

Off-line Sindhi Handwritten Character Identification

Arsha Kumari

Department of Electronic Engineering, Mehran University of Engineering and Technology Jamshoro, Sindh Pakistan
E-mail: arsharathi56@gmail.com

Din Muhammad Sangrasi*, Sania Bhatti*, Bhawani Shankar Chowdhry and Sapna Kumari***

*Department of Software Engineering Mehran UET Jamshoro, Sindh Pakistan

**Meritorious Professor, Faculty of Electrical, Electronics and Computer Engineering, MUET Jamshoro, Sindh
E-mail: din.muhammad@faculty.muett.edu.pk, sania.bhatti@faculty.muett.edu.pk, c.bhawani@ieee.org, rathisapna65@gmail.com

Received: 26 March 2019; Accepted: 19 May 2019; Published: 08 June 2019

Abstract—Handwritten Identification is an ability of the computer to receive and translate the intelligible handwritten text into machine-editable text. It is classified into two types based on the way input is given namely: off-line and online. In Off-line handwritten recognition, the input is given in the form of the image while in online input is entered on a touch screen device. The research on off-line and online handwritten Sindhi character identification is on its very initial stage in comparison to other languages. Sindhi is one of the subcontinent's oldest languages with extensive literature and rich culture. Therefore, this paper aims to identify off-line Sindhi handwritten characters. In the proposed work, major steps involve in characters identification are training and testing of the system. Training is performed using a feed-forward neural network based on the efficient accelerative technique, the Back Propagation (BP) learning algorithm with momentum term and adaptive learning rate. The dataset of 304 Sindhi handwritten characters is collected from 16 different Sindhi writers, each with 19 characters. The novelty of proposed work is the comparison of the recognition rate for the single character, two characters and three characters at a time. Results showed that the recognition rate achieved for a single character is more than the recognition rate of multiple characters at a time.

Index Terms—Off-line Handwritten, Neural Network training, Back Propagation (BP) algorithm, Sindhi Character identification.

I. INTRODUCTION

Handwritten character recognition (HCR) is advancing the communication between human and computer; it takes the world toward automation [1]. Off-line handwritten recognition is a somehow easy and fast way of inputting data to the computer. As plenty of Sindhi literature is available in Sindh literature departments in hard form, has taken too much space and will take too

much time to access any information. Therefore it is necessary to preserve that information in the digitized form so that globally everyone can access easily. Hence Sindhi HCR is a very initial step to preserve the Sindhi literature on the web to use it at the worldwide resource. Handwritten recognition is a very challenging task in computer vision and pattern recognition since every writer has a different writing style, different shape of characters and font, image quality [2]. As Sindhi is cursive language in which characters are connected to form words, hence it is a more difficult job when it comes to recognize the off-line Sindhi handwritten characters. Another problem in recognition of Sindhi characters is similarity in 1) basic shape of characters, 2) position of dots and 3) the number of dots of different character [3]. Though much work has been done on other languages such as English [4], Chinese [5], Arabic [6] and other languages but very less work has been done on Sindhi HCR at the best of the knowledge, so a lot of work is required to be done in this direction. Since Sindhi is the regional and provincial language of Pakistan spoken by 60 million people in Sindh and different areas of the world [7].

Basically, HCR is classified into two types first one is off-line and another is online. Both these types vary from one another by the way the input is given to the system. In the off-line handwritten recognition, the input is given in the form of a paper document, image etc. that will be static in nature. While in online the input is given on the touch screen device such as Tablet etc., that input will be dynamic in nature.

The proposed work main purpose is to identify the off-line handwritten Sindhi characters using the BP algorithm with adaptive learning and momentum which reduce the training time of the network. An additional contribution of this research is to perform the comparative analysis in the recognition rate for a single character and multiple characters at a time. This system is based on the graphical user interface (GUI) which is developed using MATLAB 2017a programming environment by utilizing its

computer vision and neural network tools. It is trained through the dataset of 304 characters, collected from 16 different Sindhi writers and tested by same and different writers test images. The paper is organized in the following sequence: In section II, a literature review of presented work will be enlightened. Then, in section III, a block diagram of the proposed methodology and detail of its every step with their result will be given. After that, the system performance will be evaluated on the basis of the recognition rate by testing the system using test images in section IV. In the last section, V conclusion and future work will be presented.

II. RELATED WORK

From recent work study, it has been observed that limited work has been done on Sindhi handwritten recognition which is discussed in this section.

In paper [8] Nasreen Nizamani et al. compares various thinning based pre-processing algorithms on Sindhi printed dataset which is the initial step for character recognition. Those algorithms included Z.-Suen Thinning, Thinning based morphological operation (TMBO), Skeletonization based morphological operation (SBMO) and Huang Wan Liu (H.-W.Liu) algorithm. Thinning of 4707 Sindhi characters is analyzed based on text connectivity and processing time. The results have shown the H. W. Liu thinning algorithm is suitable algorithm than other three for thinning of Sindhi characters. This algorithm preserves the connectivity of characters and it also requires moderate processing time.

Authors in the study [3] proposed work on online Sindhi Handwritten word recognition using Self-organizing map algorithm which is the type of artificial neural network. An application based on an Android platform has been developed to recognize the words written on the touch screen of a device. The dataset consists of 1200 words, collected from 60 Sindhi language native writers. This work is tested by 25 Sindhi writers and characters are recognized with an accuracy rate of 83% and the processing time of 20-30 milliseconds.

Work in the study [9] has been proposed work on two objectives: 1) complexities occurred in recognizing Sindhi characters and 2) to present experimental work of isolated Sindhi handwritten character recognition using BP neural network algorithm. For training, the dataset of total five hundred twenty input patterns is collected from five native and five non-native writers. Given the system is tested from two native writers and two non-native writers, taking 52 characters from each writer. Experimental results showed that the recognition accuracy achieved for a native writer is 91% and for non-native writers, it is 71%. Average recognition accuracy obtained is 87.75%.

A technique based on morphology is presented in paper [10]. The authors reviewed importance of Sindhi morphemes such as nature, function and structure. Also, comparative analysis has been presented between Sindhi and English language morphology to understand the Sindhi language morphology. Moreover, the paper

presented the detail of the word formation with respect to the words suffixes, prefixes and its type.

Authors in a study [11] worked on multi-digit handwritten Sindhi Numerals using SOM (Self-Organizing map). This system is able to recognize any style and different font size numerals. For the training, only one sample of each pair of multi digits numerals is required. The dataset of 4000 multi-digits numerals consisting of two digits from 10-50 is collected and for testing the system, 50 multi digits numerals are taken from 50 users. The recognition accuracy achieved from various experiments is 86.89%.

Dil Nawaz Hakro et al. in [12] presented the work on Sindhi Named entity recognition (SNER) which means to search and classify meaningful information such as the name of cities, persons, locations etc., from a given text. The system can extract and classify Sindhi words includes city names, person names and other entities. The achieved accuracy in extracting and classifying the Sindhi words is 79%.

III. PROPOSED METHODOLOGY

The proposed methodology is demonstrated on the GUI developed in MATLAB 2017a programming environment using its computer vision and neural network toolbox on a PC with window 10, a 64-bit operating system with 4GB RAM. The detailed steps deployed in the proposed work with their results are discussed in this section. The block diagram of the training and testing phases is shown in Fig.1.

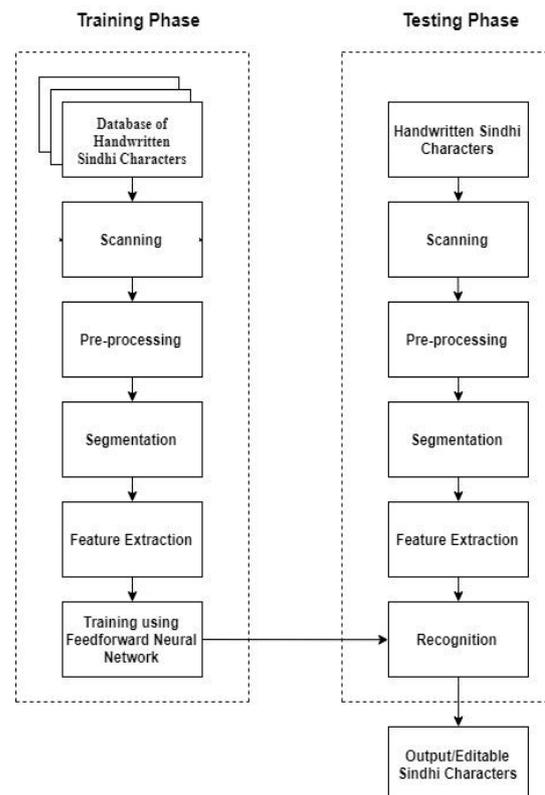
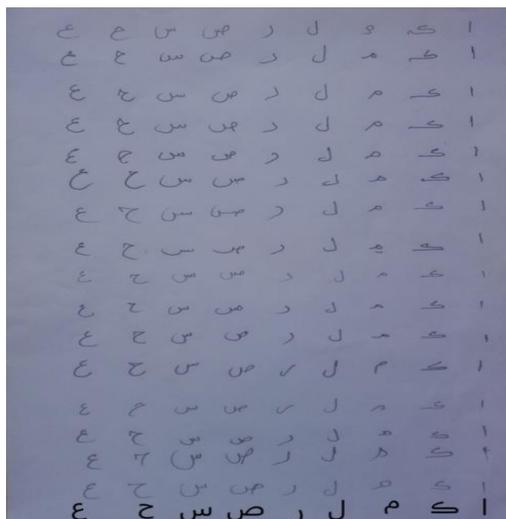


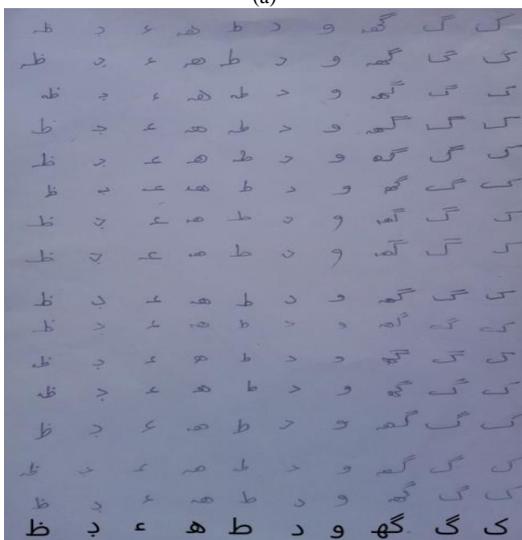
Fig.1. Training and testing phases deployed for off-line handwritten Sindhi characters identification

A. Dataset Collection

Suitable dataset is necessary for the training of the neural network. Therefore, a dataset of 304 characters is collected from 16 Sindhi writers on plain white paper by using any colour pen. Each writer allowed for writing 19 Sindhi characters. The dataset of characters is shown in Fig.2 (a). & Fig.2 (b).The proposed dataset presented many challenges because of the different position of dots, writing style, number of dots and thickness [13]. Before using the dataset for training, the preprocessing algorithm is applied to it. As there is saying about preprocessing techniques “Rubbish in, Rubbish out”. Results obtained from the preprocessed dataset are special attributes of characters mapped for training. The extracted attributes contain the numerical values that are stored in an array.



(a)



(b)

Fig.2. Dataset of 304 characters from 16 Sindhi writers (a) Dataset of 9 Characters (b) Dataset of remaining 10 characters.

B. Scanning

Handwritten images for training and testing are scanned through a mobile camera with high resolution.

During scanning the light illuminating on images must be uniform otherwise noise will occur in the scanned image. The noise will affect the accuracy of the characters recognition rate. Scanned image result is shown in Fig.3.

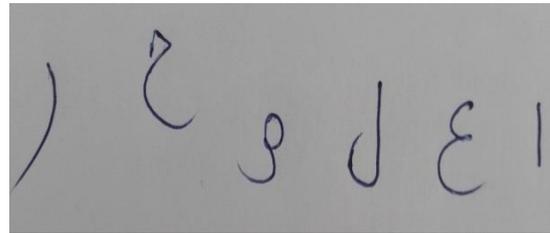


Fig.3. Scanned Image

C. Pre-processing

After scanning, all images are pre-processed which is consisted of a sequence of operations. The main objective of preprocessing is to enhance the images by modifying its attributes to make the images suitable for the recognition. Applications based on the character recognition system uses grey or binary images since high computation is required for processing colour images. It is difficult to extract characters from scanned images without applying suitable preprocessing techniques since images may contain noise or watermark; desired results for recognition purpose is a binary image with only characters [14]. Pre-processing includes several techniques which are binarization, edge detection and morphological operations (dilation and filling). All the training and testing images are passed through these steps.

1. Binarization

This step will convert the grey level image into bi-level form in which the background pixels are converted into white and foreground into black pixels. This process is called binarization and method used for it is termed as thresholding. In proposed work, Otsu’s global thresholding method [15] is employed, that will apply one threshold on the entire input image estimated from the background level intensity of image histogram. The binarization result is shown in Fig.4.



Fig.4. Result of Binarization operation

2. Edge Detection

After thresholding operation, edge detection is applied to the binary images to find the boundaries of images using a Sobel edge detection algorithm [16] and the

resultant image is shown in Fig.5.



Fig.5. Edge detection applied to Sindhi characters

3. Morphological Operations

Morphological operations process the images based on shapes. The basic operations of morphology are erosion and dilation. In which image's pixels are adjusted according to the value of a neighbour's pixel value to construct the structuring element.

3.1. Dilation

Dilation is a basic type of morphological operation, it enlarges the white region in the image, or it increases the pixels in foreground characters. It will also connect the characters which are broken. The resultant image of Sindhi characters after applying dilation is shown in Fig.6.



Fig.6. Sindhi characters after applying dilation operation

3.2. Filling

This operation fills the holes in characters which are formed in the dilated image. As in 'م' and 'ح', there are holes in the middle which are filled after this operation as shown in Fig.7.



Fig.7. Sindhi characters after applying filling operation

D. Segmentation

Segmentation goal is to extract regions which are internally similar in texture. Images are segmented into regions to allow analysis and processing of each segment independently. The pre-processed image is fragmented into individual characters by assigning a number to every character using the labelling process [17]. The result of the segmentation operation applied to the pre-processed image is shown in Fig.8.

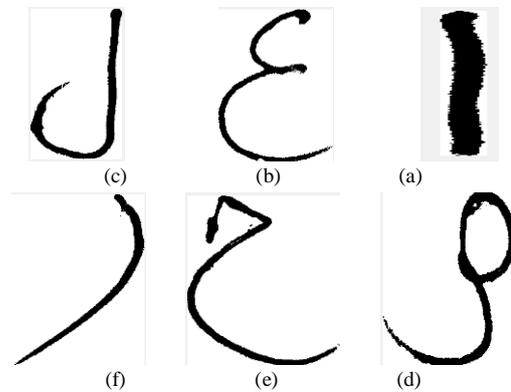


Fig.8. The result of Segmentation operation applied to the pre-processed image

E. Feature Extraction

Feature extraction plays an important role in recognizing the characters effectively. This will extract the high-level information from the segmented image to facilitate it for recognition [18]. The segmentation step provided the isolated characters, therefore, characters normalized by fixed size that is decided on the basis of the experimental or empirical technique used for feature extraction. Hence, all the pixels in images containing the important features of characters are extracted by resizing the cropped binary image into 100*290 pixels for facilitating the training and testing. Feature extraction of segmented Sindhi characters is shown in Fig.9.

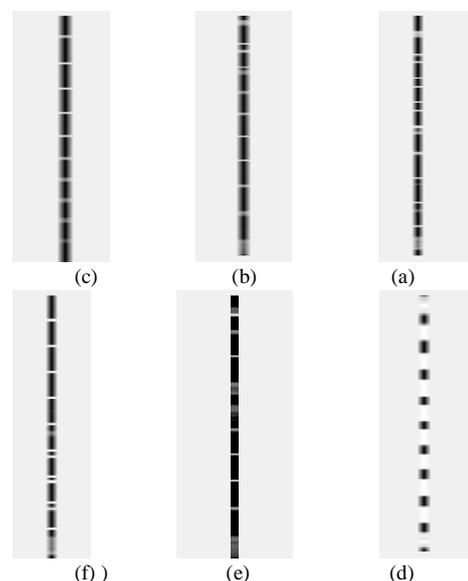


Fig.9. Feature Extraction result of each segmented Sindhi character

F. Training

Training is one of the important steps in identifying off-line Sindhi handwritten characters. In stated work, the type of model used for training and classification is BP learning with momentum and adaptive learning based on Feed Forward neural network [19]. Here the adaptive learning rate and momentum factor are employed to accelerate the training by adjusting momentum at each iteration, unlike the conventional Bp algorithm in which the longer training time and slower convergence are the main disadvantages. The BP is an optimization algorithm based on gradient descent used to reduce the system error by repetitively adjusting the weights of neurons. The process of weight adjustment continues until the minimum mean square error (MSE) is obtained. For training the network, 290 input neuron and 19 output/target neurons are used which is shown in the neural network panel in Fig.10 (b). The number of time the input neurons are used and their weights are adjusted during training is called an Epoch. In the given system 858 times neurons weight is adjusted as shown in progress panel of Fig.10 (b). Here adaptive learning and momentum are increasing the learning rate to the extent that trains the network without significantly increasing error. The time required for training the network is five seconds as shown in progress panel of Fig. 10(b). The neural network training performance shows the minimum error obtained, instated network obtained error is (0.00886). The recognition accuracy of characters will be maximum when the minimum error will be achieved. Here gradient (9.94e-06) shows the variation occurs in the error.

G. Recognition

In recognition step trained neural network is tested using the test images taken from the same and different writers. The MATLAB GUI for testing the system is shown in Fig.11.

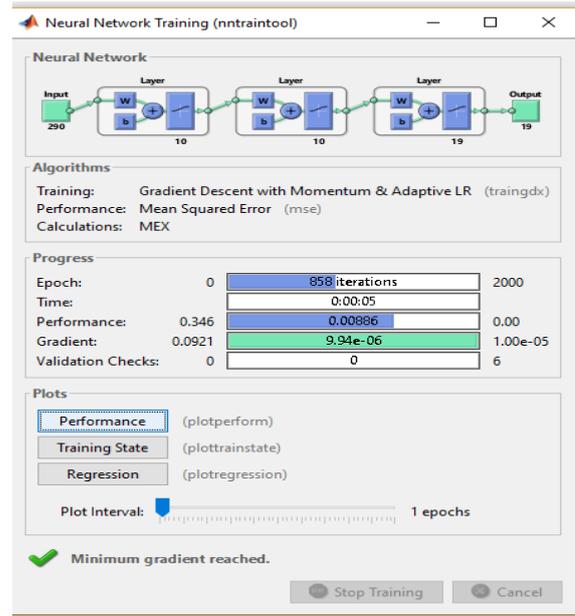


Fig.10(b). Neural Network training

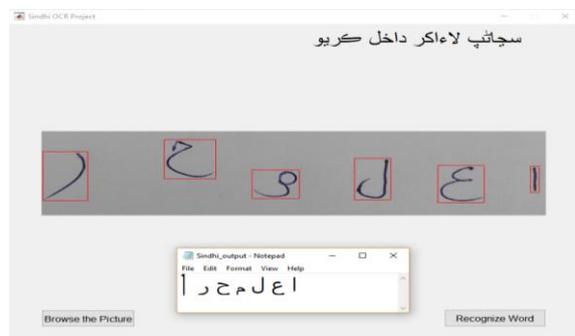


Fig.11. MATLAB GUI for testing the proposed system

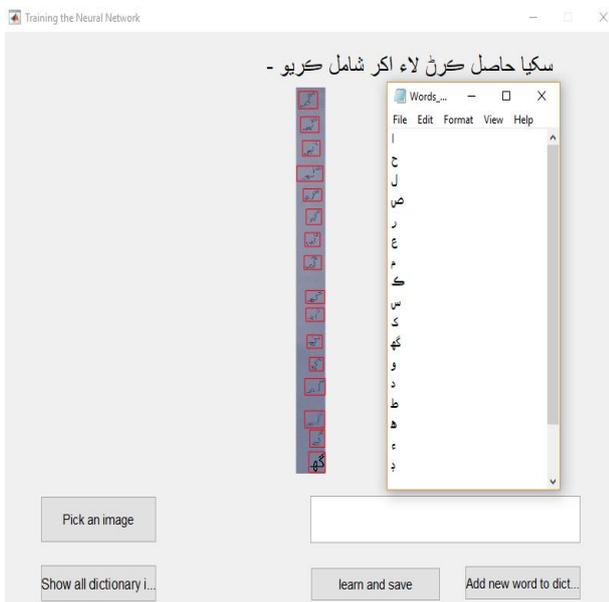


Fig.10 (a). MATLAB GUI for the training of 19 Sindhi characters

IV. RESULTS & DISCUSSION

Results are obtained after applying all the steps explained in section III on the training and test images in software MATLAB 2017a. The trained neural network is tested using test images, which are taken from the same writers and different writers. Here the same writer means from whom dataset characters are collected while the different writers mean writers from which dataset is not collected. As there is a total of 19 characters dataset is collected for the network training. So for testing, per character five test images are taken from same and different writers. In the proposed system, the recognition rate is calculated by taking the ratio of recognized characters to the total test characters taken from the writers.

$$\text{Recognition rate} = \frac{\text{Recognized characters}}{\text{Total characters taken for testing}} \quad (1)$$

When the system is tested by the same writer the recognition rate achieved for the single character is 85.2% as shown in Table 1. As every writer has their own written style so the recognition rate differs for each writer.

Table 1. System Tested by the same writers for a single character

| S. no | Single Sindhi character | Total Test images were taken for Testing | Correctly Recognized Characters | Recognition Rate (%) |
|------------------------------|-------------------------|--|---------------------------------|----------------------|
| 1 | ا | 5 | 5 | 100 |
| 2 | ح | 5 | 4 | 80 |
| 3 | ل | 5 | 5 | 100 |
| 4 | م | 5 | 4 | 80 |
| 5 | ع | 5 | 4 | 80 |
| 6 | ھ | 5 | 4 | 80 |
| 7 | ط | 5 | 4 | 80 |
| 8 | ک | 5 | 4 | 80 |
| 9 | س | 5 | 4 | 80 |
| 10 | ظ | 5 | 4 | 80 |
| 11 | ر | 5 | 5 | 100 |
| 12 | و | 5 | 5 | 100 |
| 13 | گ | 5 | 4 | 80 |
| 14 | ص | 5 | 4 | 80 |
| 15 | د | 5 | 5 | 100 |
| 16 | ڊ | 5 | 4 | 80 |
| 17 | ڪ | 5 | 4 | 80 |
| 18 | گھ | 5 | 4 | 80 |
| 19 | ء | 5 | 4 | 80 |
| Average Recognition rate (%) | | 85.2% | | |

The recognition rate obtained for different writers for single Sindhi character recognition is 81% by testing the system using

five test images for each character. The average recognition rate for the different writers is given in the following Table 2.

Table 2. System tested by different writers for single character

| S. no | Single Sindhi Characters | Total Test images were taken for Testing | Correctly Recognized Characters | Recognition Rate (%) |
|------------------------------|--------------------------|--|---------------------------------|----------------------|
| 1 | ا | 5 | 5 | 100 |
| 2 | ح | 5 | 4 | 80 |
| 3 | ل | 5 | 4 | 80 |
| 4 | م | 5 | 4 | 80 |
| 5 | ع | 5 | 4 | 80 |
| 6 | ھ | 5 | 4 | 80 |
| 7 | ط | 5 | 4 | 80 |
| 8 | ک | 5 | 4 | 80 |
| 9 | س | 5 | 4 | 80 |
| 10 | ظ | 5 | 4 | 80 |
| 11 | ر | 5 | 4 | 80 |
| 12 | و | 5 | 4 | 80 |
| 13 | گ | 5 | 4 | 80 |
| 14 | ص | 5 | 4 | 80 |
| 15 | د | 5 | 4 | 80 |
| 16 | ڊ | 5 | 4 | 80 |
| 17 | ڪ | 5 | 4 | 80 |
| 18 | گھ | 5 | 4 | 80 |
| 19 | ء | 5 | 4 | 80 |
| Average Recognition rate (%) | | 81.0% | | |

Also, the system is tested for two characters at a time taken from the same writers. Total sixteen different pairs of Sindhi characters are tested by taking five test images for each pair of Sindhi characters. The recognition

accuracy obtained for two characters identification is less than one character recognition at a time as shown in Table 3.

Table 3. System tested by the same writer for two Sindhi characters at a time

| S. No | Pair of two Sindhi characters | Total Test images were taken | Recognized characters from test images | Recognition Rate (%) |
|------------------------------|-------------------------------|------------------------------|--|----------------------|
| 1 | ا، ک | 5 | 4 | 80 |
| 2 | ا، ک | 5 | 3 | 60 |
| 3 | ل، ه | 5 | 2 | 40 |
| 4 | ط، و | 5 | 3 | 60 |
| 5 | ح، م | 5 | 2 | 40 |
| 6 | س، ا | 5 | 3 | 60 |
| 7 | ک، ع | 5 | 4 | 80 |
| 8 | ک، س | 5 | 3 | 60 |
| 9 | ص، م | 5 | 2 | 40 |
| 10 | د، ء | 5 | 3 | 60 |
| 11 | گ، ا | 5 | 3 | 60 |
| 12 | ک، ک | 5 | 3 | 60 |
| 13 | د، ع | 5 | 3 | 60 |
| 14 | ر، م | 5 | 3 | 60 |
| 15 | د، ص | 5 | 4 | 80 |
| 16 | ر، ع | 5 | 4 | 80 |
| Average recognition rate (%) | | | 61.25% | |

The proposed system is also tested for the combination of three characters at a time. The average recognition rate of 40% is achieved for the test images containing three characters as given in Table 4.

Table 4. System tested by the same writer for a combination of three characters

| S. No | Pair of three Sindhi characters | Total Test images were taken | Recognized characters from test images | Recognition Rate (%) |
|------------------------------|---------------------------------|------------------------------|--|----------------------|
| 1 | ا، ل، د | 5 | 3 | 60 |
| 2 | س، ا، ک | 5 | 2 | 40 |
| 3 | و، ا، س | 5 | 2 | 60 |
| 4 | ح، ر، د | 5 | 2 | 40 |
| 5 | ا، د، و | 5 | 3 | 60 |
| 6 | م، ک، ا | 5 | 2 | 40 |
| 7 | ه، د، ک | 5 | 1 | 20 |
| 8 | ح، ر، ل | 5 | 1 | 20 |
| 9 | ل، ا، ر | 5 | 2 | 60 |
| 10 | و، ء، ط | 5 | 2 | 40 |
| 11 | گ، ه، ر | 5 | 1 | 20 |
| 12 | ک، ل، د | 5 | 2 | 40 |
| 13 | ع، و، ک | 5 | 1 | 20 |
| 14 | ح، ا، م | 5 | 2 | 40 |
| 15 | و، گ، ه | 5 | 1 | 40 |
| 16 | ر، س، ک | 5 | 1 | 40 |
| Average recognition rate (%) | | | 40% | |

Fig.12. shows a comparison of the same writer’s recognition rate for the single character, two characters and three characters at a time. It has been analyzed that system recognition accuracy in recognizing a single character at a time is highest while recognition rate decreased when the system is tested for the images contain two characters and three characters.

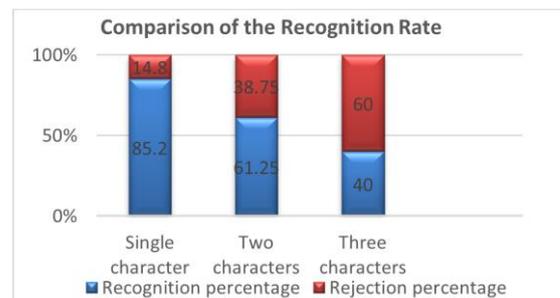


Fig.12. Comparison of recognition rate for single, two and three characters

It is concluded from experiments that at a time the system is able to recognize more than one character from test images as shown in Fig.13 (a) and Fig.13 (b). But recognition rate obtained in recognizing the single character is highest then the image contains more than one character, the recognizing accuracy decreased because there are some characters having a same basic shape like 'س' and 'ص' others 'ح' and 'ع'. Hence, these characters often recognized interchangeably. The simple shaped character such as 'ا', 'ل', 'و', 'ه' has the highest recognition rate. Meanwhile, characters like 'ک', 'م', and 'گ' has achieved less recognition rate because of complicated geometrical shape.



Fig.13 (a). Four Sindhi characters are recognized at a time from the given test image



Fig.13 (b). Two Sindhi characters recognized from a given test image

V. CONCLUSION & FUTURE WORK

In this paper off-line handwritten Sindhi character identification has been implemented on MATLAB 2017a by developing a GUI. The proposed system is first trained by collecting the dataset of 304 characters from 16 Sindhi writers; per writer 19 individual characters were written. The main contribution of the given research is the recognition rate comparison of the single character, two characters and three characters test images taken from the same writers. Based on experimental results average recognition rate achieved for the single character is 85.2%, for two characters the recognition rate 61.25% and for three characters 40% recognition rate is obtained. This decreasing trend in recognition rate is due to the complicated nature of the few characters and similar geometrical shape of some characters.

The system has some limitations such as interchangeably recognizing the characters having a similar basic shape which results in reduce of the

recognition rate. In future, work will be done to improve the recognition accuracy by increasing the characters dataset. Secondly, for improving the network training, other algorithms can be used such as proportion factor (PF) and Levenberg-Marquardt algorithm. Additionally, work can be extended for recognizing remaining Sindhi characters. Furthermore, our work opens the new window for researchers to work in the Sindhi language for words and sentences recognition.

ACKNOWLEDGEMENT

The authors are very grateful to Institute of Information and communication technologies, Mehran University of engineering and technology Jamshoro for making the research necessary for master's program, which allows the students to learn deeply about his/her field of interest. Also, we would like to express cordial thanks to every individual who supported, helped, and encouraged us during this work. Above all we are very thankful to Almighty God for all the blessings, without his guidance, this was not possible.

REFERENCES

- [1] J. Pradeep, E. Srinivasan, and S. Himavathi, "Diagonal Feature Extraction Based Handwritten Character System Using Neural Network," *Int. J. Comput. Appl.*, vol. 8, no. 9, pp. 17–22, 2010.
- [2] A. A. Sanjrani, J. Baber, M. Bakhtyar, W. Noor, and M. Khalid, "Handwritten Optical Character Recognition system for Sindhi numerals," *2016 Int. Conf. Comput. Electron. Electr. Eng. ICE Cube 2016 - Proc.*, pp. 262–267, 2016.
- [3] A. A. CHANDIO, M. LEGHARI, D. HAKRO, S. A. AWAN, and A. H. JALBANI, "A Novel Approach for Online Sindhi Handwritten Word Recognition using Neural Network," *Sindh Univ. Res. J. (Science Ser.)*, vol. 48, no. 1, pp. 213–216, 2016.
- [4] A. Yuan, G. Bai, P. Yang, Y. Guo, and X. Zhao, "Handwritten English Word Recognition based on Convolutional Neural Networks," pp. 207–212, 2012.
- [5] R. Dai, C. Liu, and B. Xiao, "Chinese character recognition: History, status and prospects," *Front. Comput. Sci. China*, vol. 1, no. 2, pp. 126–136, 2007.
- [6] A. El-Sawy and M. Loey, "Arabic Handwritten Characters Recognition using Convolutional Neural Network," *WSEAS Trans. Comput. Res.*, vol. 5, no. January, pp. 11–19, 2017.
- [7] N. A. Memon, F. Abbasi, and S. Zardari, "Glyph Identification and Character Recognition for Sindhi OCR," *Mehran Univ. Res. J. Eng. Technol.*, vol. 36, no. 4, pp. 933–940, 2017.
- [8] N. Nizamani, B. S. Chowdhry, M. Shaikh, and A. S. Saand, "Comparing Thinning Based Morphological operation on Sindhi," *IJCSNS Int. J. Comput. Sci. Netw. Secur.*, vol. 18, no. 2, pp. 126–129, 2018.
- [9] A. M. Nizamani and P. N. U. H. Janjua, "Isolated Handwritten Character Recognition in Sindhi Language using Artificial Neural Network," vol. 10, no. 1, pp. 17–23, 2012.
- [10] W. A. Narejo and J. A. Mahar, "Morphology: Sindhi morphological analysis for natural language processing applications," *2016 Int. Conf. Comput. Electron. Electr. Eng. ICE Cube 2016 - Proc.*, pp. 27–31, 2016.

- [11] A. A. CHANDIO, A. H. JALBANI, M. LEGHARI, and S. A. AWAN, "Multi-Digit Handwritten Sindhi Numerals Recognition using SOM Neural Network," *Mehran Univ. Res. J. Eng. Technol.*, vol. Volume 36, no. October, pp. 901–908, 2017.
- [12] D. N. Hakro, M. A. Hakro, and I. A. Lasharai, "Sindhi Named Entity Recognition (Sner)," *Gov. J. Polit. Sci. Vol. V, Suppl. Ed.*, vol. V.
- [13] J. J. Hull, "A Database for Handwritten Text Recognition Research," *IEEE Trans. PA'TERN Anal. Mach. Intell.*, vol. 16, no. 5, pp. 550–554, 1994.
- [14] Y. Alginahi, "Preprocessing Techniques in Character Recognition," *Character Recognit.*, pp. 1–22, 2012.
- [15] M. R. Gupta, N. P. Jacobson, and E. K. Garcia, "OCR binarization and image pre-processing for searching historical documents," *Pattern Recognit.*, vol. 40, no. 2, pp. 389–397, 2007.
- [16] R. Chandwadkar, S. Dhole, V. Gadewar, D. Raut, and S. A. Tiwaskar, "Comparison Of Edge Detection Techniques," *6th Annu. Conf. IRAJ*, no. 6, pp. 133–136, 2013.
- [17] M. Sonkusare and N. Sahu, "A SURVEY ON HANDWRITTEN CHARACTER RECOGNITION (HCR) TECHNIQUES FOR ENGLISH," *Adv. Vis. Comput. An Int. J.*, vol. 3, no. 1, 2016.
- [18] M. 'Arif, H. Hassan, D. Nasien, and H. Haron, "A Review on Feature Extraction and Feature Selection for Handwritten Character Recognition," *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, no. 2, pp. 204–212, 2015.
- [19] S. Afzal and M. A. Wani, "Comparative Study of Adaptive Learning Rate with Momentum and Resilient Back Propagation Algorithms for Neural Net Classifier Optimization," vol. 2, no. 1, 1986.



Sania Bhatti - She is working with the Department of Software Engineering, Mehran UET, Jamshoro Sindh, Pakistan. She obtained her PhD from the University of Leeds, the United Kingdom in 2010 under the scholarship of the faculty development program. Her research interests include modelling and simulation, communication networks and software engineering. She has published 21 national and international Journal papers and various international conference papers.



Bhawani Shankar Chowdhry - He is a renowned Scientist and ICT expert, Prof. BS Chowdhry is an Emeritus Professor and former Dean, Faculty of Electrical, Electronics, and Computer Engineering, Mehran UET Jamshoro. He has varied experience in diversified fields such as Electronics, Telecommunication, Microprocessor Technology, Internet of Things, and Telemedicine, Wireless sensor network etc. After completing his first degree in an electronic engineering discipline in 1983, he did his PhD in Microprocessor Based Intelligent Instrumentation from the school of Electronics and Computer Science University of Southampton, UK in 1990.



Sapna Kumari – She is Final year student, doing her B.E from software engineering Mehran UET Jamshoro. Currently she is working on her final year project titled "Indoor object identification for blind using machine learning". Her research interests are artificial intelligence, computer vision.

Authors' Profiles



Arsha Kumari – She has completed her B.E in 2016 from Department of Electronic engineering Mehran University of Engineering Technology Jamshoro. Currently, she is doing her Master's in Electronic System Engineering from Mehran UET Jamshoro. She has done many internships related to her field from

which she learnt a lot. Her research interests includes Machine Learning, Computer Vision, and Deep learning.



Din Muhammad Sangrasi – He is Assistant professor in the department of software engineering Mehran University of engineering technology Jamshoro. He has done his B.E in software engineering and M.E in a Communication system and Networks. He is currently doing his PhD in Australia.

How to cite this paper: Arsha Kumari, Din Muhammad Sangrasi, Sania Bhatti, Bhawani Shankar Chowdhry, Sapna Kumari, "Off-line Sindhi Handwritten Character Identification", *International Journal of Information Technology and Computer Science(IJITCS)*, Vol.11, No.6, pp.9-17, 2019. DOI: 10.5815/ijitcs.2019.06.02