A STUDY ON FARSI HANDWRITING STYLES FOR ONLINE RECOGNITION

Vahid Ghods [1], *Ehsanollah Kabir* [2]
[1] Department of Electrical and Computer Engineering, Semnan Branch, Islamic Azad University, Semnan, Iran.
[2] Department of Electrical and Computer Engineering, Tarbiat Modarres University, Tehran, Iran.
E-mail: v.ghods@semnaniau.ac.ir (V.Ghods), kabir@modares.ac.ir (E. Kabir)

ABSTRACT

Knowing varieties of writing a letter in a word or a subword in different handwriting styles is very beneficial in recognition specifically for online recognition. In this paper, TMU-OFS dataset consisting of 1000 frequent Farsi subwords is employed to study Farsi handwriting styles. The subwords are grouped based on their delayed strokes and their main bodies, separately. The handwriting styles in this dataset are analyzed and the wrongly spelled or incorrect structural samples are extracted. Finally, the second version of the dataset is introduced by considering the handwriting styles. The preliminarily results show a significant improvement in recognition of subwords based on their styles.

Keywords: Online handwriting, Word recognition, Farsi, Persian, Writing variety.

1.0 INTRODUCTION

One of the most influential branches of pattern recognition is handwriting recognition. In the last three decades, a number of handwriting recognition systems have been developed for Farsi and Arabic [1- 2]. In offline Farsi/Arabic handwriting, different ideas were revealed. In offline Farsi/Arabic handwriting, different ideas were revealed. An integrated OCR system for offline Farsi script was proposed in [3] and the system used information from several knowledge sources and managed them in a blackboard approach. The system was able to recognize ten popular Farsi fonts using a Font Recognizer module. It also detected English words in between Farsi sentences. In [4], many different methods have been proposed and applied to various types of images. It has provided a comprehensive review of these methods. Neuro-fuzzy inference engine has been proposed in [5] to recognize the Farsi numeral characters. In [6], the HAH manuscript's image was segmented into words, and each word was segmented into its connected parts. Then, several structural and statistical features were extracted from these connected parts. Finally, a neural network was used to learn and classify the input vectors into word classes. In [7], a new method has been presented to recognize printed Arabic characters using single hidden Markov model (HMM). The features used in the HMM were based on the arcs of the skeleton of the words. The segmentation of offline handwritten Jawi script has been presented for character recognition systems and implemented in hardware in [8]. Histogram normalization and sliding windows were used for hardware implementation of real-time segmentation.

Online handwriting recognition (OHR) has attracted more and more researchers and industrial attention in recent years [9- 10- 11- 12]. OHR systems can be broadly grouped into three categories: heuristic or structure-based methods like decision tree [13] and fuzzy schemes, Template matching methods like dynamic time warping (DTW), and statistical and learning-based methods like HMM, SVM and neural networks [14]. A number of online Farsi/Arabic handwriting recognition systems have been proposed in [15- 16- 17]. An elastic fuzzy method has been suggested in [18] and tested on a set of 1250 words resulting in 78% and 96% correct recognition without and with a dictionary checking, respectively. In [19], structural features for a multi-font Arabic/Persian OCR were presented. Using structural features, Farsi characters were divided into nine groups in [13]. A decision tree was used for group recognition, and the accuracy rate of 94% was achieved. Taking advantage of cooperation between separated horizontal and vertical signals, x(t) and y(t), yielded better overall performance compared with using 'x-y' signal in [20].
There are a few Farsi handwriting datasets. [21] introduced a very large dataset of offline handwritten Farsi digits and 102,352 binary images were extracted from about 12,000 forms. In 2006, two offline Farsi databases, IFHCDB and Farsi_CENPARMI, were available for researchers [22- 23]. As far as we know, the only presented online Farsi dataset is TMU-OFS (Tarbiat Modarres Universit-Online Farsi Subwords) [24].

Writing styles and individuality can be used for handwriting classification [25], writer verification and identification [26], and examining forensic documents [27]. In this paper, styles and varieties of online Farsi handwriting are studied and analyzed on TMU-OFS dataset. Also, the second version of TMU-OFS dataset is introduced. The results of this study are extremely beneficial to Farsi OHR.

Section 2 of this paper provides an overview of TMU-OFS dataset. Section 3 describes the subword grouping using the main bodies of the subwords and their delayed strokes. In this section, ordinary HMM training is described, and recognition processes are deduced. In Sections 4 and 5, Farsi handwriting styles and the results with analysis on TMU-OFS dataset are presented. The paper ends with some concluding remarks in Section 6.

2.0 OVERVIEW OF TMU-OFS DATASET

TMU-OFS dataset is based on the texts extracted from six years of "Hamshahry" newspaper and one year of "Keyhan" published in Iran with the total words of 313,225. The authors of [24] selected the words with the occurrence of above 30 that yields 29,739 words. In Farsi, some of the letters in a word are joined to each other. These letters create a subword; for example, "فا, ر, سی" are the subwords of "فارسی". Subwords in a word are not joined to each other. Therefore, subword recognition is usable for word recognition. The number of the unique subwords was 7,317. Volito pen and tablet from Wacom were used to collect this data. The 1000 more frequent subwords were chosen for sampling. 124 persons, 98 men and 26 women, participated in the data collection. Half of them were in post-secondary education, and the other half were studying Bachelor and Master of Science. Eight persons were left-handed. Each person wrote 100 subwords. More than 10 samples were collected for each subword and this raised the total number to more than 10,000. In this dataset, the pen pressure was not recorded.

3.0 SUBWORD GROUPING

There are some methods for subword/word classification. Two methods used in this study are based on the delayed strokes and the main bodies.

3.1 Subword grouping based on the delayed strokes

One of the most notable differences between Farsi and English scripts is the delayed strokes. The stroke(s) after the main body, first pen stroke, represent(s) delayed stroke(s) that are added to the main body. In Farsi, dots, diagonal bar (Sarkesh), vertical bar, and hat are employed as the delayed strokes. This difference creates new ideas to group and to recognize Farsi subwords and words. Grouping based on the arrangement of delayed strokes was presented in [28]. For example, subwords "بیعت, بیت, بیتا, بیحقی, جمعیت" are in one group whereby the sequence of their delayed strokes is "single-dot-down, double-dots-down, double-dots-up". [28] used neural networks for group recognition based on delayed strokes features. Then the main body of input subword was compared to the main bodies in the relevant group. An accuracy rate of above 70% was achieved in subword recognition taking into account some simple constraints in writing dots. It was supposed that single-dot (•), double-dots (━) and triple-dots ( ) were written in one pen stroke.
Some relevant delayed strokes or signs in Farsi can be seen in Table 1. The distribution of sign codes of subwords in TMU-OFS dataset is given in Table 2.

Table 1. Main signs in Farsi script

Sign Code	Sign	Corresponding characters
1	Single-dot-up	Noon, Fe, Ghein, Zaa, Zaad, Ze, Zaal, Khe (ن - ف – غ – ظ – ض – ز – ذ – خ)
2	Single-dot-down	Jim, Be (ج – ب)
3	Double-dots-up	Ghaaf, Te

		(ت – ق)
4	Double-dots-down	Ye (يـ)
5	Triple-dots-up	Shin, Zhe, Se (ث – ژ – ش)
6	Triple-dots-down	Che, Pe (پ – چ)
7	Sarkesh	Kaaf (ک)
8	Double Sarkesh	Gaaf (گ)
9	Small bar	Taa, Zaa (ط – ظ)
10	Hamzeh	(أ – (ئ ، ئـ) - ؤ)
11	Hat	Alef with Hat (آ)

Table 2. Distribution of signs of subwords in TMU-OFS dataset

Sign code	No. of subwords	Percent (%)
1	427	27.8
2	181	11.8
3	303	19.7
4	240	15.6
5	122	7.96
6	56	3.66
7	97	6.33
8	44	2.87
9	48	3.13
10	13	0.85
11	1	0.06

For example, there are 122 letters having triple-dots-up in the 1000 subwords. In Farsi, below the main body of a letter only single-dot-down, double-dots-down and triple-dots-down signs can be applied, and the other signs appear only on the top of the main bodies.
In one experiment, we divided the subwords into four groups: "no sign", "all signs up", "all signs down" and "signs up and down". Table 3 illustrates the number of subwords in each group, based on the sign location.

Table 3. No. of subwords in each group, based on the sign location

Group type	No. of subwords
no sign	134
all signs up	453
all signs down	151
signs up and down	262

If we divide the 1000 subwords based on their delayed strokes as in [28], 170 groups are formed. In each group, the sequence of the delayed strokes is the same, e.g. the sequence of the group, "جمعیت, بیحقی, بیتا, بیت, بیعت", is "single-dot-down, double-dots-down, double-dots-up". After "no sign" group that has 134 members, "single-dot-up" group has the most members. Table 4 shows the number of members for the most popular groups.

Table 4. No. of subwords in 5 most populated groups

Rank	Group	No. of subwords
1	Single-dot-up	105
2	Double-dots-up	71
3	Single-dot-down	59
4	Double-dots-down	45
5	Single-dot-up—Double-dots-up	29

Some subwords like "نچه" with "single-dot-up, triple-dots-down" and "کمیته" with "one-diagonal bar, double-dots-down, double-dots-up" have unique signs in the dataset. Therefore, these groups have one member. The frequencies of low member groups are shown in Table 5.

Table 5. No. of the least populated groups

No. of subwords in a group	No. of groups
1	80
2	35
3	10

3.2 Subword grouping based on the main body

The main body of a subword is usually written in the first stroke. We formed groups with the same main bodies, for example, subwords "تیر ، تبر ، پیر ، ببر". In this way, 655 groups were created. The details are shown in Table 6.

Table 6. Distribution of number of subwords in each group

N	12	9	7	6	5	4	3	2	1
M	2	2	3	7	8	20	25	112	476

N: Number of subwords in a group, M: Number of groups with N subwords

For example, N=5 and M=8 means that there are 8 groups that each one has 5 subwords having the same main body. In TMU-OFS dataset, there are 476 groups, M=476, with unique main body of the subwords, N=1.
HMM classifier is used for offline [29] and online [30] handwriting recognition. We did a preliminary experiment for group recognition. Only the groups with N>=4 were included in this experiment, because training of HMM needs enough samples. The number of these groups was 42, and the average number of samples in each group was 54. We used 70% of these samples for training and the rest for the test. The models of all the main bodies were calculated with the number of the states = 10 and the number of the Gaussian mixtures = 10. The accuracies of group recognition were 71% and 86% for the test set and the training set, respectively. In HMM training, the number of samples is very serious. The far values between the test set and the train set accuracies illustrated that the number of samples were not so enough. Altogether, this acceptable accuracy was achieved only for the first 42 groups (N>=4). If there were sufficient samples for other groups, this method would reveal satisfactory results.
We investigated the test samples to understand the main reason of errors in classification. One crucial error observed in the samples is that they included specific letters, e.g. "س", "ط", "ع", "ک", "م" and "ه". This error occurred because of the writing variety in the letters. One example of writing variety is shown in Table 8. All five samples of subword "سعه" existed in one group. Samples number 3, 4 and 5 of Table 8 were not recognized in their group. If the subwords using these letters separate into two or more distinct groups, two objectives are achieved, reducing errors in recognition and improving preciseness of HMM model in the training phase because of appropriate samples. Also, study on writing variety is appropriate in subword grouping based on the delayed strokes discussed in Section 3.1, because the main body of input subword is a principal criterion for final recognition. Therefore, in order to recognize handwriting especially for online recognition, varieties of usual handwriting styles should be investigated.

4.0 FARSI HANDWRITING STYLES

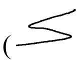
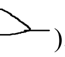
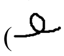
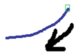
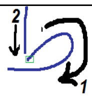
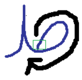
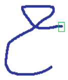
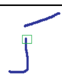
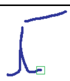
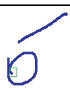
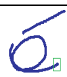
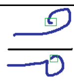
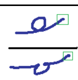
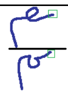
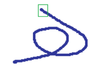
In Farsi, the handwriting style of some letters in subwords (or words) particularly in the middle of subwords has changed, and the main bodies of letters are written in two or more styles. For example, middle "He" ("ـهـ") is normally written like (), but sometimes like ().
Another example is letter "Kaaf" ("ک") before one of the letters "Alef, Laam, Kaaf" ("ا ، ل ، ک") where its shape normally transforms from () to (). These transformations are created because of easy and fast writing.
Traditional style in writing letter "Taa" ("ط") is that its vertical bar appears as the second stroke, while, when a subword starts with "Taa", this bar can be written first (). In another style, "Taa" is written in a single stroke, starting with an oval shape continued by a vertical bar written by moving the pen upward and then downward ().
Different styles of writing occur in letters "Mim, He, Sin, Eein" ("م ,ه ,س ,ع") regarding their place in subwords. In Table 7, the most famous styles of some letters are presented.

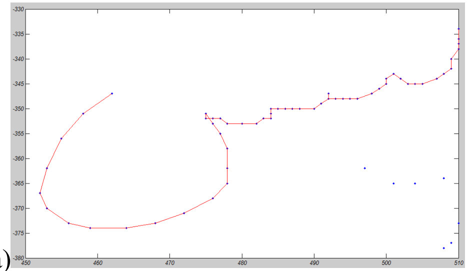
Table 7. Varieties of Farsi handwriting styles (☐ is the start point)

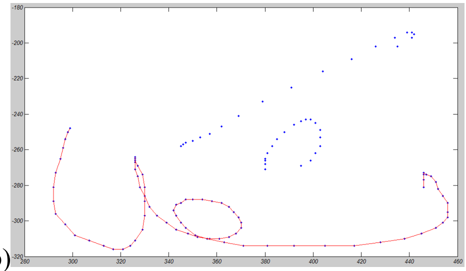
Letter	Beginning place	Middle place	Ending place
Sin (Shin)-Style 1 س(ش)			
Sin (Shin)-Style 2 س(ش)			
Taa (Zaa)-Style 1 ط(ظ)			
Taa (Zaa)-Style 2 ط(ظ)			
Taa (Zaa)-Style 3 ط(ظ)		-	-
Ein (Ghein)-Style 1 ع(غ)			
Ein (Ghein)-Style 2 ع(غ)	-		
Kaaf (Gaaf)-Style 1 ک(گ)			
Kaaf (Gaaf)-Style 2 ک(گ)			-
Mim-Style 1 م			
Mim-Style 2 م			

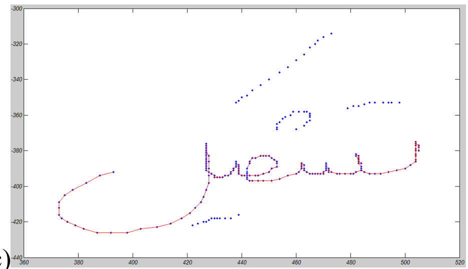
He-Style 1 ه			
He-Style 2 ه	-		

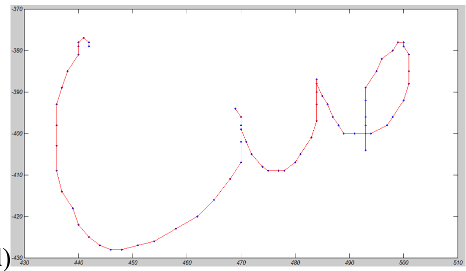
4.1 Refining TMU-OFS dataset

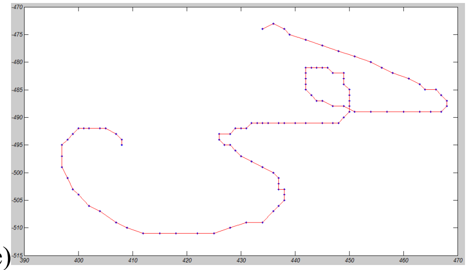
We tried to refine some problems in TMU-OFS dataset. In few classes, there were non-member samples that were eliminated. In some classes, there were spelling or structural mistakes, which were also eliminated. Some samples regarding the errors are shown in Figure 1. Some other styles of letter writing that are not included in Table 7 are demonstrated in Figure 2. Because of little usage in Farsi, we eliminated these samples, but it is possible to define the new styles for these in Table 7 (For example, Figure 1e, Figure 2a, b, c, d, e).

(a) 

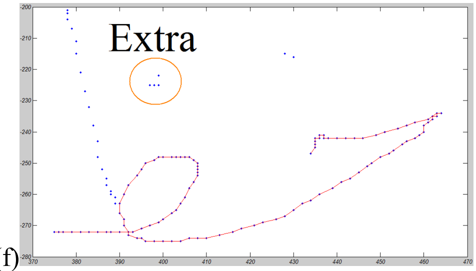
(b) 

(c) 

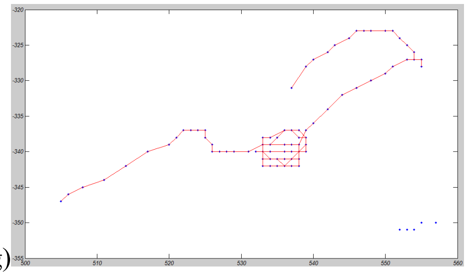
(d) 



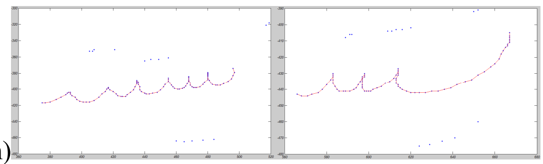
(e)



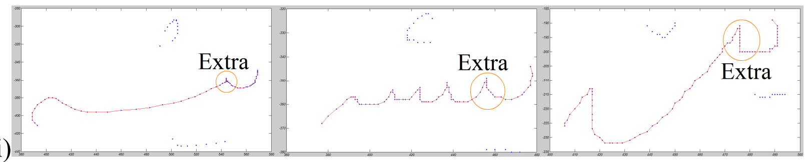
(f)



(g)

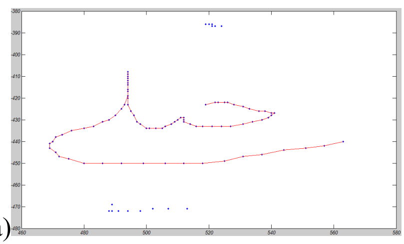


(h)



(i)

Fig. 1. Some wrongly spelled or incorrect structural samples; (a) Pas, "پس"-one extra tooth, (b) Sheka, "شکا"-wrongly written Shokola, "شکلا", (c) Tashkil, "تشکیل"-wrong style of "Kaaf"-style 2, where no "Alef, Laam, Kaaf" ("ا ، ل ، ک") comes after, (d) Saa, "صا"- one extra tooth, (e) Hen, "هن"- dot of letter "Noon" connected to its main stroke, (f) Khat, "خط"-one extra point, (g) Jome, "جمه"- circle of letter "Mim" was rotated several times, (h) Nistand, "نیستند"- one tooth missed (in both), (i) Yeshe, "یشه"- one extra tooth after "ی"



(a)

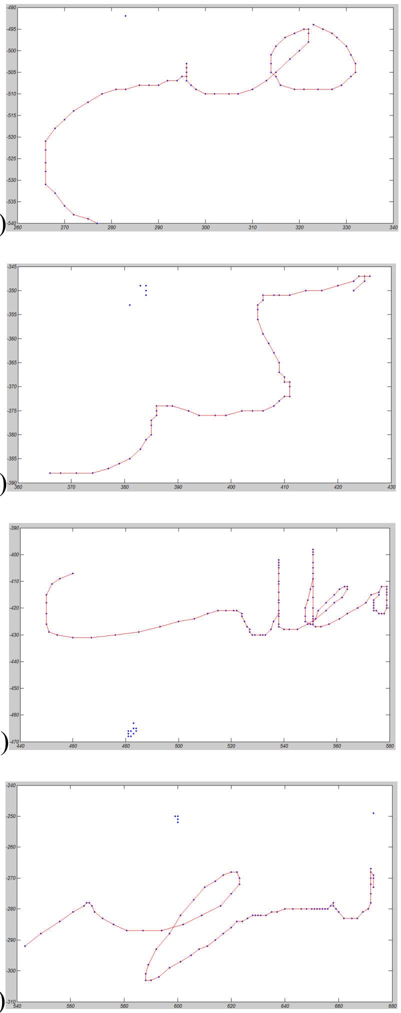


Fig. 2. Some samples written in a way not included in Table 7; (a) Kheili, "خیلی"- last "Ye" had return stroke, (b) Hen, "هن"- last "Noon" had return stroke, (c) Kaz, "کز"-"Kaaf" written in one stroke, (d) Matlab, "مطلب"-style 3 of "Taa" in the middle "Taa", (e) Noskhe, "نسخه"-middle "Khe", wrong descender

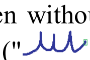
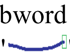
There are mixtures of subwords writing styles in the most classes, since the samples are written without any constraints. For example, subword "پسر" contains two models of writing, one with "sin" style 1 ("") and the other with "sin" style 2 (""). In some other subwords containing more letters of Table 7, more different models of writing are observed. As in Table 8, five models of writing subword Seah, "سعه" are shown.

Table 8. Different styles of writing Seah,"سعه"

Writing model	"Sin" style*	"Ein" style**	"He" style***
1 ()	1	1	1
2 ()	2	1	1
3 ()	1	2	1

4 ()	1	2	2
5 ()	2	2	1

* First letter from right
** Middle letter
*** Last letter

5.0 RESULTS AND DISCUSSION

We extended TMU-OFS dataset considering different styles of subwords writing. After dataset modification, 1000 classes for 1000 subwords were changed to 1711 classes of 1000 subwords and a second version of TMU-OFS dataset was created.

We did the experiments mentioned in Section 3.2 with the same conditions except using the second version of TMU-OFS again. The group recognition accuracy increased to 80% for the test set of the first 42 groups (mentioned in Section 3.2).

In another experiment, we used embedded training method of HMM [31] and applied it to TMU-OFS dataset and its second version separately, with corresponding basic features and equal conditions. The grouping based on the main body described in Section 3 was employed. The model of a main body was made of concatenating its letter models in embedded training. Firstly, all letters were extracted and trained by HMM. Then, the main body models were made by concatenating their letter models. The first 70% of the samples were applied for the training set and the other samples were for the test set. As shown in Table 9, the group recognition rate for the test set of TMU-OFS dataset was 42%, while this rate increased to 60% by considering 1711 classes of subwords. Some errors occurred because there were physical (natural) similarity between test classes and wrongly recognized classes (e.g. بیت, بت, شب). Exploiting the delayed strokes features can reduce these errors. The delayed strokes effect was investigated for subword modeling and lexicon reduction in [32].

Table 9. Group recognition rate

Dataset	Test set accuracy (%)	Training set accuracy (%)
TMU-OFS-I [24]	42	50
TMU-OFS-II	60	65

We investigated the varieties of letter writing styles in TMU-OFS dataset based on the main body. In Table 10 and Figure 3 the frequencies of letter writing styles mentioned in Table 7 were calculated. The values outside and within brackets are for the first and the second characters, respectively, in Table 10.

Table 10. Comparison of the handwriting styles (X= undefined)

Row	Letter	Style	Frequency at beginning place	Frequency at middle place	Frequency at ending place
1	Sin(Shin)	1	410 (303)	329 (331)	87 (109)
2	Sin(Shin)	2	203 (121)	129 (75)	13 (20)
3	Taa(Zaa)	1	61 (30)	72 (39)	48 (16)
4	Taa(Zaa)	2	28 (9)	45 (10)	0 (3)

5	Taa(Zaa)	3	12 (6)	X	X
6	Ein(Ghein)	1	397 (68)	127 (8)	38 (0)
7	Ein(Ghein)	2	X	209 (26)	62 (0)
8	Kaaf(Gaaf)	1	447 (155)	330 (204)	75 (30)
8-1	Kaaf(Gaaf) [Before Alef, Laam, Kaaf]	1	20 (9)	62 (49)	X
9	Kaaf(Gaaf)	2	24 (13)	60 (30)	X
10	Mim	1	724	541	105
11	Mim	2	362	133	144
12	He	1	380	151	677
13	He	2	X	302	195

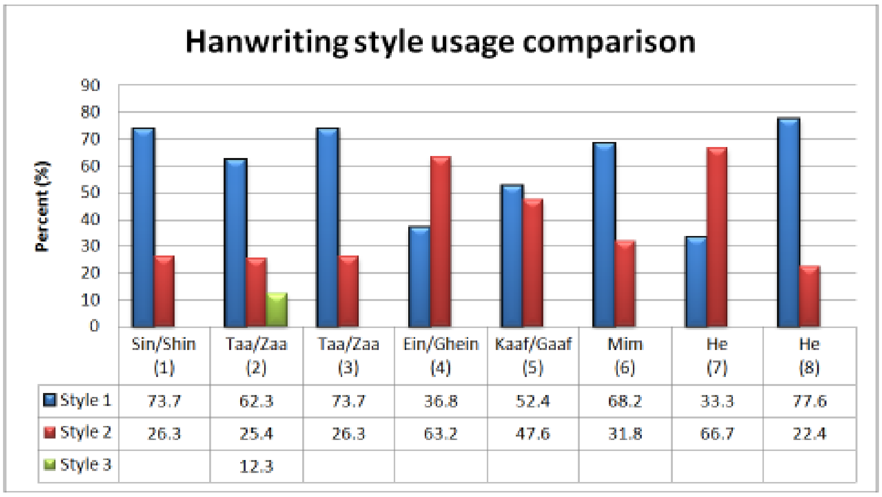


Fig. 3. Comparison of handwriting styles; (1) Total place, (2) At the beginning place, (3) Total place (1st and 2nd styles) (4) At the middle and the ending place, (5) Before "Alef, Laam, Kaaf" and at the first and the middle place, (6) Total place, (7) At the middle place, (8) At the ending place

In TMU-OFS dataset, version II, every class has a unique descriptor. The descriptors only include the codes of letters sequence of subwords. Most Farsi letters have four codes regarding their place in the subword. There are 119 unique codes for 32 letters. Some letters have the same shape for their isolated and beginning forms (or middle and ending forms). Therefore, we allocated one code to them. For example, isolated "Re" and beginning "Re" have the same code, 40. We added new codes for new styles of letter writing mentioned in Table 7. The total 167 codes are shown in Appendix A.

5.1 Analysis of writing variety

We analyzed all subwords of TMU-OFS dataset, version II, based on their main bodies. In rows 1 and 2 of Table 10, the frequencies of two writing styles of "sin" and "shin" ("س" and "ش") are given. Style 1 was utilized by 73.7% of the writers and style 2 by 26.3% of them.
Third writing style of letters "Taa" and "Zaa" ("ط" and "ظ") only occurs at the beginning place of subwords. Therefore, we considered beginning place frequency for a comparison. In beginning place, 62.3% of the writers utilized style 1, 25.4% of them used style 2 and 12.3% utilized style 3. In the other comparison between rows 3 and 4, 73.7% of the writers utilized style 1 and 26.3% used style 2.
Letters "Ein" and "Ghein" ("ع" and "غ") have one style at the beginning of subwords. In rows 6 and 7, we compared middle and ending forms of these letters. The second writing style () was applied by 63.2% of the writers. The tendency of fast writing is probably the reason of this result.
About rows 8 and 9, the second style of "Kaaf" and "Gaff" ("ک " and "گ")is normally written before letters "Alef", "Laam" and "Kaaf" ("ا" ,"ل" and "ک"), and this occurs at the beginning or middle of subwords. Therefore, we separated only letters "Kaaf" and "Gaaf" ("ک" and "گ") that they are placed before "Alef", "Laam" and "Kaaf" at the beginning and the middle of subwords to compare fairly in row 8-1. The application of the second style was 47.6%.
We extracted that 68.2% of the writers preferred to utilize anti clockwise movement for writing circle of letter "Mim" ("م").
Because of varieties in writing styles of letter "He" ("ه") regarding its position in the subword, we compared the percentage of usage at the middle and ending places, separately.

6.0 CONCLUSIONS

In the first part of this paper, two kinds of subword grouping based on their delayed strokes and main body were studied, and preliminary results of recognition were reported.
In the second part, varieties of Farsi handwriting styles were discussed. For handwriting recognition, especially online, study of varieties of styles is essential. After introducing some famous styles of letter writing in subwords, TMU-OFS dataset, version II was created. The preliminarily results showed a significant accuracy rate improvement in subword (word) recognition by grouping them based on their writing styles. Researchers can have access to the new version of the dataset by emailing their request to the authors.

REFERENCES

[1] M. Dehghan, K. Faez, M. Ahmadi, and M. Shridhar, "Handwritten Farsi (Arabic) Word Recognition: A Holistic Approach Using Discrete Hmm," Pattern Recognition. vol. 34. 2001, pp. 1057-1065.

[2] A. Amin, *Machine Recognition of Handwritten Arabic Words by the Irac Ii System*, in *Proceedings of the 6th International conference on pattern recognition*. 1982: Munich, Germany. p. 34-36.

[3] H. Khosravi and E. Kabir, "A Blackboard Approach Towards Integrated Farsi Ocr System," International Journal on Document Analysis and Recognition. vol. 12, no. 1. 2009, pp. 21-32.

[4] L.M. Lorigo and V. Govindaraju, "Offline Arabic Handwriting Recognition: A Survey," IEEE Transactions on Pattern Analysis and Machine Intelligence. vol. 28, no. 5. 2006, pp. 712-724.

[5] G.A. Montazer, H.Q. Saremi, and V. Khatibi, "A Neuro-Fuzzy Inference Engine for Farsi Numeral Characters Recognition," Expert Systems with Applications. vol. 37, no. 9. 2010, pp. 6327-6337.

[6] Z.A. Aghbari and S. Brook, "Hah Manuscripts: A Holistic Paradigm for Classifying and Retrieving Historical Arabic Handwritten Documents," Expert Systems with Applications. vol. 36, no. 8. 2009, pp. 10942-10951.

[7] A.H. Hassin, X.-L. Tang, J.-F. Liu, and W. Zhao, "Printed Arabic Character Recognition Using Hmm," Journal of Computer Science and Technology. vol. 19, no. 4. 2004, pp. 538-543.

[8] Z. Razak, K. Zulkiflee, N.M. Noor, R. Salleh, and M. Yaacob, "Off-Line Handwritten Jawi Character Segmentation Using Histogram Normalization and Sliding Window Approach for Hardware Implementation," Malaysian Journal of Computer Science. vol. 22, no. 1. 2009, pp. 34-43.

[9] C.L. Liu, S. Jaeger, and M. Nakagawa, "Online Recognition of Chinese Characters: The State-of-the-Art," IEEE Transactions on Pattern Analysis and Machine Intelligence. vol. 26, no. 2. 2004, pp. 198-213.

[10] S. Jaeger, C.L. Liu, and M. Nakagawa, "The State of the Art in Japanese Online Handwriting Recognition Compared to Techniques in Western Handwriting Recognition," International Journal on Document Analysis and Recognition. vol. 6, no. 2. 2003, pp. 75-88.

[11] P. Artieres, *Stroke Level Hmm for on-Line Handwriting Recognition*, in *Proceedings of the 8th international workshop on frontiers in handwriting recognition*. 2002.

[12] L. Jin, Y. Gao, G. Liu, Y. Li, and K. Ding, "Scut-Couch2009—a Comprehensive Online Unconstrained Chinese Handwriting Database and Benchmark Evaluation," International Journal on Document Analysis and Recognition. vol. 14, no. 1. 2011, pp. 53-64.

[13] V. Ghods and E. Kabir, *Feature Extraction for Online Farsi Characters*, in *Proceedings of the 12th international conference on frontiers in handwriting recognition (ICFHR)*. 2010: Kolkata, India. p. 477-482.

[14] R. Plamondon and N. Srihari, "On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey," IEEE Transactions on Pattern Analysis and Machine Intelligence. vol. 22, no. 1. 2000, pp. 63-84.

[15] M.S. Baghshah, S.B. Shouraki, and S. Kasayi, *A Novel Fuzzy Approach to Recognition of Online Persian Handwriting*, in *Proceedings of the 5th international conference on intelligent systems design and applications (ISDA'05)*. 2005: Poland.

[16] J. Sternby, J. Morwingb, J. Anderssonb, and C. Fribergb, "On-Line Arabic Handwriting Recognition with Templates," Pattern Recognition. vol. 42, no. 12. 2009, pp. 3278-3286.

[17] H.E. Abed, M. Kherallah, V. Märgner, and A.M. Alimi, "On-Line Arabic Handwriting Recognition Competition, Adab Database and Participating Systems," International Journal on Document Analysis and Recognition. vol. 14, no. 1. 2011, pp. 15-23.

[18] R. Halavati and S.B. Shouraki, "Recognition of Persian Online Handwriting Using Elastic Fuzzy Pattern Recognition," International Journal of Pattern Recognition and Artificial Intelligence. vol. 21, no. 12. 2007, pp. 491-513.

[19] M. Kavianifar and A. Amin, *Preprocessing and Structural Feature Extraction for a Multi-Fonts Arabic/Persian Ocr*, in *Proceedings of the 5th international conference on document analysis and recognition*. 1999. p. 213–216.

[20] V. Ghods, E. Kabir, and F. Razzazi, "Decision Fusion of Horizontal and Vertical Trajectories for Recognition of Online Farsi Subwords," Engineering Applications of Artificial Intelligence. vol. 26, no. 1. 2013, pp. 544-550.

[21] H. Khosravi and E. Kabir, "Introducing a Very Large Dataset of Handwritten Farsi Digits and a Study on Their Varieties," Pattern Recognition Letters. vol. 28, no. 10. 2007, pp. 1133-1141.

[22] S. Mozaffari, K. Faez, F. Faradji, M. Ziaratban, and S.M. Golzan, *A Comprehensive Isolated Farsi/Arabic Character Database for Handwritten Ocr Research*, in *Proceedings of the 10th international workshop on frontiers in handwriting recognition*. 2006. p. 385-389.

[23] F. Solimanpour, J. Sadri, and C. Suen, *Standard Databases for Recognition of Handwritten Digits, Numerical String, Legal Amounts, Letters and Dates in Farsi Language*, in *Proceedings of the 10th international workshop on frontiers in handwriting recognition*. 2006. p. 3-7.

[24] S.M. Razavi and E. Kabir, *A Dataset for Online Farsi Handwriting*, in *6th national conference on intelligent systems (in Farsi)*. 2004: Kerman, Iran.

[25] M.E. Dehkordi, N. Sherkat, and T. Allen, "Handwriting Style Classification," International Journal on Document Analysis and Recognition. vol. 6. 2003, pp. 55–74.

[26] L. Schomaker, M. Bulacu, and K. Franke, *Automatic Writer Identification Using Fragmented Connected-Component Contours*, in *Proceedings of the 9th international workshop on frontiers in handwriting recognition*. 2004: Japan. p. 185-190.

[27] S.N. Srihari, S.H. Cha, H. Arora, and S. Lee, "Individuality of Handwriting," Journal of Forensic Sciences. vol. 47, no. 4. 2002, pp. 1–17.

[28] S.M. Razavi and E. Kabir, "A Simple Method for Online Farsi Subword Recognition," Journal of electrical and computer engineering (in Farsi). vol. 2. 2006, pp. 63-72.

[29] T. Plötz and G.A. Fink, "Markov Models for Offline Handwriting Recognition: A Survey," International Journal on Document Analysis and Recognition. vol. 12. 2009, pp. 269–298.

[30] H. Sajedi, M. Jamzad, H. Sameti, and B. Babaali, "A Grouping-Based Method for on-Line Farsi Discrete Character Recognition Using Hidden Markov Model," *12th international conference of computer society of Iran*, 2007, p. Pages

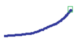
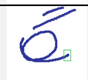
[31] L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," Proceedings of the IEEE. vol. 77. 1989, pp. 257–285.

[32] V. Ghods, E. Kabir, and F. Razzazi, "Effect of Delayed Strokes on the Recognition of Online Farsi Handwriting," Pattern Recognition Letters. vol. 34. 2013, pp. 486-491.

Appendix

Appendix A. Allocated codes to letters ('-' denotes a reserved code)

Letter	Isolated		Beginning place		Middle place		Ending place	
Alef	آ	1	آ	1	آ	1	آ	1
Alef	ا	2	ا	2	ﺎ	3	ﺎ	3
Be	ب	4	ﺑ	5	ﺑ	6	ب	7
Pe	پ	8	ﭘ	9	ﭘ	10	پ	11
Te	ت	12	ﺗ	13	ﺗ	14	ت	15
Se	ث	16	ﺛ	17	ﺛ	18	ث	19
Jim	ج	20	ﺟ	21	ﺟ	22	ج	23
Che	چ	24	ﭼ	25	ﭼ	26	چ	27
He	ح	28	ﺣ	29	ﺣ	30	ح	31
Khe	خ	32	ﺧ	33	ﺧ	34	خ	35
Daal	د	36	د	36	د	37	د	37
Zaal	ذ	38	ذ	38	ذ	39	ذ	39
Re	ر	40	ر	40	ر	41	ر	41
Ze	ز	42	ز	42	ز	43	ز	43
Zhe	ژ	44	ژ	44	ژ	45	ژ	45
Sin	س	46	ﺳ	47	ﺳ	48	س	49
Shin	ش	50	ﺷ	51	ﺷ	52	ش	53
Saad	ص	54	ﺻ	55	ﺻ	56	ص	57
Zaad	ض	58	ﺿ	59	ﺿ	60	ض	61
Taa	ط	62	ط	63	ط	64	ط	65
Zaa	ظ	66	ظ	67	ظ	68	ظ	69
Ein	ع	70	ﻋ	71	ﻌ	72	ع	73
Ghein	غ	74	ﻏ	75	ﻐ	76	غ	77
Fe	ف	78	ﻓ	79	ﻓ	80	ف	81
Ghaaf	ق	82	ﻗ	83	ﻗ	84	ق	85
Kaaf	ک	86	ﻛ	87	ﻛ	88	ك	89

Gaaf	ﮒ	90	ﮔ	91	ﮕ	92	ﮓ	93
Laam	ل	94	ﻟ	95	ﻠ	96	ﻞ	97
Mim	م	98	ﻣ	99	ﻤ	100	ﻢ	101
Noon	ن	102	ﻧ	103	ﻨ	104	ﻦ	105
Vaav	و	106	و	106	ﻮ	107	ﻮ	107
He	ه	108	ﻫ	109		110		111
Ye	ی	112	ﻳ	113	ﻴ	114	ي	115
Hamze	ء	116	ﺋ	117	ﺌ	118	أ	119
Sin – 2		120		121		122		123
Shin - 2		124		125		126		127
Taa – 2	ط	128	ط	129	-	130	-	131
Zaa – 2	ظ	132	ظ	133	-	134	-	135
Taa – 3	ط	136	ط	137	ط	138	ط	139
Zaa - 3	ظ	140	ظ	141	ظ	142	ظ	143
Ein – 2	-	144	-	145	ﻌ	146	ع	147
Ghein- 2	-	148	-	149	ﻐ	150	غ	151
Kaaf - 2	-	152		153		154	-	155
Gaaf - 2	-	156		157		158	-	159
Mim - 2	م	160	ﻣ	161	ﻤ	162	م	163
He - 2	-	164	-	165	ﻬ	166	ﻪ	167