Online Persian/Arabic Writer Identification using Gated Recurrent Unit Neural Networks

Mahsa Aliakbarzadeh[1], Farbod Razzazi[2*]
1- Department of Electrical and Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran.
Email: mahsaaliakbarzadeh@gmail.com
2- Department of Electrical and Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran.
Email: razzazi@srbiau.ac.ir (Corresponding author)

Received: March 2020 Revised: May 2020 Accepted : June 2020

ABSTRACT:
Conventional methods in writer identification mostly rely on hand-crafted features to represent the characteristics of different handwritten scripts. In this paper, we propose an end-to-end model for online text-independent writer identification on Persian/Arabic online handwritten scripts by using Gated Recurrent Unit (GRU) neural networks. The method does not require any specific knowledge for handwriting data analysis. Because of the exclusive ability of deep neural networks, we just represented our data by Random Strokes (RS) representations, which are differential horizontal and vertical coordinates extracted from different handwritings with a predefined length. This representation is a context independent representation. Therefore, this writer identification at RS level is more general than character level or word level in identification systems, which require character or word segmentation. The RS representation is then fed to a GRU neural network to represent the sequence for final classification. All RS features of a writer are then classified independently, and in the final stage, the posterior probabilities are averaged to make the final decision. Experiments on KHATT database, which consists of online handwritings of Arabic writers, gave us 100% accuracy on 10 writers and 76% accuracy on 50 writers, which is much better than previous works on online Persian/Arabic writer identification.

KEYWORDS*:* Handcrafted Features, End-to-End Identification, GRU, Online Writer Identification.

1. INTRODUCTION

Writer identification is a behavioral biometric system competing with voice and face identification systems [1]. The aim of a writer identification system is to identify the writer of a text among a number of writers which has previously been introduced to the system. Writer identification should not be mistaken by writer verification or handwritten recognition systems. Contrary to writer identification, in writer verification, the goal is to answer a yes/no question and verify a specified writer. We also have handwritten recognition which recognizes the text and has nothing to do with writer identification [2], [3].

Writer identification systems can be divided into offline and online models. The main difference in online and offline models is the input data. In offline models which are the prior approaches in writer identification, the input is the images of a text ;while, in online models; in addition to the text image, the temporal information (e.g. x-y coordinates, pressure, azimuth, and altitude) are available. Therefore, better results might be achieved in the latter approach [4], [5].

Writer identification systems can also be categorized into text dependent and text independent systems. Text dependent writer identification systems give better results as the input should be exactly the same for all writers; while, in text independent writer identification systems, as there is no restriction in text content, it is hard to achieve high accuracy, and of course, text independent models are more practical [6], [7].

Writer identification systems can be widely used in forensic and bank authentication applications. In addition, by growing the usage of pen-based and touch-based gadgets, online writer identifications are becoming more popular than offline models.

Deep temporal neural networks have recently been used for time series identification and behavior analysis. Their results have been very glancing in different applications. However, generally, using these types of neural networks are computationally very complex.

In this paper, we present a novel high performance, computationally light model for online writer identification using Gated Recurrent Units. The structure of this paper is as follows: In the second section, the related state of the art is briefly studied. In the third

Paper type: Research paper
DOI: https://doi.org/10.29252/mjee.14.3.9
How to cite this paper: M. Aliakbarzadeh, F. Razzazi, **“Online Persian/Arabic Writer Identification using Gated Recurrent Unit Neural Networks”**, *Majlesi Journal of Electrical Engineering*, Vol.14, No. 3, pp. 73-79, 2020.

section, the proposed method is presented. The results are provided in the fourth section and they are compared with the previous methods. In the final section, the paper is concluded.

2. RELATED WORK

Previous work in writer identification systems have mostly used handcrafted features for feature extraction. The features of handwriting identification systems can be divided into textural-based and grapheme-based features. Textural-based features are the statistical information about the slant and curvature of the text; while in grapheme-based feature extraction, local information, (e.g. the way someone write a letter) are extracted [8],[9].

In online approaches, temporal parameters like horizontal and vertical coordinates and pen pressure are used to extract the features vector. In a study, the histograms of velocity, pressure and azimuth of different strokes were used as textural features [10]. Another study have used shape primitive features which give direction and curvature of the contour [6]. In another study, allograph prototype matching has been used as a grapheme-based feature extraction approach [3]. Although local features has achieved higher accuracy, they require a segmentation technique as a preprocessing step.

In addition to hand-crafted feature extraction models, new studies in both online and offline writer identification have employed deep neural networks to dispose the usage of sophisticated handcrafted features. In 2016, Weixin Yang et.al have used convolutional neural networks for online writer identification in English language and reached 98.5% accuracy on 134 writers [11]. As Covolutional Neural Network (CNN) is suitable to deal with fixed-size images, a stroke and character segmentation stage has been required as a preprocessing step. Therefore, There have still been some hand-crafted features to achieve high accuracy for CNN-based writer identification systems. In a recent study in offline writer identification, CNN has been used as an end-to-end approach. They achieved 99.9% accuracy on 100 offline writers [12].

Xu-Yao Zhang et.al have proposed a completely end-to-end writer identification system by using a Long Short Term Memory(LSTM) recurrent neural network. In that study, data is represented by Random Hybrid Stroke (RHS) representation. Then, RHS data is fed into a bidirectional LSTM to extract useful features. The decision is devoted to an ensemble-based model. It has reached 100% accuracy on 133 English writers and 99.5% accuracy on 186 Chinese writers [13].

There are a few research in Arabic/Persian online writer identification. In a research carried out by Tamour Dib et al, Beta Elliptic algorithm was used to extract a handcrafted features vector. They classified the features vector by a multi layer feed forward neural network. They have achieved 91% accuracy on only 19 writers [14].

In the present study, a GRU neural network is employed to model the handwritten sequential data. By inspiring Xu-Yao Zhang work, we used GRU neural network to design an online Persian/Arabic writer identification system. First, from x-y coordinates of the pen tip, we acquired some context independent random RS sequence representations from the handwritten data of a writer, involved in handwriting. Moreover RS representations include data augmentation within it which is necessary for any deep neural network. At last, when every RS data is classified separately, an ensemble based model was used for final writer identification.

3. PROPOSED GRU-BASED WRITER IDENTIFICATION SYSTEM

In this section, the proposed model is discussed in detail. First, the RS method for data representation is described. Then we shortly formulate GRU structure mathematically for author authentication. The role of GRU is discussed as a feature extraction module. Finally, the ensemble-based method is utilized as the final classifier.

3.1. Data Representation

Electronic devices, such as tablet and smart phones that have pen capturing capability, can gather handwriting information including horizontal and vertical coordinates as well as azimuth and the pressure of the pen tip. Here, we only used x and y coordinates for RS representation. Therefore, for every sample of each writer, we have a long sequence of coordinates called S which can be formulated as:

$$S = [\, [x_1, y_1], [x_2, y_2], \ldots, [[x_n, y_n]\,], \tag{1}$$

Where, x_i and y_i are coordinates of the pen in the i[th] sample. We can transform S (point-level) into another differential sequence ΔS, which shows the direction of the pen movements:

$$\begin{aligned} \Delta S = [\, & [x_2 - x_1 , y_2 - y_1] , \\ & \ldots [x_i - x_{i-1} , y_i - y_{i-1}] , \\ & \ldots [x_n - x_{n-1} , y_n - y_{n-1}] \,] \end{aligned} \tag{2}$$

ΔS can be a text line or a text page. As it may be a very long sequence, we may not have sufficient text for every writer to train our deep neural network. Therefore, we randomly sample multiple short length continuous subsequences from each ΔS and call it as an RS sequence, standing for a random stroke. Normally, the word stroke in writer identification systems denotes a segment between a pen-down and the corresponding pen-up. However here, the strokes show a differential predefined fixed length sequence between two sampling points.

$$RS= [[\Delta x_0, \Delta y_0], \ldots, [\Delta x_i, \Delta y_i], \ldots, [\Delta x_k, \Delta y_k]] \quad (3)$$

We can obtain as many RS samples as we want from each text to train our deep neural network. Each RS sample is a vector of the fixed length N, which means that our GRU model input would be an Nx2 matrix.

It is clear that RS level representation does not require any domain knowledge of handwriting, so it can be considered as a text and language independent method.

3.2. Feature Extraction by GRU Neural Network

Recurrent Neural Networks are special artificial neural networks which can model dynamic time behavior; Therefore, they can be used as a modeling tool for sequential data (e.g. speech recognition, handwriting recognition and image captioning). In a Recurrent Neural Network (RNN), a hidden state is produced which results in a hidden vector at each time step. The hidden state h_i is a function of input x_i as well as the previous state h_{i-1} :

$$h_t= f(x_t,h_{t-1}) \quad (4)$$

For every time step, by passing h_t through a soft-max function, the output y_t is generated and we can also generate the vector $[y_1,y_2,\ldots, y_k]$ which can be used in sequence-to-sequence applications. However, in writer identification, we only need a fixed length vector of the last state.

Vanilla RNN cannot be used for long time sequences due to vanishing gradient problem. As shown in Fig.1, GRU as an extension of recurrent neural network, is robust against vanishing gradient phenomenon and can learn long term dependencies according to (5) to (8):

$$z_t = \sigma_g (W_z x_t + U_z h_{t-1} + b_z) \quad (5)$$

$$r_t = \sigma_g (W_r x_t + U_r h_{t-1} + b_r) \quad (6)$$

$$\dot{h}_t = \tanh(W_h x_t + U_h(r_t \odot h_{t-1})+ b_h) \quad (7)$$

$$h_t = z_t \odot h_{t-1} + (1- z_t) \odot \dot{h}_t \quad (8)$$

Where, x_t is input vector, h_t is output vector, z_t is update gate vector, r_t is reset gate vector and W, U and b are parameter matrices and vector, also the operator $\odot$ denotes the Hadamard product.

For automatic feature extraction, we exploit a GRU neural network stage to extract the features without any background information about useful information for classifying writers. Although we experimented different extensions of GRU including bidirectional and multilayer GRU, vanilla GRU was just enough to reach a reasonable accuracy. By passing the input vector, through GRU network, a feature vector with Nx2 dimension of the last state was fed into a fully connected layer for final classification. The final fully connected layer has only one layer with the number of neurons equal to the number of classes. In addition, we used sigmoid as the nonlinear function for final classification. All weights and biases for both GRU and fully connected layer were initialized by random value drawn from a zero-mean Gaussian distribution function. The cell and the hidden state of GRU were initialized as zero. In addition, to train our network, we used backpropagation algorithm with ADAM optimization function, which is based on adaptive estimation of lower order moments. It is clear that there is not enough explicit information for a simple classifier to classify the writers in RS representation. However, trusting GRU ability to extract suitable temporal features to discriminate writers, we did not have to extract handcrafted features as it has been performed in previous methods. The schematic of our model is depicted in Fig. 2 where the process of described writer identification is briefly shown.

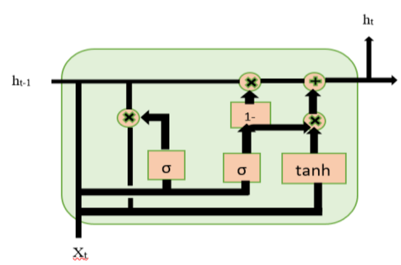

Fig. 1. GRU Internal Structure.

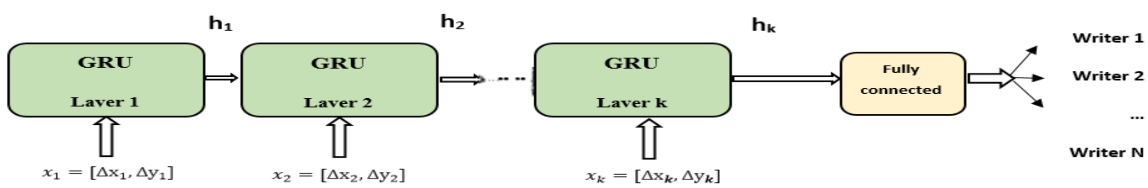

Fig. 2. The Proposed GRU Architecture Model.

3.3. Ensemble based method

For every registered writer, we have a text or some lines of their online handwriting data which is then transferred into a set of RS representation. We can obtain as many RS samples as we want by randomly sampling each handwriting to have a large dataset of random strokes to train our deep neural network model. After trainig our model with these RS samples, every test data is classified seperately by the trained model. However, each of these RS representation can not represent the behavior of a writer adequately. Indeed, the information of an RS is too raw to be able to achieve a high performance writer identifier. Therefore, we used an ensemble based method to reach to the desired accuracy. In fact, each RS sample presents a short part of a handwritig which cannot discriminate a writer character. In ensemble based processing stage, a mapping of the posterial probablity functions of a set of RS samples are added up toghether to give an ensemble-based prediction, which is the final writer identification result. The schematic of ensemble-based method is represented in Fig. 3.

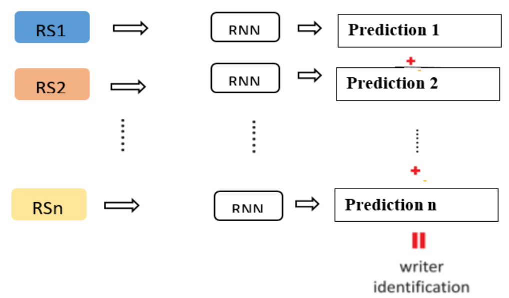

Fig. 3. Final writer identification based on ensemble based model.

As shown in Fig. 3, each RS data is fed to the RNN, then the outputs which are one-hot vectors are averaged to define the target writer. In our experiment by 90 RS data we obtained the desired result.

4. EXPERIMENTS AND RESULTS

In this section, first we show the impact of RS numbers in ensemble based decision. Next, we present the result of different model parameters to empirically optimize the model accuracy. Then we compare the results of our method with other approaches in writer identification.

4.1. Database

In this research, as there is not an appropriate Persian online handwriting database, we used Online-KHATT database, which consists of 10040 lines of online Arabic handwriting [15]. we selected 50 writers for our writer identification task in this dataset. The writers with sufficient length data in both training and test sets are selected for this purpose. In total, there are four or five lines available in average for each writer in the train data, and half of it for test data.

4.2. Implementation Details

For each of these 50 writers, we extracted 2000 RS sequences from the train set, and 1000 RS sequences from test set. In addition, each RS sequence is a 2xN vector. We selected the length of the vectors empirically according to Fig 4. The curve shows that the value of N equal to 100 and 110 are good choices. Moreover, we optimized the RNN size dimension empirically into 200. The optimization algorithm is ADAM with mini batch size of 128 and the learning rate is set to 0.001. Furthermore, we used sigmoid as the soft-max function in the fully connected layer. After optimizing the mentioned parameters, by an ensemble based method for evaluation, we reached the valuable accuracy of 76% in ONLINE KHATT dataset which outperforms other state of the art approaches in online Arabic writer identification.

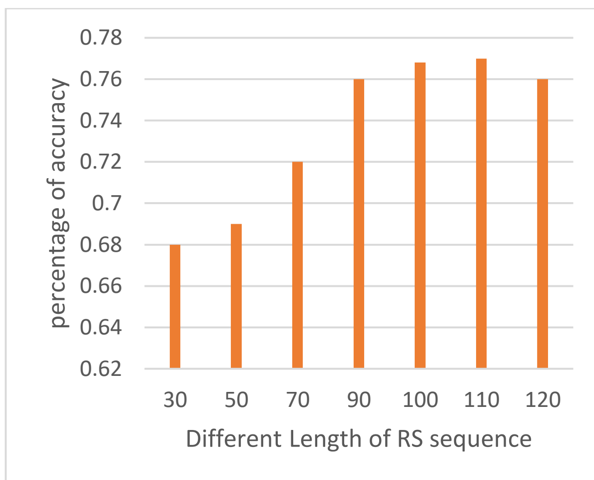

Fig. 4. The effect of RS length on accuracy.

4.3. The Effect of the Total Number of Random Strokes

By utilizing an ensemble based model, we need some RS samples in order to make the prediction. As shown in table 1, a few number of RS samples cannot discriminate different writers well. Hence, in order to identify each writer, at least a set of 90 RS samples is required to gain the maximum accuracy. As the length of each RS sample reaches 100, considering the fact that four or five RS samples is often extracted for each Arabic word, we require at least 15 words to identify the writer.

Table 1. The effect of RS numbers on accuracy.

Number of RS	RNN Size	Accuracy
5	200	44%
10	200	56%
30	200	64%
60	200	69%
90	200	76%
100	200	76%

4.4. Other Alternatives of GRU

In order to show the performance of GRU, we substituted GRU neural network with two other networks. In the first experiment, we used SVM as an ordinary classifier. To have a fair comparison, we used the same training and test arrangements in the assessment. We reached the maximum accuracy of 5% on the test data, which is not comparable with the accuracy achieved by GRU. The reason that SVM classifier could not identify the writers is its disability in automatically extracting desirable features from the raw data.

We also replaced DCNN with GRU to see if convolutional neural network could perform as well as a recurrent neural network or not. After RS representation, as shown in Fig. 5, we fed it to a DCNN with four convolutional layers, each of which was accompanied by a Max pooling (MP) layer. The first layer consisted of 8 filters, the second layer had 16 filters, the third one had 32 filters, and finally the last one consisted of 64 layers. We fixed the size of the convolutional filters to 3×3 with a stride of 1 pixel. The window size of max-pooling was 2×2 (MP2) with a stride of 2 pixels. The results can be seen in Table 2. As it is shown in the table, although the accuracies were better than SVM classifier, they were not comparable with our GRU network.

This result might be due to the fact that CNN is designed to extract features from images. While, we fed the CNN a sequential data of the length 2x100, this was suitable for a recurrent neural network like GRU rather than a convolutional network. Furthermore, we decided to exploit CNN by feeding the synthesized image of our RS data to the network. In this approach, GRU can concatenate on the extracted features from CNN. Therefore, there are more abstract and enriched features in the GRU stage. We converted each RS sequence to an image of size 120*120 handwritten image and gave it to the proposed CNN. However, we could not achieve any noticeable result in this approach and the best outcome was only 45% on ten writers. That could be due to the RS structure which cannot segment letters in Arabic handwriting. Fig. 6 shows the sample images of each RS sequence.

In the last experiment, in order to assess the capability of the GRU network, we replaced GRU with an LSTM neural network, which is roughly the same as GRU except for it had the output layer, which makes it a little more complicated. Although the accuracy of the LSTM neural network was roughly the same as our GRU neural network (76% on 50 writers), its computational performance is not comparable, in both training and testing phases. LSTM Training time for 100000 RS samples on a corei3-5005u 2GHz processor and Intel(R) HD Graphics 5500 GPU took 23 hours. The testing time was 20ms for each sample, while the training time with GRU was about 13 hours and the test time was only 13ms. The result of alternative replaced architectures can be seen in Table 3.

Table 2. Accuracies achieved by CNN (Feeding RS vector as the input).

Number of Classes	accuracy
10	92%
20	51%
30	38%
40	12%
50	8%

Table 3. Accuracies Achieved by Alternating Different Networks with GRU.

Alternative Network	Accuracy(on 50 writers)
SVM	5%
DCNN	8%
DCNN(RS Image Representation)	26%
LSTM	76%
GRU	**76%**

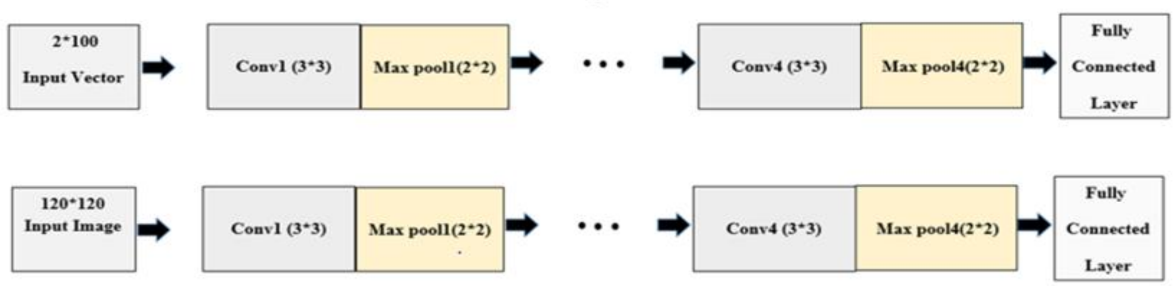

Fig. 5. The structure of proposed convolutional neural network (once each RS vector is fed to CNN. The corresponding 120*120 image is fed to CNN).

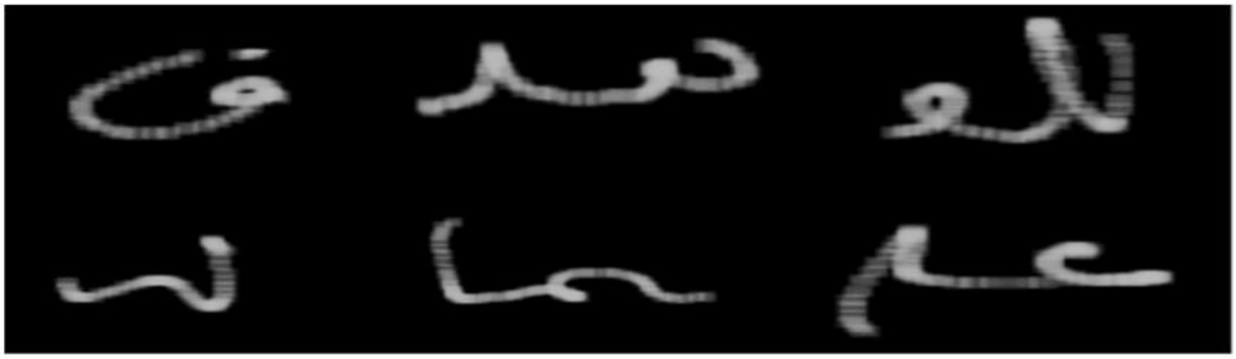

Fig. 6. Visualization of an RS sequence (each RS sample is a vector of length 100 consisting x and y coordinates).

4.5. Comparison with Other State-of-the-Art Approaches

In Table 4, there is a list of different online writer identification systems on different datasets. As handcrafted features cannot guarantee the extraction of optimum features, the accuracy achieved by featured-based approaches is lower than deep neural network approaches. Although we could not reach the accuracy achieved by the state of the art approaches on English dataset; however, our method outperforms the previous methods on Arabic datasets. In addition, the implementation complexity of the approach is very low as there is no need for any sophisticated preprocessing on the dataset. In fact, the research on online Arabic writer identification is very rare. One of them, which has been conducted by Mariam Gargouri et.al could reach the accuracy of 12.7% on 55 writers, and the last research which has been done by Dhieb had 91% accuracy on only 19 writers [15], [16].

5. CONCLUSION

In this research, we developed an end-to-end writer identification approach on Persian/Arabic handwritten scripts based on GRU deep recurrent neural network. The method was more efficient than previous methods on online Persian/Arabic writer identification both on accuracy and complexity aspects. We noticed that GRU neural network outperforms convolutional neural networks as the feature extractor for handwriting trajectory images. In fact, this was the first step in utilizing GRU deep neural network in Persian/Arabic online writer identification. For future research, we are going to optimize this structure for writer verification application and script recognition tasks. Furthermore, by using an appropriate augmentation method, this method can be used for signature verification task.

Table 4. Comparison of different writer identification approaches.

Writer	Accuracy	Database	Approach	Language
Liwicki et al	82%	BIT(133C)	Handcrafted features	English
Liwicki et al	80%	BIT(187C)	Handcrafted features	Chinese
Yang et al.	98.5%	BIT(133C)	DCNN	English
Zhang et al	100%	BIT(133C)	BDLSTM	English
Our method	100%	KHATT(10C)	GRU	Arabic
	76%	KHATT(50C)	GRU	
Thameur Dhieb	91%	ADAB(19C)	MLP Beta Eliptic model	Arabic
Mariem GARGOURI	12.7%	ADAB(55C)	Point based feature.	Arabic

REFERENCES

[1] Emadi, M., & Jafar Pour, M., " **Improving Human's Finger Knuckle Identification using High Order Zernike Moments**", *Majlesi Journal of Electrical Engineering*, Vol. 14(1), pp. 119-125, 2020.

[2] V.Moradi, F.Razzazi, A.Behrad " **Recognition of Handwritten Persian Two-digit Numerals Using a**

Novel Hybrid SVM/HMM algorithm", *Majlesi Journal of Electrical Engineering*, Vol. 10(3), 2016.

[3] R. Hanusiak, L.S. Oliveira, E. Justino, R. Sabourin, "**Writer verification using texture-based features, International Journal on Document Analysis and Recog- nition** ", pp. 213–226.580, 2012.

[4] Li. Bangy, Zhenan Sun, Tieniu Tan, "**Hierarchical shape primitive features for online text-independent writer identification**", *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, pp. 986-990,2009.

[5] M. Bulacu, L. Schomaker, A. Brink, "**Text-independent writer identification and verification on offline Arabic handwriting**", *in: International Conference on Document Analysis and Recognition*, pp. 769–773, 2007.

[6] Li, Bangy, Zhenan Sun, Tieniu Tan, "**Hierarchical shape primitive features for online text-independent writer identification**", *Proceedings of the International Conference on Document Analysis and Recognitio*n, ICDAR, pp. 986-990,2009.

[7] X. Wu, Y. Tang, W. Bu," **Offline text-independent writer identification based on scale invariant feature transform**", *IEEE Transactions on Information Forensics and Security*, Vol. 9 (3), pp. 526–53, 2014.

[8] Guoxian TAN, "**Writing style modelling based on grapheme distributions: application to on-line writer identification**", *Thesis in Nanyang Technological university*, 2013.

[9] D. Bertolini, LS Oliveira, E Justino, R Sabourin, "**Texture based descriptor for writer identification and verification** ", *in Elsevier*, 2013

[10] Li. Bangy, Zhenan Sun, Tieniu Tan, "**Online Text-Independent Writer Identification Based on Stroke's Probability Distribution Function**", *International Conference on Biometrics Springer*, Vol. 4642, pp. 201-210,2007.

[11] Weixin Yang, Lianwen Jin, Manfei Liu, "**Deep Writer ID: An End- to-End Online Text-Independent Writer Identification System**", In *IEEE Intelligent Systems*, 2016.

[12] Hung Tuan Nguyen, Cuong Tuan Nguyen, "**Text-independent writer identification using convolutional neural network**", *Pattern Recognition Elsevier*, 2019.

[13] Xu-Yao Zhang, Guo-Sen Xie,"**End-to-End Online Writer Identification With Recurrent Neural Network**", *IEEE transactions on human-machine systems*, Vol. 47, No. 2, 2017.

[14] Thameur Dhieb, Wael Ouarda," **Online Arabic Writer Identification based on BetaElliptic Mode**", 1*5th International Conference on Intelligent Systems Design and Applications*, At Marrakesh, Morroco, 2016

[15] http://onlinekhatt.ideas2serve.net/

[16] Mariem GARGOURI, Slim KANOUN, "**Text-independent Writer Identification on Online Arabic Handwriting**", *12th International Conference on Document Analysis and Recognition*, 2013