

KNN and ANN-based Recognition of Handwritten Pashto Letters using Zoning Features

Sulaiman Khan¹, Hazrat Ali², Zahid Ullah^{3*}, Nasru Minallah⁴, Shahid Maqsood⁵, and Abdul Hafeez^{6*}

Computer Science, University of Swabi, Pakistan¹,

Department of Electrical Engineering, COMSATS University Islamabad, Abbottabad Campus, Abbottabad Pakistan²,

Electrical Engineering, CECOS University, Pakistan³

Computer Science, UET Jalozai, Pakistan^{4,6},

Industrial Engineering, UET Jalozai, Pakistan⁵

Abstract—This paper presents an intelligent recognition system for handwritten Pashto letters. However, handwritten character recognition is challenging due to the variations in shape and style. In addition to that, these characters naturally vary among individuals. The identification becomes even daunting due to the lack of standard datasets comprising of inscribed Pashto letters. In this work, we have designed a database of moderate size, which encompasses a total of 4488 images, stemming from 102 distinguishing samples for each of the 44 letters in Pashto. Furthermore, the recognition framework extracts zoning features followed by K-Nearest Neighbour (KNN) and Neural Network (NN) for classifying individual letters. Based on the evaluation, the proposed system achieves an overall classification accuracy of approximately 70.05% by using KNN, while an accuracy of 72% through NN at the cost of an increased computation time.

Keywords—KNN; deep neural network; OCR; zoning technique; Pashto; character recognition; classification

I. INTRODUCTION

In this modern technological and digital age, optical character recognition (OCR) systems play a vital role in machine learning and automatic recognition problems. OCR is a section of the software tool that converts printed text and images to machine-readable form and enables the device to recognize images or text like humans. OCR systems are commercially available for separate languages, which include Chinese, English, Japanese, and others. However, limited OCR-based systems are available for cursive languages such as Persian and Arabic and are not highly robust. To the best of our knowledge, commercial OCRs do not exist for carved Pashto letters recognition except in research labs.

Handwritten letters recognition is a daunting task mainly because of variations in writing styles of different users. Handwritten letters recognition can be done either offline or online. Online character recognition is simpler and easier to implement due to the temporal-based information such as velocity, time, number of strokes, and direction for writing. Besides, the trace of the pen is a few pixels wide, so thinning techniques are not viable here. On the other hand, the implementation of an offline recognition system implementation is even laborious due to high variations in writing and font styles of different users.

Pashto is a major language of Pashtun tribe in Pakistan and the official language of Afghanistan. In census of 2007–2009, it was estimated that about 40–60 millions of people around the world are native speakers of this language.

Pashto letters can be shaped into six different formats, which make the recognition process challenging. Furthermore, the count of character dots and occurrence of these dots that varies between letters make the problem challenging.

Research shows the use of high-level features based on the structural information of letters. An OCR-based system using deep learning network model that incorporates Bi- and Multi-dimensional short-term memory for printed Pashto text recognition has been suggested [1].

A web-based survey shows that Pashto script contains a considerable number of unique ligature [2]. Such ligature poses challenges on the implementation of an OCR-based system for identifying carved Pashto letters. As printed letters contain a constant shape/style and font size; thus, the said technique fails due to higher variations in style and font in case of inscribed letters. Riaz et al. [3] has presented the development of an OCR system for cursive Pashto script using scale invariant feature transform and principle component analysis. This work presents a system for handwritten Pashto letters recognition, which has the following key contributions:

- Designed and developed a medium-sized database of 4488 (102 samples for each letter) for further research work.
- Provided a baseline result for the identification of inscribed Pashto letters using KNN and deep Neural Network and zoning features.
- Evaluated and provided comprehensive results through the proposed system for handwritten Pashto letters recognition, which may help the researchers to further explore this area.

The proposed approach is efficient, simple, and cost-effective. This paper is divided in seven sections: Section II explains the related work. Section III captures the background information about the classifiers and feature extraction algorithm used in this research work. Section IV delineates the methodology. Section V discusses about the feature extraction, which is very important in the area of pattern recognition and

*Corresponding Author: zahidullah@cecos.edu.pk, abdul.hafeez@uetpeshawar.edu.pk

machine learning while section VI demonstrates the experimental results followed by the conclusions and future work in Section VII.

II. RELATED WORK

Pashto, Persian, Urdu, and Arabic are sister languages. Several diverse approaches are suggested by different researchers for developing an OCR system for these languages. However, Pashto script contains 44 letters, greater than Arabic script which are 28 though comprehensive, Persian script comprising of 32 letters, and Urdu script encompassing 38 letters. Pashto language encapsulates all the letters from the Urdu script with additional seven letters. This additional seven letters in Pashto makes the traditional OCRs incapable to recognize handwritten Pashto letters. Some of the closely related work on the prescribed languages is mentioned below.

Abdullah et al. [4] presented an OCR system for Arabic handwriting recognition based on Neural Network classifier for classifying an IFN—ENIT dataset. Ahmad et al. [5] presented a novel approach of gated bidirectional long short term memory (GBLSTM) for recognition of printed Urdu Nastaliq text, which is a special form of Neural Network based on ligature information of the printed text. Ahmed et al. [6] used a one dimensional BLSTM for handwritten Urdu letter recognition where a medium size database for handwritten Urdu letters collected from 500 people was developed.

Alotaibi et al. [7] suggested an algorithm to develop an OCR that can check the originality and similarity of online Quranic contents where Quranic text is a combination of diacritics and letters. For diacritic detection, they used region-based algorithms and projection method is used for letter detection. The results of the similarity indices are compared with standard Mushaf Al Madina benchmark. Boufenar et al. [8] presented the concept of supervised learning technique named Artificial immune system based on zoning technique for isolated carved Arabic letters recognition. Jameel and Kumar [9] suggested the use of B spline curves as a feature extractor for offline Urdu character recognition. Naz et al. [10] [11] presented the use of multi-dimensional recurrent Neural Network based on statistical features for Urdu Nastaliq text recognition. Rabi et al. [12] performed a survey on different OCR systems for handwritten cursive Arabic and Latin script recognition where it was concluded that the results of contextual sub character of Hidden Markov Models were proven with high accuracy for handwritten Latin and Arabic script recognition.

Rouini et al. [13] presented the use of dynamic random forest classifier based on surf descriptor feature extraction technique. Sahlol et al. [14] inspected different classifiers Genetic algorithm (GA), Particle Swarm optimization (PSO), Grey Wolf optimization (GWO), and BAT algorithms (BAT) for handwritten Arabic characters recognition. After testing each algorithm, it was concluded that GWO provides prominent results for handwritten Arabic characters recognition. As Sindhi language is a super set of Arabic language, Shaikh et al. [15] developed an OCR system for text recognition using an approach based on segmentation.

M. Kumar et al. [16] presented a comprehensive survey of Indic and non-Indic scripts on letters and numeral recognition. Zayene et al. [17] presented a novel approach for Arabic video

text recognition using recurrent Neural network. This system suggests a segmentation free method mainly based on a multi-dimensional version of long short term memory combined with a connectionist temporal classification layer. Veershetty et al. [18] suggested the concept of an optical character recognition (OCR) system for handwritten script recognition based on KNN, SVM, and linear discriminant analysis (LDA) classifiers. For feature extraction, they used a technique based on Radon and wavelet transform, and words were extracted using morphological dilation methods.

Malviya et al. [19] carried out a comparative study of various feature extractions techniques named Zernike moments, projection histogram, zoning methods, template machine, and chain coding technique and classification algorithms such as SVM and Artificial Neural Network (ANN) have been discussed. Some vital parameters are selected based on sample size, data types, and accuracy. Bhunia et al. [20] presented a novel approach for word level Indic-script recognition using character level data in input stage. This approach uses a multimodal Neural Network that accepts both offline and online data as an input to explore the information of both online and offline modality for text/script recognition. This multi-modal fusion scheme combines the data of both offline and online data, which indeed a real scenario of data being fed to the network. The validity of this system was tested for English and six Indian scripts. Obaidullah et al. [21] carried out a comprehensive survey for the development of an OCR system for Indic script recognition in multi-script document images. Multiple pre-processing techniques, feature extraction techniques, and classifiers used in script recognition were discussed.

The literature review shows that a little work is available on the development of an OCR system for the recognition of printed Pashto letters; however, there is no OCR system developed for automatic recognition of handwritten Pashto letters. All the above mentioned algorithms perform well for the specified languages but fail in recognizing the handwritten Pashto letters owing to the extra number of letters in the character set. In this paper, we present a robust OCR system for the recognition of handwritten Pashto letters having the key benefits mentioned above.

III. BACKGROUND STUDY

This section describes the detail of the character modeling for Pashto script, classification techniques followed by KNN, and Neural Network classifiers.

A. Pashto

Pashto is the language of Pashtuns, often pronounced as Pakhto/Pukhto/Pushto and is the official language of Afghanistan and a major language of Pashtun clan in Pakistan. In Persian literature, it is known as Afghani while in Urdu or Hindi literature, it is known as Pathani. Pashto has two major dialects namely soft dialect and hard dialect. Both of these dialects are phonologically differ from each other. The soft dialect is called southern while the hard dialect is known as northern. In soft dialect i-e., southern, Pushto is spelled as Pashto while in hard dialect i-e., in northern, it is spelled as Pukhto or Pakhto. The word Pashto is followed as a representation for both hard and soft dialects. The Kandahari form of

Pashto Character Set					
Name	Alphabet	Name	Alphabet	Name	Alphabet
Alif	ا	Re	ر	Qa'uf	ق
Be	ب	Rhe	ړ	Ka'f	ک
Pe	پ	Ze	ز	Ga'f	گ
The	ت	Jey	ژ	Laam	ل
Te	ټ	Ghey	ګ	Meem	م
Se	س	Seen	س	Noon	ن
Jeem	ج	Sheen	ش	Rnoon	ښ
Zeem	خ	Heen	ه	Wow	و
Che	چ	Swa'd	ص	Kha	ه
Seem	ش	Dwa'd	ض	Yee	ي
Hey	ح	Thwe	ط	Yey	ې
Khe	خ	Zwe	ظ	Ye	ی
Dhaal	د	Ain	ع	Yay	ی
Daal	ډ	Ghain	غ	Yai	ئ
Zaal	ذ	Fey	ف		

Fig. 1: Pashto characters dataset.

TABLE I: Urdu Specific Letters Representation in Pashto Script

Urdu Letter	Pashto equivalent
ٹ	ټ
ڈ	د
ڑ	ړ
گ	ک
ے	ی

Pashto dialect, also known as Pata Khazana, is considered as standard spelling system for Pashto script.

Pashto script consists of 44 letters shown in Fig. 1. The Name represents letter name while Alphabet represents letters shape in isolated form. It has borrowed all the letters from Persian script, i.e., 32 letters that has further borrowed the entire letter set, i.e., 28 letters from Arabic script. That is why Pashto is known as a modified pattern of Perso-Arabic characters. Urdu script adopts all 32 letters from Persian script with 6 additional letters. Pashto script encapsulates all the Urdu characters with minor change in these 6 special characters for Urdu script as shown in Table I. It encompasses additional 7 characters, especially to Pashto script forming a dataset of 44 characters as shown in Table II. In order to make a word in Pashto script, two or more than two isolated letters are combined to form a word. While defining a word, a letter shape changes w.r.t its position (start, middle or end) in the word as shown in the Table III. Both Naksh and Nastaliq is followed for Pashto script writing; however, Naksh is considered as standard writing style for Pashto script.

TABLE II: Pashto Specific Letters

7 Special Pashto characters						
خ	ش	ښ	ښ	ي	ې	ئ

TABLE III: Change in Letters Shape W.R.T. its Position in Word

Isolated Letter	Contextual form		
	Start	Middle	End
و	و	و	و
ظ	ظ	ظ	ظ

B. K-Nearest Neighbor (KNN)

KNN is a supervised learning tool used in regression and classification problems. In training phase, KNN uses multi-dimensional feature vector space that assigns a class label to each training sample. Many researchers have suggested the use of KNN classifier in text/digits recognition and classification such as Hazra et al [22] who presented the concept of KNN classifier for both handwritten and printed letters recognition in English language based on sophisticated feature extractor technique.

For online handwritten, Gujarati character recognition Naik et al. [23] suggested the use of SVM with polynomial, linear, and RBF kernel, KNN with variant values of K and multi-layer perception (MLPs) for stroke classification based on hybrid feature set. Selamat et al. [24] suggested the use of hybrid KNN algorithms for web paged base Arabic language identification and classification. They carried out the results based on SVM, back propagation neural network, KNN, and hybrid KNN. Zhang et al. [25] presented the use of KNN for visual category recognition based on text, color, and particularly shape in a homogeneous framework. Hasan [26] presented the concept of KNN classifier for Arabic(Indian) digits recognition using multi-dimensional features, which consist of discrete cosine transform (DCT) and projection methods.

KNN generates classification results by storing all the available cases and stratify new classes based on a similarity measure (distance functions). Pashto contains 44 letters in its character set so there are 44 classes to be classified. In short, it is a multi-class recognition problem. Fig. 2 represents a basic multi-class KNN model. In Fig. 2 class1, class2, and class3 represent 3 different classes. In our case, it contains 44 classes as there are 44 letters in Pashto character dataset.

C. Neural Network (NN)

NN has performed a vital role in the recognition and classification problems. Inspired from human nervous system, ANN is composed of layered architecture—input, hidden, and output layer. It contains a network of neurons connected through weighted connections that accepts input, performs processing, and produce detailed patterns. Machine learning (ML) has been widely used in a variety of applications. ML has been used in scheduling tasks in real time through cloud computing in the form of genetic algorithms [27]. Another

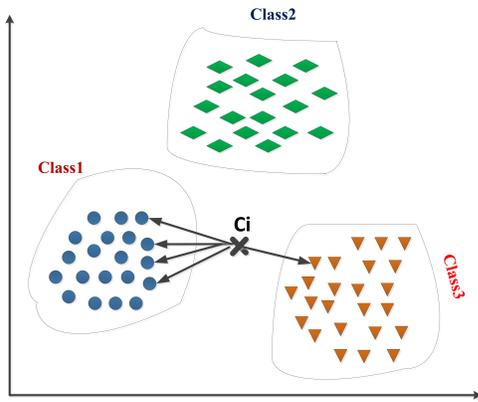


Fig. 2: Basic multi-class KNN basic model.

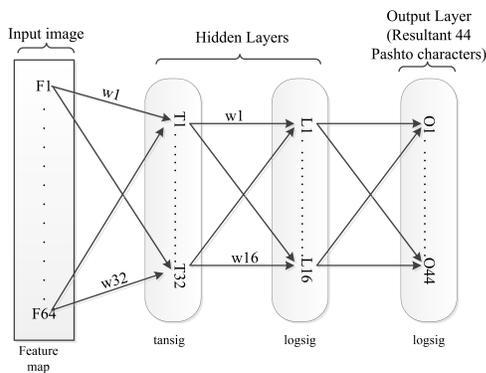


Fig. 3: Neural Network for handwritten Pashto character recognition.

study shows the use of ML models in genomics [28]. The goal is to detect variations and errors in Genomics datasets that entail higher variations. Decision trees and tabu search have been utilized in order to learn the dispatching rules for smart scheduling [29]. To explore the active learning, exponential gradient exploration has been studied [30].

Owing to NN's high identification and recognition abilities especially in text recognition problems, multiple researchers have suggested the use of this model, some of which are mentioned here. Jameel et al. [31] carried a review paper on Urdu character recognition using NN. In this paper, they suggested the use of B-Spline curves as a feature extractor technique for Urdu characters recognition. Zhang et al. [32] presented the use of recurrent NN for drawing and recognition purposes of Chinese language. Patel et al. [33] suggested the use of ANN for handwritten character recognition based on discrete wavelet transform as a feature extractor technique, which is based on accurate level of multi-resolution technique. A basis NN diagram for HPLR system is shown in Fig. 3. In this research work, a NN classifier is selected with two hidden layers and one input and output layer. A feature map of 16 distinct values based on zoning technique are fed at input layer and the expected results are calculated at the output layer.

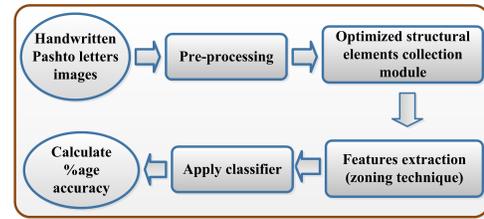


Fig. 4: The proposed Pashto handwritten letter recognition system.

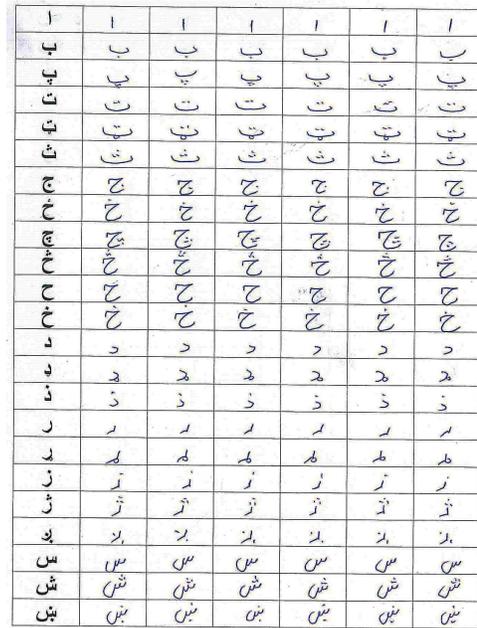


Fig. 5: First 23 handwritten Pashto characters.

IV. THE PROPOSED METHODOLOGY

The proposed OCR system for the recognition of handwritten Pashto letters is divided into three main steps as shown in Fig. 4.

- Database development for the handwritten Pashto letters.
- Feature selection/extraction.
- Classification and recognition using KNN and NN classifiers.

A. Database development for the handwritten Pashto letters

A medium size handwritten character database of 4488 characters (contains 102 samples for each letter) is developed by collecting handwritten samples from different individuals. These samples are collected on an A4 size paper divided into 6 columns for collecting a letter variant samples from same person. These samples are further scanned into computer readable format as shown in as shown in Fig. 5 and Fig. 6.

1) Letters Extraction: The letters are extracted in order to create a database. A few extracted letters are shown in

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
ك	ك	ك	ك	ك	ك	ك
گ	گ	گ	گ	گ	گ	گ
ل	ل	ل	ل	ل	ل	ل
م	م	م	م	م	م	م
ن	ن	ن	ن	ن	ن	ن
ښ	ښ	ښ	ښ	ښ	ښ	ښ
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
ى	ى	ى	ى	ى	ى	ى
ې	ې	ې	ې	ې	ې	ې
ځ	ځ	ځ	ځ	ځ	ځ	ځ

Fig. 6: Remaining 21 handwritten Pashto characters.

TABLE IV: A Table with Handwritten Sliced Pashto Characters

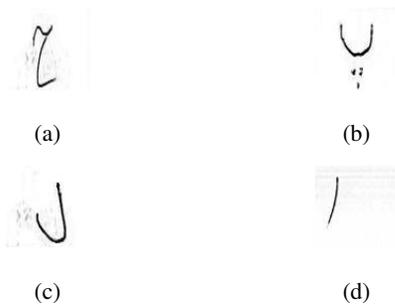


Table IV. All these extracted letters are resized into a fixed size of 44x44. This fixed size of the character helps in generating a uniform sized feature vector.

Each extracted letter in Table IV is hugely affected with dark spots i.e., noise, which is removed using thresholding. During the data collection phase, the handwritten character position varies in the 64x64 region/box. The reason is that letters can be written on top, left, right, and bottom of the box varying from person to person. We have centralized all the letters. Post-thresholding and centralizing results are captured in Table V.

V. FEATURE EXTRACTION

Selecting an astute, informative and independent feature is a crucial step for effective classification. This paper presents the concept of zoning method as a feature extractor technique for the recognition of handwritten Pashto letters.

A. Zoning Technique

This research work uses a 4x4 static grid to extract each letter features as shown in Fig. 7. By applying this zoning

TABLE V: A Table with Thresholded and Centralized Results on Sliced Characters

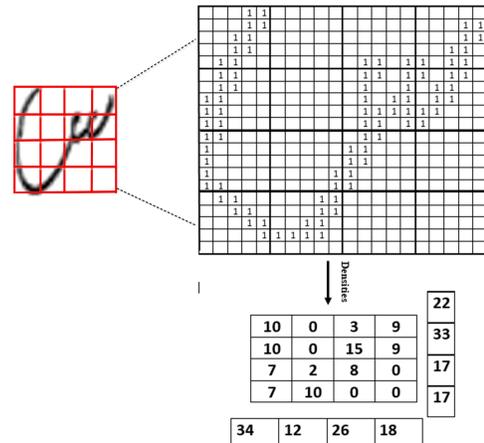
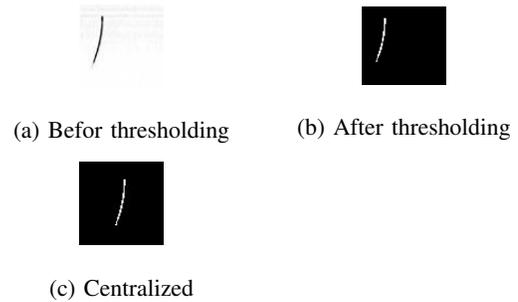


Fig. 7: Zoning feature extraction.

grid, it superimposes the pattern/character image and divides it into 16 equal zones. In each zone, the density of the letter is extracted that represents the ratio of the black pixels forming the letter on the total size of zone [34]. In this way, a feature map for all 4488 letters is obtained for classification.

After applying this technique a feature vector of 16 real values formed for each sample because we focus on zones not on the number of pixels.

VI. RESULTS

This section summarizes the results obtained after applying KNN and NN classifiers to handwritten Pashto letters for classification/recognition.

A. Classification Accuracy of K-Nearest Neighbours

The results of the KNN classifier for Pashto script recognition are shown in Fig. 8. The results are carried out using KNN classifier based on zoning features. The total image features for the Pashto letters is divided into a ratio of (2:1) for training and testing phases. The databases consists 102 samples for each Pashto letter. Thus, 68 letters features are selected for training phase and the remaining 34 letters features are selected for testing phase. An overall accuracy of about 70.05% is obtained for KNN, lesser than ANN, which is 72%.

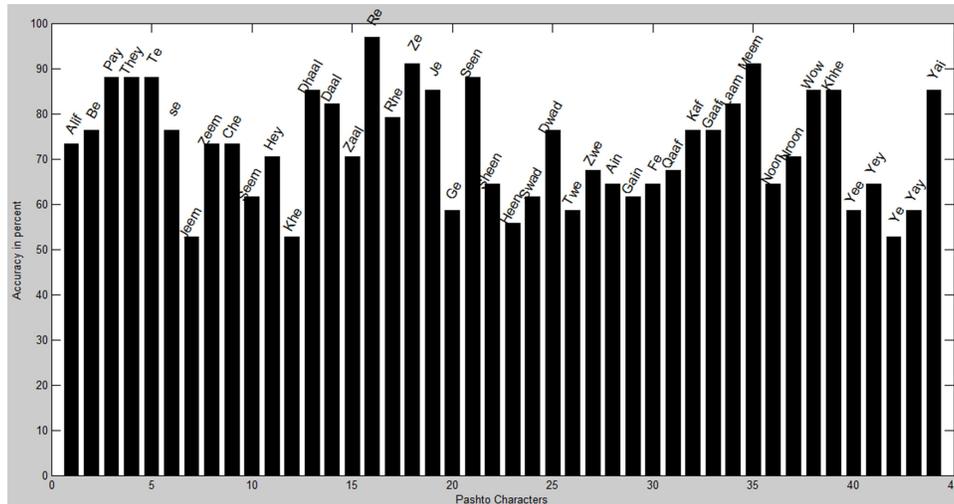


Fig. 8: KNN classifier accuracy results for HPLR system.

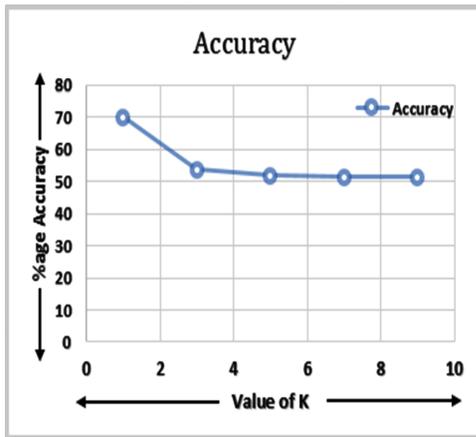


Fig. 9: KNN accuracy results for different values of K.

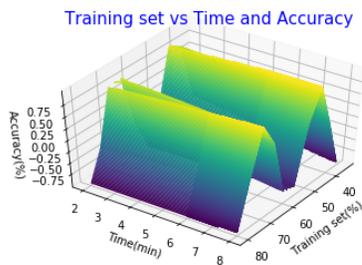


Fig. 10: NN results of HPLR system.

The accuracy of the KNN classifier is tested for different nearest neighbor values of K and it was detected that accuracy varies when the value of K increases, as the occurrence of other class features causes miss-classification. Fig. 9 represents the accuracy results drawn for varying values of K. High accuracy of KNN classifier is noted for the value of K equals to 1, because of high values of K causes the occurrence of other

class features that cause miss-classification.

B. Classification Accuracy of Neural Network Classifier

The feature map is divided into 2:1 for training and test data. NN classifier achieves an accuracy of about 72 % better than KNN classifier. Fig. 10 represents the overall result of NN classifier for Pashto letter recognition problem.

The efficiency of the classifier is tested for different size of training and test samples vs. time. The data is split into (training, test) sets of (35%, 65%), (40%, 60%), (45%, 55%), (50%, 50%), (55%, 45%), (60%, 40%), (65%, 35%), (70%, 30%), (75%, 25%), and (80%, 20%). The corresponding time and accuracy results are generated in Fig. 10. Where it is explained that when there is an increase in the training size, accuracy of the system increases. However, increasing the training size adversely affects the simulation time. A higher accuracy rate of 72% is carried out for 80% of training and 20% of test set.

Furthermore, the NN results based on varying epoch size for different training and test sets are also shown in Fig. 11. It is evident that as the number of epoch increases for given training and test sets, accuracy of the system increases. The mean square error error rate and gradient in the shape for handwritten Pashto letters is shown in Fig. 12.

VII. CONCLUSIONS AND FUTURE WORK

In this paper, an OCR system for automatic recognition of Pashto letters is developed by using KNN and NN classifiers based on zoning feature extractor technique. Experimental results show an accuracy of 70.07% for KNN while 72% for NN. Contributions include the provision of handwritten Pashto letters database as a resource for future research work and the experimental results, which will provide a baseline accuracy for future models tested on the data.

In future, we aim to extend and evaluate our technique for a larger database of Pashto script using an increasing number

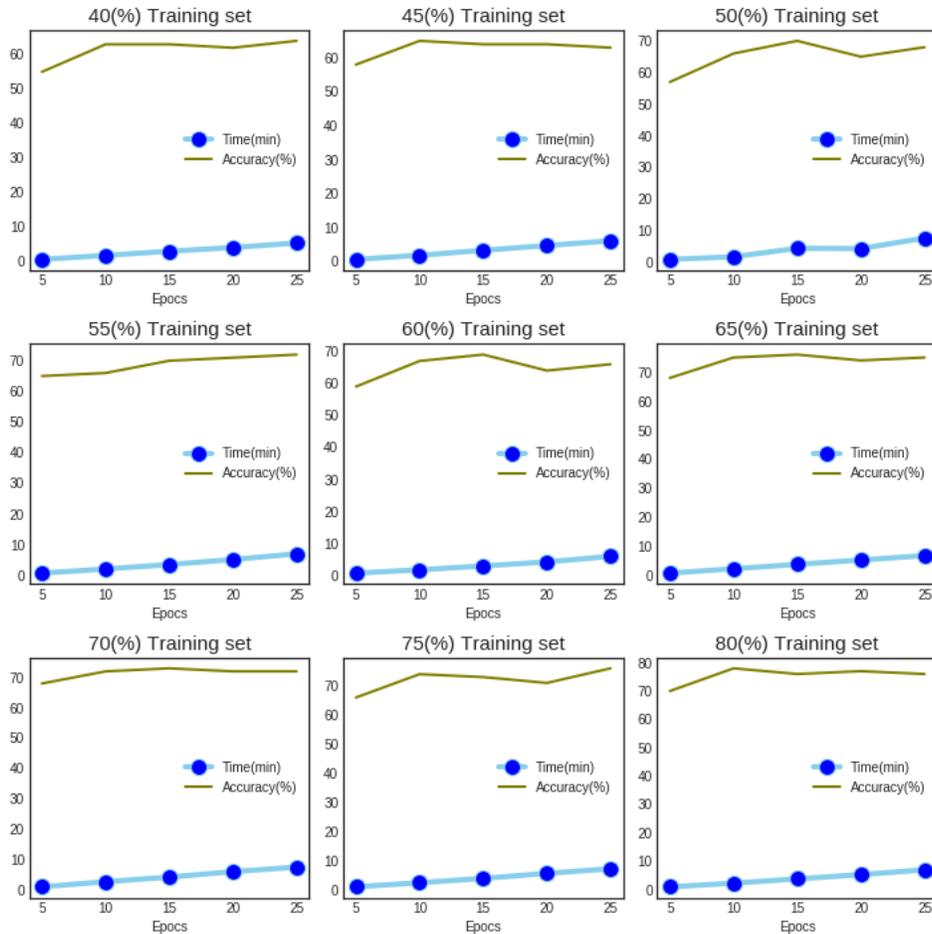


Fig. 11: NN classifier accuracy and time results for varying training and test sets.

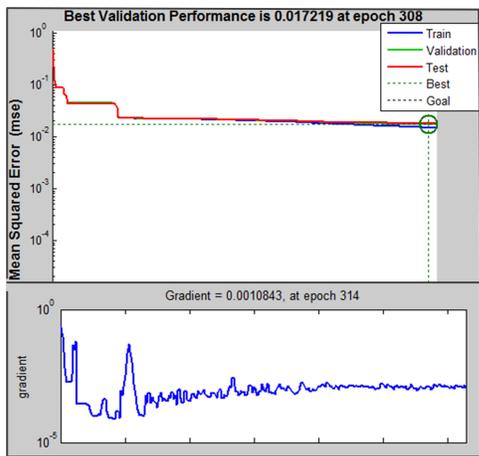


Fig. 12: MSE and gradient results of HPLR system.

REFERENCES

- [1] R. Ahmad, M. Z. Afzal, S. F. Rashid, M. Liwicki, A. Dengel, and T. Breuel, "Recognizable units in pashto language for ocr," in *Document Analysis and Recognition (ICDAR), 2015 13th International Conference on*. IEEE, Conference Proceedings, pp. 1246–1250.
- [2] R. Ahmad, M. Z. Afzal, S. F. Rashid, M. Liwicki, T. Breuel, and A. Dengel, "Kpti: Katib's pashto text imagebase and deep learning benchmark," in *Frontiers in Handwriting Recognition (ICFHR), 2016 15th International Conference on*. IEEE, Conference Proceedings, pp. 453–458.
- [3] R. Ahmad, S. Naz, M. Z. Afzal, S. H. Amin, and T. Breuel, "Robust optical recognition of cursive pashto script using scale, rotation and location invariant approach," *PLoS one*, vol. 10, no. 9, p. e0133648, 2015.
- [4] A. Abdullah, B. Agal, C. Alharthi, and D. Alrashidi, "Arabic handwriting recognition using neural network classifier," *Journal of Fundamental and Applied Sciences*, vol. 10, no. 4S, pp. 208–212, 2018.
- [5] I. Ahmad, X. Wang, Y. hao Mao, G. Liu, H. Ahmad, and R. Ullah, "Ligature based urdu nastaleeq sentence recognition using gated bi-directional long short term memory," *Cluster Computing*, pp. 1–12, 2017.
- [6] S. B. Ahmed, S. Naz, S. Swati, and M. I. Razzak, "Handwritten urdu character recognition using one-dimensional lstm classifier," *Neural Computing and Applications*, pp. 1–9, 2017.
- [7] F. Alotaibi, M. T. Abdullah, R. B. H. Abdullah, R. W. B. O. Rahmat, I. A. T. Hashem, and A. K. Sangaiah, "Optical character recognition for quranic image similarity matching," *IEEE Access*, vol. 6, pp. 554–562, 2018.

of hidden layers coupled with different feature extractor techniques to achieve a higher accuracy. Furthermore, our goal is to extend the proposed model for the connected letters.

- [8] C. Boufenar, M. Batouche, and M. Schoenauer, "An artificial immune system for offline isolated handwritten arabic character recognition," *Evolving Systems*, vol. 9, no. 1, pp. 25–41, 2018.
- [9] M. Jameel and S. Kumar, "Offline recognition of handwritten urdu characters using b spline curves: A survey," *International Journal of Computer Applications*, vol. 157, no. 1, 2017.
- [10] S. Naz, A. I. Umar, R. Ahmad, S. B. Ahmed, S. H. Shirazi, and M. I. Razzak, "Urdu nastaliq text recognition system based on multi-dimensional recurrent neural network and statistical features," *Neural Computing and Applications*, vol. 28, no. 2, pp. 219–231, 2017.
- [11] S. Naz, A. I. Umar, S. B. Ahmed, R. Ahmad, S. H. Shirazi, M. I. Razzak, and A. Zaman, "Pak. j. statist. 2018 vol. 34 (1), 47-53 statistical features extraction for character recognition using recurrent neural network," *Pak. J. Statist.*, vol. 34, no. 1, pp. 47–53, 2018.
- [12] M. Rabi, M. Amrouch, and Z. Mahani, "A survey of contextual handwritten recognition systems based hmms for cursive arabic and latin script," *International Journal of Computer Applications*, vol. 160, no. 2, 2017.
- [13] K. Rouini, K. Jayech, and M. A. Mahjoub, "Off-line arabic handwriting recognition using dynamic random forests," 2017.
- [14] A. T. Sahlol, M. Elhoseny, E. Elhariri, and A. E. Hassanien, "Arabic handwritten characters recognition system, towards improving its accuracy," in *Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), 2017 IEEE International Conference*. IEEE, Conference Proceedings, pp. 1–7.
- [15] N. A. Shaikh, Z. A. Shaikh, and G. Ali, "Segmentation of arabic text into characters for recognition," in *International Multi Topic Conference*. Springer, Conference Proceedings, pp. 11–18.
- [16] M. Kumar, M. Jindal, R. Sharma, and S. R. Jindal, "Character and numeral recognition for non-indic and indic scripts: a survey," *Artificial Intelligence Review*, pp. 1–27, 2018.
- [17] O. Zayene, S. M. Touj, J. Hennebert, R. Ingold, and N. E. B. Amara, "Multi-dimensional long short-term memory networks for artificial arabic text recognition in news video," *IET Computer Vision*, 2018.
- [18] C. Veershetty, R. Pardeshi, M. Hangarge, and C. Dhawale, *Radon and Wavelet Transforms for Handwritten Script Identification*. Springer, 2018, pp. 755–765.
- [19] P. Malviya and M. Ingle, "Feature extraction and classification techniques in character recognition systemsa comparative study," in *Proceedings of International Conference on Recent Advancement on Computer and Communication*. Springer, Conference Proceedings, pp. 527–538.
- [20] A. K. Bhunia, S. Mukherjee, A. Sain, A. Bhattacharyya, A. K. Bhunia, P. P. Roy, and U. Pal, "Indic handwritten script identification using offline-online multimodal deep network," *arXiv preprint arXiv:1802.08568*, 2018.
- [21] S. M. Obaidullah, K. Santosh, N. Das, C. Halder, and K. Roy, "Handwritten indic script identification in multi-script document images: A survey," *International Journal of Pattern Recognition and Artificial Intelligence*, 2018.
- [22] T. K. Hazra, D. P. Singh, and N. Daga, "Optical character recognition using knn on custom image dataset," in *Industrial Automation and Electromechanical Engineering Conference (IEMECON), 2017 8th Annual IEEE, Conference Proceedings*, pp. 110–114.
- [23] V. A. Naik and A. A. Desai, "Online handwritten gujarati character recognition using svm, mlp, and k-nn," in *Computing, Communication and Networking Technologies (ICCCNT), 2017 8th International Conference on*. IEEE, Conference Proceedings, pp. 1–6.
- [24] A. Selamat, I. M. I. Subroto, and C.-C. Ng, "Arabic script web page language identification using hybrid-knn method," *International Journal of Computational Intelligence and Applications*, vol. 8, no. 03, pp. 315–343, 2009.
- [25] H. Zhang, A. C. Berg, M. Maire, and J. Malik, "Svm-knn: Discriminative nearest neighbor classification for visual category recognition," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2, June 2006, pp. 2126–2136.
- [26] A. K. A. Hassan, "Arabic (indian) handwritten digits recognition using multi feature and knn classifier," *Journal of University of Babylon*, vol. 26, no. 4, pp. 10–17, 2018.
- [27] A. Mahmood and S. A. Khan, "Hard real-time task scheduling in cloud computing using an adaptive genetic algorithm," *Computers*, vol. 6, no. 2, p. 15, 2017.
- [28] M. Krachunov, M. Nisheva, and D. Vassilev, "Application of machine learning models in error and variant detection in high-variation genomics datasets," *Computers*, vol. 6, no. 4, p. 29, 2017.
- [29] A. Shahzad and N. Mebarki, "Learning dispatching rules for scheduling: A synergistic view comprising decision trees, tabu search and simulation," *Computers*, vol. 5, no. 1, p. 3, 2016.
- [30] D. Bouneffouf, "Exponentiated gradient exploration for active learning," *Computers*, vol. 5, no. 1, p. 1, 2016.
- [31] M. Jameel, S. Kumar, and A. Karim, "A review on recognition of handwritten urdu characters using neural networks," *International Journal*, vol. 8, no. 9, 2017.
- [32] X.-Y. Zhang, F. Yin, Y.-M. Zhang, C.-L. Liu, and Y. Bengio, "Drawing and recognizing chinese characters with recurrent neural network," *IEEE transactions on pattern analysis and machine intelligence*, 2017.
- [33] D. Patel, T. Som, and M. Singh, *Wavelet-Based Recognition of Handwritten Characters Using Artificial Neural Network*. IGI Global, 2017, pp. 1043–1060.
- [34] S. Nebti and A. Boukerram, "Handwritten characters recognition based on nature-inspired computing and neuro-evolution," *Applied intelligence*, vol. 38, no. 2, pp. 146–159, 2013.