Mehran University Research Journal of Engineering and Technology
Vol. 40, No. 1, 152 - 159, January 2021
p-ISSN: 0254-7821, e-ISSN: 2413-7219
DOI: 10.22581/muet1982.2101.14

Zernike Moments Based Handwritten Pashto Character Recognition Using Linear Discriminant Analysis

Sardar Jehangir[1a], Sohail Khan[1b], Sulaiman Khan[1c], Shah Nazir[1d], Anwar Hussain[1e]

RECEIVED ON 16.05.2019, ACCEPTED ON 03.03.2020

ABSTRACT

This paper presents an efficient Optical Character Recognition (OCR) system for offline isolated Pashto characters recognition. Developing an OCR system for handwritten character recognition is a challenging task because of the handwritten characters vary both in shape and in style and most of the time the handwritten characters also vary among the individuals. The identification of the inscribed Pashto letters becomes even palling due to the unavailability of a standard handwritten Pashto characters database. For experimental and simulation purposes a handwritten Pashto characters database is developed by collecting handwritten samples from the students of the university on A4 sized page. These collected samples are then scanned, stemmed and preprocessed to form a medium sized database that encompasses 14784 handwritten Pashto character images (336 distinguishing handwritten samples for each 44 characters in Pashto script). Furthermore, the Zernike moments are considered as a feature extractor tool for the proposed OCR system to extract features of each individual character. Linear Discriminant Analysis (LDA) is followed as a recognition tool for the proposed recognition system based on the calculated features map using Zernike moments. Applicability of the proposed system is tested by validating it with 10-fold cross-validation method and an overall accuracy of 63.71% is obtained for the handwritten Pashto isolated characters using the proposed OCR system.

Keywords: Linear Discriminant Analysis, Zernike Moments, 10-Fold Cross-Validation, Pashto, OCR.

1. INTRODUCTION

In the last decades, a lot of research has been reported on machine learning and pattern identification problems. Optical Characters Recognition (OCR) is a significant problem of research for the researchers in the pattern recognition. OCR converts images of text into computer readable format. State of the art techniques are suggested for different languages like English, Chinese, Arabic, Hindi, Dari, Persian and other around the world and high accuracy results are calculated for these languages. Cursive script languages like Arabic, Pashto and Urdu are the open research fields due to complexity in writing and word formation. Also, writing styles of these languages are varying for different peoples, and even it varies slightly for the same person on different occasions. These are the main problems that encounter hurdles in attaining state of the art performances in cursive-script based languages.

As per the study of literature, no research work has been reported for handwritten Pashto characters recognition. Boufenar *et al.* [1] presented an Artificial Immune Recognition (AIR) system using 7 types of features including both statistical and structural features for offline Arabic letters recognition. Abandah and Anssari [2] presented the concept of recognizing handwritten Arabic letters using Normalized Central Moments (NCMs) and Zernike Moments (ZMs) features based on Support Vector

[1] Department of Computer Science, University of Swabi, KP, Pakistan. Email: [a]sardarjehangir88@gmail.com, [b]mukhlisdagai@gmail.com, [c]engr.sulaiman88@gmail.com (Corresponding Author), [d]shahnazir@uoswabi.edu.pk, [e]anwar@uoswabi.edu.pk

Machines (SVM). Bhuiyan and Alsaade [3] presented a hybrid neural network approach for Arabic character recognition. The system is composed of a Bidirectional Associative Memory and a Multi-Layer Perceptron (BAMMLP). Classification/Accuracy results are calculated for the system in less than 1ms. Oujaoura *et al.* [4] suggested a method for offline Arabic letters identification using three feature techniques including zernike moments in conjunction with neural networks. Zernike moments surpass rest of the two in recognition rate. Sulaiman *et al.* [5] presented the use of KNN and artificial neural network for handwritten Pashto characters recognition bases on zoning feature extraction tool.

Naz *et al.* [6] investigates the study of Urdu Nastali'q text recognition using multiple geometrical features and Multi-Dimensional Long Short-Term Memory neural networks (MDLSTM). Jameel *et al.* [7] presented the concept of basis spline (B-Spline) curves as feature extractor and Neural Network for the recognition of isolated Urdu characters. Jameel and Kumar [8] also proposed basis spline curves in conjunction with Artificial Neural Network (ANN) for offline handwritten Urdu characters. Ahmed *et al.* [9] presented an algorithm for Urdu letters identification using Bidirectional Long Short-Term Memory (BLSTM) system. They also introduced a new database Urdu-Nasta'liq handwritten dataset. This paper presents an OCR system for isolated handwritten Pashto characters recognition. Pashto language has character set of 44 letters. It shares the same cursive style as that of Arabic, Persian and Urdu. Pashto language text is written from right-to-left side.

The paper is organized as follow; Section 2 gives the details of related work to Pashto text recognition. Section 3 gives detail about the Pashto script. While section 4 describes the proposed methodology of the OCR system based on Zernike moments as a feature extractor for the isolated handwritten Pashto characters recognition. Section 5 explains the results of the proposed research followed by conclusion in section 6.

2. RELATED WORK

For the last few decades, the handwritten character recognition is a prominent research problem in the field of image processing and machine learning. Significant improvements in OCR system for languages like English, Chinese and Japanese have been made. The languages like Arabic, Persian and Urdu still needs an effective handwritten character recognition system. The main problem associated with these languages is its cursive writing style. Pashto language shares same cursive nature. Several studies on automatic OCR system have been reported for languages like Arabic, Persian and Urdu, but no work has been reported on handwritten Pashto characters. A little work has been reported for printed letters recognition in Pashto language like Ahmad *et al.* [10] investigates the study of developing optical character recognition system for Pashto printed characters using k nearest neighbors.

Tavoli *et al.* [11] proposed a new feature extractor for the recognition of Arabic and Persian words, namely Statistical Geometric Components of Straight Lines (SGCSL) technique. Abandah and Anssari [2] presented the concept of recognizing handwritten Arabic letters using NCMs and ZMs features based on SVM as a classifier. Bhuiyan and Alsaade [3] presented a hybrid neural network approach for Arabic character recognition which contained a bidirectional associative memory and a multilayer perceptron (BAMMLP). Boufenar *et al,* [1] presented an Artificial Immune Recognition (AIR) using 7 types of features including both statistical and structural features for offline handwritten Arabic character recognition.

Sahlol *et al.* [12] presented an Arabic OCR system using a number of optimizers. CENPARMI dataset was used for testing of the system using three classifiers Linear Discriminant Analysis (LDA), SVM and Random Forest Trees (RFT). Oujaoura et al. [4] suggested a method for offline Arabic character recognition using three feature extractors in conjunction with neural networks. Zernike moments surpasses rest of the two in recognition rate. Aranian *et al.* [13] proposed a hybrid approach using artificial neural network, genetic algorithm and quantum genetic algorithm for the identification of Persian handwritten characters. They also performed feature dimensionality reduction on the datasets. Shafique *et al.* [14] suggested the concept of neural network for the recognition of handwritten Sindhi characters. Naz *et al.* [15] proposed entity recognition system in Urdu language using hybrid unigram and bigram approaches based on IJCNLP NE dataset and CRL NE dataset.

Naz *et al.* [16] presents a hybrid approach for Urdu nastali'q text recognition using hierarchical combination of Convolutional Neural Networks (CNN) and MDLSTM. They tested the system on Urdu Printed Text line Images (UPTI) dataset producing state of the art recognition results. Naz *et al.* [17] suggested geometrical features and multi-dimensional long short-term memory for Urdu Nasta'liq text recognition using sliding window technique. Naz *et al.* [18] suggested the use of zoning features and 2DLSTM networks for identification of Urdu text recognition. The system performance was evaluated on UPTI dataset. Ahmed et al. [9] presented an algorithm for Urdu letters identification using the BLSTM. They introduced a new database called the Urdu Nastali'q Handwritten Dataset (UNHD).

This paper presents an OCR system for offline isolated Pashto character using Zernike moments as feature extractor technique and LDA as a classification tool.

3. PASHTO SCRIPT

Pashto is the official language of Afghanistan and a major language of Pashtun tribe in northern areas (Khyber Pakhtunkhwa) of Pakistan. In census 2007 – 2009, it was estimated that about 40 – 60 millions of people around the globe are native speakers of this language [19]. It consist both in hard dialect and soft dialect. The soft dialect is termed as Southern while the hard dialect is known as Northern. Both are differ from each other on phonological basis. It is cursive in nature, and had borrowed all the characters of the Arabic script, Persian script and Urdu script with some modification and additional six characters specific to Pashto script to made 44 character dataset and is shown in Fig. 1.

4. PROPOSED METHODOLOGY

Any handwritten OCR system consists of mainly three major steps that are; the input (handwritten character images), a feature extractor tool (to calculate astute features from the handwritten characters) and a classification tool for the recognition purpose. For the proposed Handwritten Pashto Characters Recognition (HPCR) system, we have developed a handwritten Pashto characters database which is developed for input, Zernike moments is considered for feature extraction purposes, and linear discriminant analysis selection for recognition purpose. **Fig. 2** shows the proposed methodology of HPCR system.

1	ا	16	ر	31	ق
2	ب	17	ړ	32	ک
3	پ	18	ز	33	ګ
4	ت	19	ژ	34	ل
5	ټ	20	ږ	35	م
6	ث	21	س	36	ن
7	ج	22	ش	37	ڼ
8	ځ	23	ښ	38	و
9	چ	24	ص	39	ه
10	څ	25	ض	40	ي
11	ح	26	ط	41	ې
12	خ	27	ظ	42	ی
13	د	28	ع	43	ۍ
14	ډ	29	غ	44	ئ
15	ذ	30	ف		

Fig. 1: Pashto Character Set

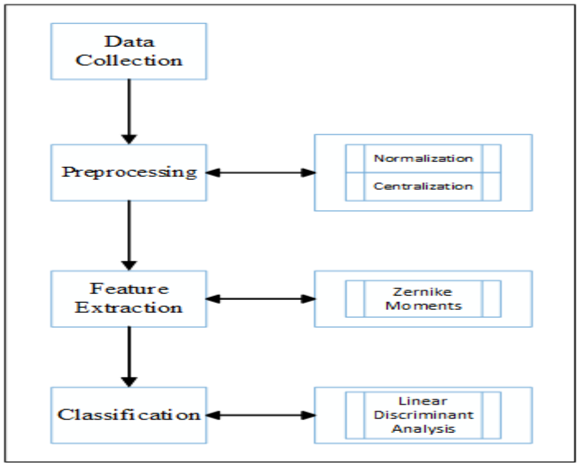


Fig. 2: Block Diagram for Proposed OCR

4.1 Database Development

There is no handwritten Pashto characters database available for simulation purpose. A database of 14784 handwritten Pashto characters samples is developed by collecting samples from students and teachers in University, varying in age, gender and educational backgrounds. Table 1 shows age-wise distribution of samples collected for handwritten Pashto characters.

Table 1: Age-Wise Distribution From People in Data Collection

Age	Gender		Total
	Male	Female	
Below 30 years	40	30	70
Above 30 & Below 50 years	30	15	45
Above 50 years	20	15	35
Total number of people involved			150

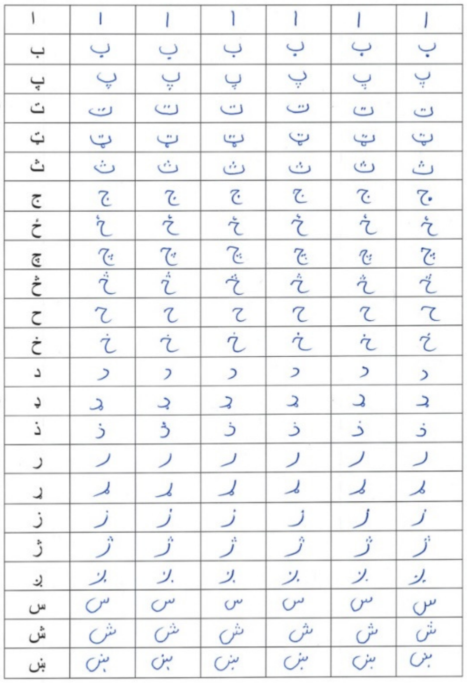
Pashto script consists of 44 characters, it is more difficult to set all the characters in one page, so the handwritten samples are calculated in two pages. First 23 characters are calculated on one page that is shown in Fig. 3, while the remaining 21 on the other page as shown in Fig. 4. The page is divided into six columns to get variant samples from different students.

These scanned images are then further processed in order to extract the individual character samples by applying preprocessing steps (that are discussed in the next section) to develop a database for the proposed OCR system.

4.2 Preprocessing

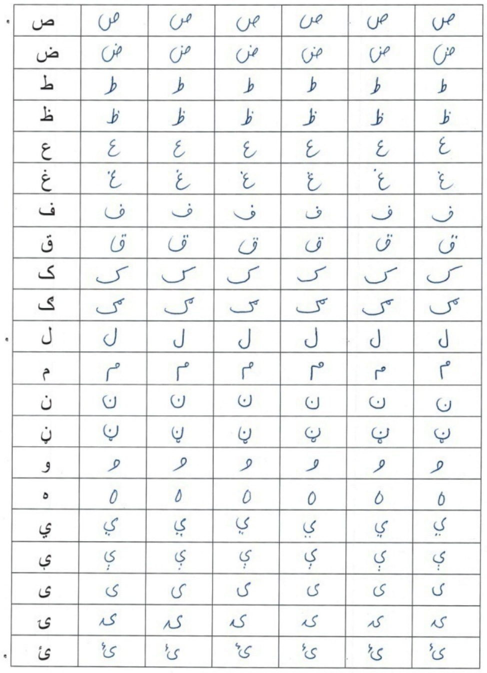
In order to extract uniform features, it is necessary to apply some filtering techniques to character database to make the images uniform. The database images contained noise and also characters appeared at different locations (left, right, top, bottom). The images are normalized to a fixed size character images.

In this research work, the noise (black dots) are removed using thresholding. In our experimental case, we come with an optimum threshold value of 30. After noise removal, some morphological operations of erusion and dilation were followed to fix the skeleton of the handwritten characters in the sliced images. By applying morphological operation, all the characters are centralized in sliced images and converted to a fixed size of 80 × 80. The result obtained is shown in Fig. 5.



ا	ا	ا	ا	ا	ا	ا
ب	ب	ب	ب	ب	ب	ب
پ	پ	پ	پ	پ	پ	پ
ت	ت	ت	ت	ت	ت	ت
ټ	ټ	ټ	ټ	ټ	ټ	ټ
ث	ث	ث	ث	ث	ث	ث
ج	ج	ج	ج	ج	ج	ج
ځ	ځ	ځ	ځ	ځ	ځ	ځ
چ	چ	چ	چ	چ	چ	چ
څ	څ	څ	څ	څ	څ	څ
ح	ح	ح	ح	ح	ح	ح
خ	خ	خ	خ	خ	خ	خ
د	د	د	د	د	د	د
ډ	ډ	ډ	ډ	ډ	ډ	ډ
ذ	ذ	ذ	ذ	ذ	ذ	ذ
ر	ر	ر	ر	ر	ر	ر
ړ	ړ	ړ	ړ	ړ	ړ	ړ
ز	ز	ز	ز	ز	ز	ز
ژ	ژ	ژ	ژ	ژ	ژ	ژ
ږ	ږ	ږ	ږ	ږ	ږ	ږ
س	س	س	س	س	س	س
ش	ش	ش	ش	ش	ش	ش
ښ	ښ	ښ	ښ	ښ	ښ	ښ

Fig. 3: First 23 Characters of Pashto Character Set



ص	ص	ص	ص	ص	ص	ص
ض	ض	ض	ض	ض	ض	ض
ط	ط	ط	ط	ط	ط	ط
ظ	ظ	ظ	ظ	ظ	ظ	ظ
ع	ع	ع	ع	ع	ع	ع
غ	غ	غ	غ	غ	غ	غ
ف	ف	ف	ف	ف	ف	ف
ق	ق	ق	ق	ق	ق	ق
ک	ک	ک	ک	ک	ک	ک
ګ	ګ	ګ	ګ	ګ	ګ	ګ
ل	ل	ل	ل	ل	ل	ل
م	م	م	م	م	م	م
ن	ن	ن	ن	ن	ن	ن
ڼ	ڼ	ڼ	ڼ	ڼ	ڼ	ڼ
و	و	و	و	و	و	و
ه	ه	ه	ه	ه	ه	ه
ي	ي	ي	ي	ي	ي	ي
ې	ې	ې	ې	ې	ې	ې
ی	ی	ی	ی	ی	ی	ی
ۍ	ۍ	ۍ	ۍ	ۍ	ۍ	ۍ
ئ	ئ	ئ	ئ	ئ	ئ	ئ

Fig. 4: Last 21 Characters of Pashto Character Set

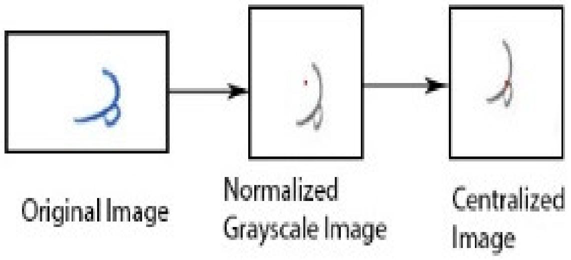


Fig. 5: Resultant Character after Preprocessing Steps

3. Feature Extraction

Feature extraction is a process by which the essential characteristics of the actual character image is represented in much lower dimensional space. For feature set generation, we have used Zernike moments and symmetric methods introduced by Khotanzad and Hong [20, 21]. Kef *et al.* [22] presented a novel based approach for Arabic characters recognition using Zernike moments for feature extraction purpose. Kan and Srinath [23] proposed developed an OCR system for invariant characters recognition systems based on Zernike moments and orthogonal Fourier-Millen moments. Singh *et al.* [24] suggested the use of moments for handwritten digits recognition. For this research work, Zernike moments are considered as a feature extraction tool. Zernike moments provide a set of features that are invariant under orthogonal, size and displacement transformation. This tool extracts the features by drawing a circle with a radius one (x2 + y2 = 1) at the center of the image as center of the unit circle. Zernike function is calculated as in equation 1,

$$V^{pq}(x, y) = R_{pq}(\rho)e^{jq\theta} \quad (1)$$

R_{pq} is the orthogonal radial polynomial given in equation 2,

$$R_{pq}(\rho) = \sum_{s=0}^{\frac{p-|q|}{2}} (-1)^s \frac{(p-s)!}{s!\left[\frac{p+|q|}{2}-s\right]!\left[\frac{p-|q|}{2}-s\right]!} p^{p-2s} \quad (2)$$

Here we are calculating zernike moments for digital image with order p and repetition q (q-p must be even), so zernike moments are given by equation (3):

$$Z_{pq} = \frac{p+1}{\pi}\sum_x \sum_y f(x, y)V_{pq}^*(\rho, \theta), \qquad x^2 + y^2 \leq 1 \quad (3)$$

Thirty-six Zernike features are calculated for an individual image up to order 10 and repetition 10.

4.4 Classification

Classification is the most important step in an OCR system development. An efficient linear discriminant analysis (LDA) classifier is proposed to classify the handwritten Pashto characters based on the Zernike features map calculated. LDA is a generalization of the Fisher discriminant analysis introduced by Ronal Fisher a British statistician and geneticist [25]. Hasasneh *et al.* [24] presented the use of unsupervised Deep Belief Network (BDN) for offline handwritten Arabic characters recognition. Naz *et al*, [25] suggested a Ghost Character Theory (GCT) for multilingual Arabic characters recognition. GCT uses the concept of back propagation neural network (BPNN). As the Arabic text follows the Nasah and Nasta'liq style, so the proposed system developed by Naz *et al*, [25] ultimately works for Urdu scripts as these are the sibling languages. Memon *et al.* [26] developed an OCR for glyphs and Sindhi characters recognition. In this approach the glyphs are successfully identified from scanned images and the characters are recognized. Awan *et al.* [14] presented the concept of neural network for the recognition of handwritten Sindhi characters based on zoning feature extraction algorithm. However all these techniques works good and are highly applicable for the problems addressed, but unfortunately all these techniques fails due to large characters dataset in Pashto script and varying characters in this language. This paper suggests the use of LDA technique for classification purposes in the proposed OCR system. LDA is generally a dimensionality reduction tool and outperforms in multi-class problems. This technique works by picking a new dimension that gives the maximum separation between the means of the projected classes, and minimum variance within each projected class. Fig. 5 presents a generalized model of the LDA classification tool.

Pashto script consists of 44 characters (in other words it contains 44 classes) and it ultimately specifies a multi-class problem. Fig. 6 represents a conventional multi-class LDA classifier system.

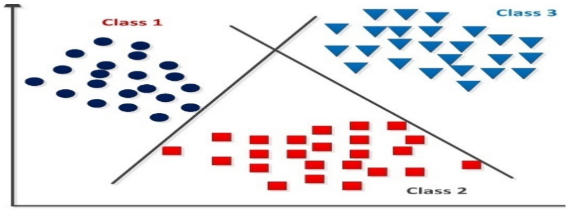


Fig. 6: Conventional LDA Model for HPCR System

5. RESULTS AND DISCUSSIONS

Recognition results are generated based on LDA classification tool. These recognition results are simulated based on the Zernike feature map calculated in the previous step. For training and testing purposes, the calculated feature map is divided into 2: 1. Based on this ratio, an accuracy of 63.71% is calculated on 10-fold cross-validation method. LDA is tested for variant training and test sets. The training sets starts from 50%, 55%, 60%, 65%, 70%, 75% and 80%, while the remaining are considered as test set. For the proposed variant training and test set accuracy is calculated that is shown in Fig. 7.

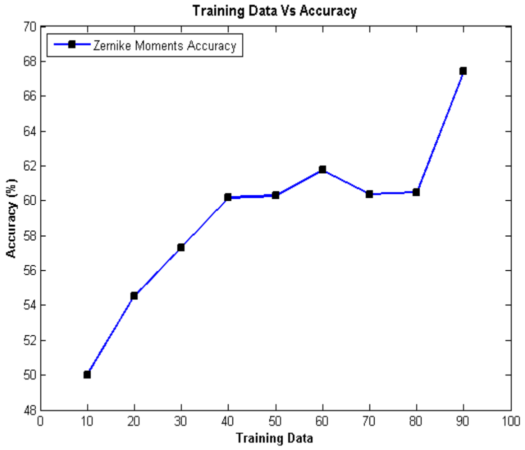


Fig. 7: Training Set vs Accuracy Graph

It is evident from the Fig. 7, that as the training set increases, the recognition accuracy of the proposed PHCR system also increases.

Time consumed for each variant training and test sets is calculated for the proposed PHCR system, and a graph is generated based on the variant sets vs accuracy and time consumption. Classification accuracy and time consumption graph is plotted based on varying training and test sets that is shown in Fig. 8.

It is evident from Fig. 8, that when the training set increases the accuracy of the proposed PHCR system increases along with time consumption for the proposed PHCR system. After performing simulations for varying training and test sets based on the Zernike features map, 63.71% is the highest accuracy results achieved for the proposed PHCR system.

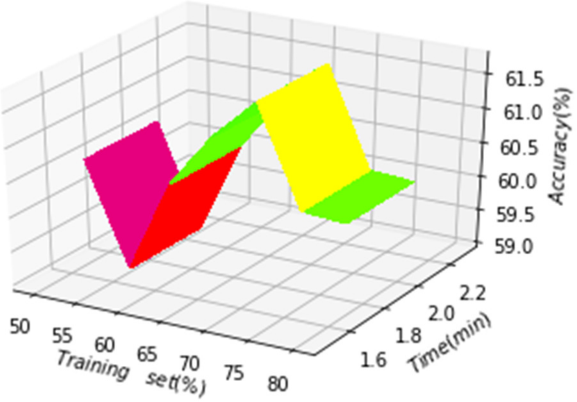


Fig. 8: Training Set Vs Time And Accuracy Graph

6. CONCLUSION

In this paper, an OCR system for recognition of handwritten Pashto characters is used. A Medium size database of 14784 characters is developed by collecting samples from different people varying in age, gender and educational backgrounds in University. Zernike moments invariants are considered as a feature extractor tool in the proposed OCR system. While an efficient LDA classifier is used to classify the individual character images. An accuracy result of 63.71% was calculated using 10-fold cross validation.

In future, we tend to improve the accuracy of the system by using different combinations of features and classifiers methods. Also, we want to increase database samples to achieve high accuracy, and to improve this work for word/script recognition.

ACKNOWLEDGEMENT

The authors acknowledge the University of Swabi for providing all the resources for conducting the research and participants of experts for providing different shapes of Pashto text for the datasets.

REFERENCES

[1] Boufenar C., Batouche M., Schoenauer M., "An artificial immune system for offline isolated handwritten arabic character recognition", *Evolving Systems*, Vol. 9, pp. 25-41, 2018.

[2] Abandah G., Anssari N., "Novel moment features extraction for recognizing handwritten Arabic letters," *Journal of Computer Science*, Vol. 5, p. 226, 2009.

[3] Bhuiyan M.A.A., Alsaade F.W., "On Arabic Character Recognition Employing Hybrid Neural Network," *International Journal of Advanced Computer Science and Applications*, Vol. 8, pp. 96-101, 2017.

[4] Oujaoura M., El Ayachi R., Fakir M., Bouikhalene B., Minaoui B., "Zernike moments and neural networks for recognition of isolated Arabic characters", *International Journal of Computer Engineering Science*, Vol. 2, pp. 17-25, 2012.

[5] Khan S., Ali H., Ullah Z., Minallah N., Maqsood S., Hafeez A., "KNN and ANN-based Recognition of Handwritten Pashto Letters using Zoning Features", *International Journal of Advanced Computer Science and Applications*, Vol. 9, 2018.

[6] Naz S., Umar A. I., Ahmed S. B., Ahmad R., Shirazi S. H., Razzak M. I., Zamari A., "Statistical Features Extraction For Character Recognition Using Recurrent Neural Network," *Pakistan Journal of Statistics*, Vol. 34, No.1, 2018.

[7] Jameel M., Kumar S., Karim A., "A Review on Recognition of Handwritten Urdu Characters Using Neural Networks", *International Journal of Advanced Research in Computer Science*, Vol. 8, 2017.

[8] Jameel M. Kumar S., "Offline Recognition of Handwritten Urdu Characters using B Spline Curves: A Survey", *International Journal of Computer Applications*, Vol. 157, 2017.

[9] Ahmed S. B., Naz S., Swati S., Razzak M. I., "Handwritten Urdu character recognition using one-dimensional BLSTM classifier," *Neural Computing and Applications*, pp. 1-9, 2017.

[10] Ahmad N., Khan A. A., Abid S. A. R., Yasir M., "Pashto Isolated Character Recognition Using K-NN Classifier," *Sindh University Research Journal* (Science Series), Vol. 45, 2013.

[11] Tavoli R., Keyvanpour M., Mozaffari S., "Statistical geometric components of straight lines (SGCSL) feature extraction method for offline Arabic/Persian handwritten words recognition", *IET Image Processing*, Vol. 12, pp. 1606-1616, 2018.

[12] Sahlol A. T., Elhoseny M., Elhariri E., Hassanien A. E., "Arabic handwritten characters recognition system, towards improving its accuracy", *Proceedings of the IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)*, pp. 1-7, 2017.

[13] Aranian M. J., Sarvaghad-Moghaddam M., Houshmand M., "Feature dimensionality reduction for recognition of Persian handwritten letters using a combination of quantum genetic algorithm and neural network," *Majlesi Journal of Electrical Engineering*, Vol. 11, 2017.

[14] Awan S. A., Abro Z. H., Jalbani A. H., Hameed M., "Handwritten Sindhi Character Recognition Using Neural Networks," *Mehran University Research Journal of Engineering and Technology*, Vol. 37, p. 6, 2018.

[15] Naz S., Umar A. I., Razzak M. I., "A hybrid approach for NER system for scarce resourced language-Urdu: Integrating n-gram with rules and gazetteers," *Mehran University Research Journal of Engineering & Technology*, Vol. 34, p. 349, 2015.

[16] Naz S., Umar A. I., Ahmad R., Siddiqi I., Ahmed S. B., Razzak M. I., Shafait F., "Urdu Nastaliq recognition using convolutional–recursive deep learning", *Neurocomputing*, Vol. 243, pp. 80-87, 2017.

[17] Naz S., Umar A.I., Ahmad R., Ahmed S. B., Shirazi S. H., Razzak M. I., "Urdu Nasta'liq text recognition system based on multi-dimensional recurrent neural network and statistical features", *Neural Computing and*

Applications, Vol. 28, pp. 219-231, 2017.

[18] Naz S., Ahmed S. B., Ahmad R., Razzak M. I., "Zoning features and 2DLSTM for Urdu Text-Line Recognition", *Procedia Computer Science*, Vol. 96, pp. 16-22, 2016.

[19] Khan S., Ali H., Ullah Z., Minallah N., Maqsood S., Hafeez A., "KNN and ANN-based Recognition of Handwritten Pashto Letters using Zoning Features", *Machine Learning*, Vol. 9, 2018.

[20] Khotanzad A., Hong Y. H., "Invariant image recognition by Zernike moments", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, pp. 489-497, 1990.

[21] Khotanzad A., Hong Y. H., "Rotation invariant image recognition using features selected via a systematic method", *Pattern Recognition*, Vol. 23, pp. 1089-1101, 1990.

[22] Kef M., Chergui L., Chikhi S., "A novel fuzzy approach for handwritten Arabic character recognition", *Pattern Analysis and Applications*, Vol. 19, pp. 1041-1056, 2016.

[23] Kan C., Srinath M. D., "Invariant character recognition with Zernike and orthogonal Fourier–Mellin moments," *Pattern Recognition*, Vol. 35, pp. 143-154, 2002.

[24] Singh P. K., Sarkar R., Nasipuri M., "A study of moment based features on handwritten digit recognition", *Applied Computational Intelligence and Soft Computing*, Vol. 2016, 2016.

[25] Fisher R. A., "The use of multiple measurements in taxonomic problems", *Annals of Eugenics*, Vol. 7, pp. 179-188, 1936.

[26] Memon N. A., Abbasi F., Zardari S., Glyph Identification and Character Recognition for Sindhi OCR, *Mehran University Research Journal of Engineering and Technology*, Vol. 36, No. 4, 2017.