recognition is proposed in this study. The proposed network includes five main steps. The structure of the proposed network is demonstrated in Fig. 6.

As depicted in Fig. 6, the cells involved in the proposed network are the granule cells, interneurons, and mossy cells. Moreover, the granule cells are the essential cells in this network which are affected by the input cells, internal inhibitor cells, and mossy cells. The final activity of the granule cells determines the network output. These cells are excited by the input and mossy cells and inhibited by inhibitor cells. The proposed network employs two excitation steps and one inhibition step, which are a prominent feature. These two levels of excitation and one level of inhibition increase the accuracy of recognition. In the following sections, various phases of Fig. 7 with the feedforward relations are entirely described in steps one to five. The feedback learning relation is defined in the fifth step. All relations of the proposed network are presented based on the computational model of the DG in the brain [51].

$$f(g_{in}^i(j)) = e^{-\frac{(g_{in}^i(j) - c)^2}{\sigma}}$$ (3)

In Equation 3, c and σ are the center of the function and the standard deviation, respectively.

2) Step2: Inhibition from interneurons

The purpose of the second step is to turn off some of the granule cells in each layer utilizing the interneurons, intending to transfer only some of the granule cells, and not all of them, to the next step. Fig. 8 illustrates this step in the proposed network.

The inhibitory neurons cause inhibition on the granule cells and turn some of them off. In Fig. 8, the granule cells that are silenced are shown in dark colors. The remainder of the following operations is performed on the cells which survive from local inhibition. Removing several cells from the network output. These cells are excited by the input and mossy cells. The final activity of the granule cells determines the network output. These cells are excited by the input and mossy cells and inhibited by inhibitor cells. The proposed network employs two excitation steps and one inhibition step, which are a prominent feature. These two levels of excitation and one level of inhibition increase the accuracy of recognition. In the following sections, various phases of Fig. 7 with the feedforward relations are entirely described in steps one to five. The feedback learning relation is defined in the fifth step. All relations of the proposed network are presented based on the computational model of the DG in the brain [51].

$$f(g_{in}^i(j)) = e^{-\frac{(g_{in}^i(j) - c)^2}{\sigma}}$$ (3)

In Equation 3, c and σ are the center of the function and the standard deviation, respectively.

2) Step2: Inhibition from interneurons

The purpose of the second step is to turn off some of the granule cells in each layer utilizing the interneurons, intending to transfer only some of the granule cells, and not all of them, to the next step. Fig. 8 illustrates this step in the proposed network.

The inhibitory neurons cause inhibition on the granule cells and turn some of them off. In Fig. 8, the granule cells that are silenced are shown in dark colors. The remainder of the following operations is performed on the cells which survive from local inhibition. Removing several cells from the network output. These cells are excited by the input and mossy cells. The final activity of the granule cells determines the network output. These cells are excited by the input and mossy cells and inhibited by inhibitor cells. The proposed network employs two excitation steps and one inhibition step, which are a prominent feature. These two levels of excitation and one level of inhibition increase the accuracy of recognition. In the following sections, various phases of Fig. 7 with the feedforward relations are entirely described in steps one to five. The feedback learning relation is defined in the fifth step. All relations of the proposed network are presented based on the computational model of the DG in the brain [51].

$$f(g_{in}^i(j)) = e^{-\frac{(g_{in}^i(j) - c)^2}{\sigma}}$$ (3)

In Equation 3, c and σ are the center of the function and the standard deviation, respectively.

2) Step2: Inhibition from interneurons

The purpose of the second step is to turn off some of the granule cells in each layer utilizing the interneurons, intending to transfer only some of the granule cells, and not all of them, to the next step. Fig. 8 illustrates this step in the proposed network.

The inhibitory neurons cause inhibition on the granule cells and turn some of them off. In Fig. 8, the granule cells that are silenced are shown in dark colors. The remainder of the following operations is performed on the cells which survive from local inhibition. Removing several cells from the network output. These cells are excited by the input and mossy cells. The final activity of the granule cells determines the network output. These cells are excited by the input and mossy cells and inhibited by inhibitor cells. The proposed network employs two excitation steps and one inhibition step, which are a prominent feature. These two levels of excitation and one level of inhibition increase the accuracy of recognition. In the following sections, various phases of Fig. 7 with the feedforward relations are entirely described in steps one to five. The feedback learning relation is defined in the fifth step. All relations of the proposed network are presented based on the computational model of the DG in the brain [51].
3) **Step 3: Excitation from the mossy cells**

This step reflects the influence of granule cells on mossy cells, and in contrast, the excitation of mossy cells against the same granule cells. Figure 9 shows the operations performed at this step.

![Figure 9. Excitation of granule cells by mossy cells](image)

As depicted in Fig. 9, mossy cells excite the active granule cells, which are shown in white in Fig.9. Excitation of the granule cells by mossy cells is figured as:

$$g_{e-m}^i (j) = \beta_{ac} \sum_{m} Y_m W_{ij}$$  \hspace{1cm} (5)

In Equation 5, $g_{e-m}^i (j)$ demonstrates the excitation of the $j$th granule cell in the $i$th layer and $\beta_{ac}$ is a constant value. $Y_m$ is the activity of the $m$th mossy cell calculated by relation 6.

$$Y_m = \sum_j Y_j$$  \hspace{1cm} (6)

The movement of this step will make the remaining active granule cells stronger than before.

4) **Step 4: Calculation of the output of the proposed network**

After all of the stages above, the remaining active granule cells are used for calculating the system's output. Fig. 10 shows the final layers of granule cells. In Fig. 10, the silenced granule cells are shown in black, and the output of the network (active granule cells) is shown in white.

![Figure 10. The output granule cells of the proposed network](image)

The final activity of each granule cell (output of the proposed network) in each layer is determined as follows:

$$V_j^i (j) = v_{rest}^j (j) + g_{e-pp}^i (j) - g_{in}^i (j) + g_{e-m}^i (j)$$  \hspace{1cm} (7)

In Equation 7, $V_j^i$ represents the potential of the $i$th granule cell in the $j$th layer. $v_{rest}^j (j)$ represents the resting potential of granule cells $i$ in the layer $j$, which is considered a constant. $g_{e-pp}^i (j)$ indicates the excitation from the input. The inhibition from the interneurons is represented as $g_{in}^i (j)$. $g_{e-m}^i (j)$ shows excitation from the mossy cells into the granule cells, which were explained in the subsections given above.

For two similar Arabic word images, the number of the proposed network's granule cells is similar. Opposite conditions occur for two different Arabic word images. Fig. 11 and Fig. 12 illustrate the results of the proposed network for various similar and dissimilar words.

![Figure 11. Final activity of granule cells for two styles of an Arabic word](image)

![Figure 12. Final activity of granule cells for two dissimilar Arabic words: رواد and شماخ](image)

As shown in Fig. 11, the last activity of the granule cells for the two styles of the word رواد are very similar. The same granule cells of a different layer of the proposed network are silenced, and the same cells are active. Also, the output of the active granule cells was similar. Fig. 12 shows the results of the proposed network for two different words رواد and شماخ. As is shown, the same granule cells of the proposed network were not compatible.

5) **Step 5. Learning in the proposed network**

Weights of connections between the granule cells and input cells need to be updated after every network trail. Granule cells in the proposed network use the “winner take all” structure [51], meaning that more potent cells silence the rest of the cells. The fact that there is one inhibition step in the structure of the proposed network also confirms this issue. Therefore, competitive learning is the best approach for training the proposed network. As a result, the weights of vectors of granule cells are trained by competitive learning at the end of each network implementation trail.

IV. **EXPERIMENTAL RESULTS**

The proposed network was evaluated using images with high overlaps. Recognition of handwritten Arabic character images with high overlaps requires high
accuracy and speed. The simulation of this study was carried out using three well-known datasets for Arabic characters and digit: IFN/ENIT database [52], which is one of the challenging handwriting Arabic datasets used for handwritten Arabic recognition; MADBase [53], and CMATERdb3.3.1 [54] for handwritten Arabic digit recognition. The number of layers of the granule cells was determined to be three. The number of granule cells in each layer was considered as 15. The numbers of mossy and HIPP cells in various experiments were considered as 7 and 10, respectively. Table 2 shows the initial values of the parameters in the proposed network.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\gamma_{\text{era}}$</th>
<th>$\beta_{\text{MC}}$</th>
<th>$\beta_{\text{HIPP}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial value</td>
<td>0.3</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

### A. Datasets

1) **IFN/ENIT**

IFN/ENIT provides training and testing images to support handwriting recognition. It comprises about 2200 (300 dpi) binary images of handwriting samples from approximately 411 different writers. Twenty-six-thousand binary word images represent town/village names extracted from the forms and are saved individually in five different sets. These five sets are labeled a, b, c, d, and e [52]. Formerly, sets a, b, and c were used for training, while set d was used for testing. The proposed network was tested using sets a, b, and c for training and set d for testing. Experiments were also performed using sets a, b, c, and d for training, while set e was used for testing.

2) **CMATERdb3.3.1**

CMATER [54] is the pattern recognition database repository created at the CMATER laboratory in Jadavpur University using special kinds of pre-formatted datasheets filed by people of different age groups. For various scripts, individual digits and characters are extracted from the sheets. CMATERdb3.3.1 [54] contains 3000 separate handwritten Arabic numeral images. Each of them is a 32*32 pixel RGB image.

3) **MADBase**

The MADBase is a modified version of the ADBase benchmark with the same format as the MNIST benchmark [55]. MADBase is composed of 70,000 digits written by 700 writers. This dataset contains 60000 training and 10000 testing images, and each image is a 28*28 pixel RGB.

### B. Experiments on IFN/ENIT dataset

1) **Network training and testing**

As shown earlier, the IFN/ENIT included four sets. In this study, based on the work of [3], images of the same city name dataset were grouped in a folder. Therefore, the whole dataset was organized in folders instead of having raw data images. Each folder represented a class with a city name and contained all the city's images written by different writers.

The training set of IFN/ENIT was split, and 30% of the training set was considered as a validation set. Table 3 shows the number of training epochs in various experiments.

<table>
<thead>
<tr>
<th>Training/testing</th>
<th>Recognition accuracy</th>
<th>Number of epochs</th>
</tr>
</thead>
<tbody>
<tr>
<td>abc/d</td>
<td>98.7</td>
<td>20</td>
</tr>
<tr>
<td>abed/e</td>
<td>98.03</td>
<td>30</td>
</tr>
</tbody>
</table>

The recognition of handwritten Arabic characters was performed at a significant speed using the proposed network in the early training steps. This significant speed was due to the stages of inhibition in the proposed network. The silenced weak granule cells reduced the number of subsequent operations of the proposed network in the later steps. Only the weights of granule cells that survived from inhibition were updated, so the network's training was accelerated. Table 4 indicates the number of granule cells before and after the inhibition step in the proposed network.

<table>
<thead>
<tr>
<th>Layers</th>
<th>Number of GCs</th>
<th>Number of GCs after the inhibition stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>11</td>
</tr>
<tr>
<td>Total</td>
<td>45</td>
<td>31</td>
</tr>
</tbody>
</table>

A network with 15 granule cells in three layers was trained. That is, there were a total of 45 granule cells in this network. As it is clear from Table 4, the number of granule cells were reduced drastically after the interneurons' inhibition. After the inhibitions, the number of granule cells was reduced as much as 31.11% of all granule cells. Due to this reduction, recognition of handwritten Arabic words was performed in the early phase of training. This significant result can be attributed to the proposed network structure, which uses one inhibition step and two excitation steps. This feature of the proposed network decreases the epochs of the learning stage.

2) **The generalized capability of the proposed method**

Noise is one of the essential issues in handwritten recognition. In this study, generalization capability was analyzed using noisy test data. Various levels of noise were applied to the test set using pepper noise. The proposed network was compared for handwritten Arabic with and without noise. Table 5 illustrates the results of this assessment for the noise level of 10.
Table 5 indicates that the recognition was accomplished with high accuracy despite applying noise to the test samples. Given these results, the accuracy of recognizing an image was not significantly decreased by noise.

3) Comparison with other methods

As stated earlier, convolutional neural networks are among the most prominent deep learning methods used to feature extraction and classification. In this subsection, the proposed architecture (HMM-CNN-proposed network) is compared with (CNN-CNN) architecture. The results are shown in Table 6.

Table 6 shows the superiority of the proposed architecture in recognition of handwritten Arabic. For better evaluation, the proposed network was compared with other methods. These methods can be divided into two classes. The first class uses CNN as a feature extractor, and the second class performs CNN as a classifier. The experiment was performed based on sets a, b, and c as training sets, while set d was used for testing. Table 7 illustrates a comparison of the proposed network with other methods. As demonstrated in Table 7, five methods were compared with the proposed network. Three of the compared methods used CNN as a feature extractor. These methods used well-known classifiers: SVM, AlexNet, HMM, Dynamic Bayesian Network, and BLSTMCCTC. In contrast, two of the compared methods used CNN as a classifier. The former used PCOH as a feature extractor, and the latter used CNN as both a feature extractor and a classifier. The results show that the proposed network outperforms all the compared methods by an accuracy recognition of 98.7%.

C. Experiments on CMATERdb3.3.1 dataset

As stated above, this dataset contains 3000 separate handwritten numeral images. Training of the proposed network was performed using 2000 images. The rest of the image samples were used for the test. Similar to the process in [57], images were inverted before feeding into the proposed network. As a result, the numerals were in a white foreground on the backdrop of the black background. Since edges are a very important feature in character recognition, the black background makes edge detection easier. The result of the experiment showed a good accuracy recognition of 99.8 using this dataset.

D. Experiments on MADBase dataset

The black background makes the edge detection more straightforward in digit recognition problems. Similar to what was used for experiments on CMATERdb3.3.1, the images were inverted before being fed into the proposed network. Sixty-thousand samples of this dataset were used for training. The test was performed using 10000 samples, and the accuracy of 99.7 was achieved by the proposed network. Table 8 shows the comparison of state-of-the-art methods with the proposed method. As represented in Table 8, promising results for recognizing digit samples were achieved by the proposed network compared to other CNN-based methods.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Feature extraction</th>
<th>Handwritten Arabic recognition</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>[59] Maalej, Kherallah 2018</td>
<td>CNN</td>
<td>BLSTM-CTC</td>
<td>92.21</td>
</tr>
<tr>
<td>[34] Elleuch et al. 2016</td>
<td>CNN</td>
<td>SVM</td>
<td>92.95</td>
</tr>
<tr>
<td>[35] Amrouch, Rabi 2017</td>
<td>CNN</td>
<td>HMM</td>
<td>88.95</td>
</tr>
<tr>
<td>[55] Sudholt, Fink 2016</td>
<td>PCOH</td>
<td>CNN</td>
<td>92.14</td>
</tr>
<tr>
<td>[28] Poznanski, Wolf 2016</td>
<td>CNN</td>
<td>CNN</td>
<td>97.07</td>
</tr>
<tr>
<td>Proposed network</td>
<td>CNN</td>
<td>DG OF HIPPOCAMPUS</td>
<td>98.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reference</th>
<th>Dataset</th>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>[56] Ashiquzzama n, Tushar 2017</td>
<td>CMATERdb 3.3.1</td>
<td>CNN</td>
<td>97.4</td>
</tr>
<tr>
<td>[57] Ashiquzzama n et al. 2019</td>
<td>CMATERdb 3.3.1</td>
<td>CNN</td>
<td>99.4</td>
</tr>
<tr>
<td>[58] Finjan 2020</td>
<td>MADBase</td>
<td>CNN</td>
<td>99.6</td>
</tr>
</tbody>
</table>
V. RESULTS AND DISCUSSION

The proposed network for the recognition of handwritten Arabic digits is modeled based on the computational function of the DG in the brain. HMM was used for character modeling, and feature extraction was performed using a convolutional neural network. Deep learning requires large datasets for training because of the huge number of parameters needed to be tuned by a learning algorithm [2]. A huge variety of styles for Arabic characters makes handwritten Arabic word recognition more challenging. In this study, the capability of the DG of the brain is modeled for solving the problem with handwritten Arabic words. Separating patterns with high overlaps are attributed to the DG of the brain. Therefore, modeling the functionality of pattern separation in this part of the brain can help us to recognize handwritten Arabic characters with high overlaps. The proposed network has two excitation stages and one inhibition stage. Inputs and mossy cells excite the granule cells. In contrast, interneurons inhibit them. The final activity of granule cells after excitations and inhibition was considered as the output of the network. These special excitations and inhibition phases are considered as one of the prominent features of the proposed network. These excitation and inhibition steps play an essential role in the accuracy of handwritten Arabic recognition. Some of the granule cells, the most important cells in the proposed network, are silenced due to the inhibition phase in the proposed network. So, the number of remaining operations in the network was decreased. As is observed, this reduction enhances the speed and accuracy of the recognition. The results show that the total reduction of granule cells reaches approximately half. As depicted in Table 4, the number of 45 granule cells in the three layers is reduced to 11, 9, and 11, respectively. The results of the experiments on the IFN/ENIT dataset show the superiority of the proposed network as compared with other CNN-based methods. The proposed network was also tested against noisy handwritten Arabic words. The results indicated that the proposed network could recognize various handwritten Arabic noisy images with acceptable accuracy.

The proposed network was compared with other CNN-based methods. Primarily, three methods were used for comparison: 1) methods that used CNN for feature extraction and other classification methods. 2) methods that extracted features using various methods and utilized CNN for classification. 3) methods that used CNN both for feature extraction and classification. When CNN was used as a feature extractor together with Bidirectional Long Short-Term Memory (BLSTM) and was followed by a Connectionist Temporal Classification layer (CTC) as a classifier, the accuracy of 92.21% was achieved [59]. When CNN was used as a feature extractor and SVM was used as a classifier (Elleuch et al. 2016), the accuracy was improved up to 92.95%. The accuracy recognition of 88.95 was reported when CNN was used as a feature extractor, and HMM was used as a classifier [35]. When CNN was employed as a classifier for feature extraction using the Pyramidal Histogram of Characters (PHOC), the achieved recognition rate was 92.14%. The best result has previously been reported by [28] as 97.07. The current study used CNN for feature extraction and classification of handwritten Arabic characters. The results in Table 7 show that the proposed network is more accurate than all the compared methods. The achieved accuracy in the proposed network boosted the best method up to $(2.3 - 1.3) / 2.3 \times 100 = 43.47\%$ on IFN/ENIT dataset. In addition, the proposed network was evaluated using CMATERdb3.3.1 and MADBase datasets. Significant results were achieved on these two Arabic digit datasets. According to the experimental results given in Table 8, the proposed network boosted the state-of-the-art methods up to $(0.6 - 0.2) / 0.6 = 66.6\%$ and $(0.4 - 0.37) / 0.4 \times 100 = 25\%$ on CMATERdb3.3.1 and MADBase datasets, respectively. In addition to measuring the accuracy of recognition, the confidence interval and CER for all the experiments were computed based on [60]. Table 9 shows the character error rate values and the confidence interval on all experimental results. Experimental results and statistical analyses show that quite a uniform confidence interval was achieved for all the experiments performed on the used datasets.

Standard deviation is mostly used in the area of research and is regarded as a very good measure of dispersion in a series. In this study, the mean and the sample standard deviation and a confidence interval of 95% (statistical significance level of 5%) were calculated for the experimental results, as represented in Table 10. Besides, a pair of t-tests at a significant level of 0.05 was performed on all of the used datasets.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>CER</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFN/ENIT</td>
<td>1.63±0.33</td>
<td>(1.3, 1.97)</td>
</tr>
<tr>
<td>CMATERdb3.3.1</td>
<td>0.35±0.15</td>
<td>(0.2, 0.5)</td>
</tr>
<tr>
<td>MADBase</td>
<td>0.45±0.15</td>
<td>(0.3, 0.6)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dataset</th>
<th>p-value</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFN/ENIT</td>
<td>0.00220</td>
<td>0.4738</td>
</tr>
<tr>
<td>CMATERdb3.3.1</td>
<td>0.00095</td>
<td>0.2121</td>
</tr>
<tr>
<td>MADBase</td>
<td>0.00073</td>
<td>0.1626</td>
</tr>
</tbody>
</table>

The null hypothesis was rejected for all the performed experiments. The results of these tests are represented in Table 10. All these results show the superiority of the proposed network.
The efficient architecture proposed in this study comprises HMM for character modeling, CNN for feature extraction, and an intelligent DG-based network for handwritten Arabic recognition. The proposed network is modeled based on the functionality of pattern separation in the DG of the brain. Separating patterns with high overlaps is the responsibility of a part of the brain referred to as the DG. The brain's modeling capability attributed to this region can enhance handwritten Arabic recognition because of its inherent challenging problems. Gathering the convolutional neural network with an intelligent DG-based network can be helpful for the recognition of handwritten Arabic with many overlaps. The proposed network performs two excitation stages and one inhibition stage, thereby enhancing the accuracy of recognition. In the inhibition stage, some of the cells are made silent. These silenced cells are removed from both the operations and the training phase of the proposed network. Hence, the number of remaining operations is decreased. This reduction plays an essential role in the accuracy and the number of epochs in the proposed network. Good accuracy was achieved using three well-known datasets resulting in the least CER and uniform confidence intervals. Moreover, the results proved the proposed network's capability to recognize noisy handwritten Arabic images at various noise levels.

For future work, the functionality of CA3 in the hippocampus can be combined with the functionality of the dentate gyrus. CA3 plays an important role in pattern completion. A combination of pattern separation and pattern completion can be used in the recognition of damaged word images. We also are going to evaluate the performance of our proposed network by substituting other methods instead of CNN for feature extraction.

REFERENCES


She has also awarded to Tonic Inhibitory Control of Dentate Gyrus Granule Cells by α-Containing GABA Receptors Reduces Memory Interference, JNeurosci, 2015, vol. 35, no. 40, pp. 13698-13712.


Nazal Modhej received her B.Sc. degree in Computer Engineering from Shahid Chamran University, Ahvaz, Iran, in 2005 and M.Sc. degree in Computer Engineering from Science and Research Branch of the Islamic Azad University, Ahvaz, Iran, in 2010 and Ph.D. degree in Computer Engineering at Islamic Azad University Karaj branch, Karaj, Iran in 2021. She is the faculty member of the Islamic Azad University of Susangerd from 2005. Her research interest includes Intelligent Systems, Neuroscience and Learning.

Mohammad Teshnehlab received his B.Sc. degree from Stony Brook University, USA, in 1980, the M.Sc. degree from Oita University, Japan, in 1990, and the Ph.D. degree from Saga University, Japan, in 1994. He is a faculty member of Electrical Eng. Department of K. N. Toosi University of Technology. Professor Teshnehlab is a member of the Industrial Control Center of Excellence and founder of Intelligent Systems Laboratory (ISLab). He is also a co-founder and member of Intelligent Systems Scientific Society of Iran (ISSSI) and a member of the editorial board of the Iranian Journal of Fuzzy Systems (IJFS), International Journal of Information & Communication Technology Research (IJICTR), and Scientific Journal of Computational Intelligence in Electrical Engineering. His research areas are Artificial Rough and Deep Neural Networks, Fuzzy Systems and Neural Nets, Optimization, and Expert Systems.

Azam Bastanfard received her B.Sc. in Mathematics Applied in The Computer at Islamic Azad University Karaj branch in 1997 and M.Sc. degree in Computer Engineering from the Tokyo Institute of Technology in 2001 and Ph.D. degree in Computer Science from the Tokyo Institute of Technology in 2004. She had a Postdoctoral position at MIRALab, University of Geneva. Since 2006 she is an Assistant Professor at the Islamic Azad University of Karaj. Her research areas are Artificial Intelligence, Computer Games, Image Processing and Multimedia. She has also awarded “The national elite woman.”

Somayeh Raeisdana received her B.Sc., M.Sc., and Ph.D. degrees in Biomedical Engineering from Amirkabir University of Technology, Tarbiat Modares University, and Science and Research Branch of Islamic Azad University in 2003, 2005 and 2010, Tehran, Iran respectively. She is an Assistant Professor at the Islamic Azad University of Qazvin. Her research interest includes the Brain and Neuroscience.