Isolated Handwritten Character Recognition in Sindhi Language using Artificial Neural Network

Ali Muhammad Nizamani and Prof.Naeem Ul Hassan Janjua
SZABIST,
Karachi, Pakistan

Abstract: **- In comparison with the other Unicode-based characters languages, Sindhi language is different in terms of "Shape", "cursive style", and "position of character". These behaviours lead to present linguistic information and create difficulties in writing and printing and also create more complexities for document digitalization. The objective of this study was to (i) summarize the main characteristic of Sindhi language writing style.(ii) identify key problems in Sindhi handwritten character recognition.(iii) suggest a neural network recognition model. Sample data of Sindhi alphabets (fifty two characters) are collected from native and non-native writers and train by the system. The back propagation training mechanism is used for classification which is designed and trained to recognize any set of character. The input pattern is first fixed by image processing techniques used by a dynamic link library. Proposed model were tested with native and non-native character datasets. This study also provides comparative analysis of the learning rate of native and non-native writer's inputs and provides percentile learning differences. The model was robust and flexible which can easily be extended to different other character sets. The network exhibited recognition rate, approaching to near fair level with reasonable noise tolerance**.

Keywords**: Neural Network, Feature Extraction, Supervised learning.**

I. INTRODUCTION

The functionality of computer software is required to receive and interpret handwriting input from the source of paper, photographs or from the touch screens. Handwriting recognition is the functionality of a Software system which receives and understands handwritten input from sources such as paper documents, photographs, touch-screens and other devices. A complete handwriting recognition system also deals with formatting and it executes correct segmentation process for characters and finds the most correct results. Once individual characters are extracted, the recognition engine starts to recognize particular corresponding computer character. Handwriting recognition is ability of computer programs to convert human written text into soft copy format. This process has two ways.

1. By scanning written text
2. By taking input from input device

Most scanning suites offer some form of optical character recognition which allows to scanning in handwritten inputs and translated into digitalized form. Many archivists used character recognition procedure to convert the Literature of handwritten historical documents in digital form, this process

helps to make searchable and easily accessible documents. Physical papers are very common to use most of the time in daily life, because paper is considered as a comfortable source and very feasible medium to make an archive of date. The demand for physical documents will continue for many years to come. Hence there is a great demand for the software techniques that can automatically extract, analyze and store information from the physical documents for later retrieval. Lot of literature in Sindhi language exists in different local libraries. These documents are not really available on web to use it as worldwide resource because there is no proper method digitization is available. These data archives also hold many of handwritten documents of different great poet, authors which really need to be digitized for global promotion.

II. COMPLEXITIES IN ISOLATED SINDHI HANDWRITTEN CHARACTER RECOGNITION

Comparatively optical character recognition algorithms for handwritten text are much more complex in nature. Handwriting has several significant variations from one user to another user. Similarly printed fonts inputs are also different from each other, so the level of accuracy of hand written character recognition (HCR) is difficult to interpret by OCR software. Handwriting recognition system has several complexities which are discussed below.

A. Variability in writing Style of different writers

Different writers write same statements in different styles. This behavior creates complexity for recognition-algorithm to understand input pattern, similarly the variability of same author writes differently over the period of time [1].

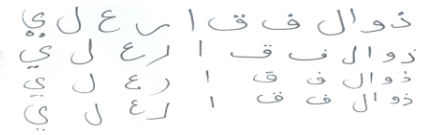

Figure 1 variability in writing styles [self]

According to figure 1, the word "ذوالفقار" is written by different writers and every writer has their own ways to sketch / write character. [1]

B. Similar geometric / topological structure of character

Characters which hold similar geometric structure create problem for recognition-algorithm. According to figure -2

Figure 2 Similar Topological Structure of character [self]

Characters geometrical shapes are similar. In this case, if input "ج" is given to input pattern and if that image contains some noise after the dot of Character then it may be possible that pattern recognition-algorithm may consider as "jhay, or chay" or any other similar geometric shape character.[1]

C. Low quality of input pattern / poorly written

Poorly written input characters can also create problems to recognition-algorithm. Suppose input pattern character

contains some shaded drop ink then its nature becomes complex to understand because it may be different from the pattern which is given at the training time of algorithm. [1]

D. Nonnative writers

Nonnative writers can't write characters smoothly as compared to tradition form. So in this case it needs to be retraining every input pattern for the non-native writers. [1]

E. Other Language Features

Sindh language text scripted cursively which means that graphemes are connected with each other. Connection of these characters interrupted at the end of the word. It is very necessary for the handwritten character recognition system to perform segmentation of characters.

F. Complex dotting structure of Sindhi Character

Dots are very important in Sindhi characters. They distinguish characters from each other and separate the signature of the character. They are extensively utilized to differentiate characters which hold similar graphemes. As shown in figure 3.

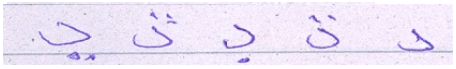

Figure 3.Complex dotting structure of Sindhi characters

According to figure 3 similar signature or geometrical shapes contain only different dots and positions of dots. Handwritten recognition algorithm can be interrupted in case of noisy inputs of above complex nature characters. For example character "Day" contain only one dot on below line where as character "Dhay" used two dots on below line. In case of above example noisy input which is not clear enough can create problem for recognition algorithm. Figure 4 represents similar characters demonstration.

G. Figure4. Different characters with different dot positions

Characters presented in figures having similar geometric shapes but they hold only differences of dots and their positions.

III. Approaches For Isolated Handwritten Character Recognition

A. LeNet Using Neural Network

The LeNet approach for Arabic character recognition contains two stages to recognize a character. In first stage it reads main body or shape of a character and in second stage it reads auxiliary information of character like dots etc. Proposed method follows this approach to recognize single character at a time. Every true pixel demonstrates a neural net known with a single "unit" or a single Neuron. Input pattern is fixed in 20 x 20 pixel field. For every input pattern, extra blank pixels are assigned around the border of input. This extra blank field helps to reduce edge effect problem during calculations. As first two layers of figure 1 shows darker shaded area which represents greater activity. This third stage of this approach finds the noise input against various dots of input character to remove the ambiguity. [2]

B. Segmentation Process

Segmentation process used to cut individual characters separately. Every segmented image then moves for classification by recognizer. The complexity of segmentation process depends on the input pattern quality and the nature of string. Characters are connected with tentative cuts. Segmentations process locates these entotic cuts / gaps between blobs of ink. As shown in figure.5

Figure 5 Tentative Cuts [2]

Entotic cuts are the block of ink which is comparatively more highlighted every time at the starting and the ending of character. The segmented results then passed to LeNet for marking score against every character. After scoring every character, back propagation learning algorithm is used to train neural network. In this proposed approach of recognition, author used multi-layer feed forward network. The training process takes input pattern. Output layer preserves output after network training process. The error rate will identify at the output layer. It calculates average difference between input and output layer pattern. To minimize the error rate weights at hidden layer are re-adjusted. Capability of learning process and the accuracy depends on the updating weights against every learning rate. [2]

C. Isolated Handwritten Characters Recognition for Gurumukhi Script

In the Gurumukhi isolated handwritten character recognition system, author completes whole process of recognition in two stages. In stage one, features are extracted which can uniquely identify the character. Similarly performance of Gurumukhi script character recognition system depends on how well the features are extracted to uniquely recognize a character. Similarly classifiers recognized the character by following the extracted features. Steps followed by Gurumukhi character recognition are discussed below [1].

D. Feature extraction method

Performance of character recognition system depends on the feature extraction method. So it is very important to have good feature extraction method to achieve high level accuracy in recognition system. [1]

E. Zoning approach

Input pattern given to Character recognition system is divided into different overlapping and non-overlapping zones. Zoning approach calculates the density of every object pixel. The zone which is denser, is calculated in terms of number of pixels. In this way every zone's pixels calculated and then sum of those pixels further divided by total number of pixels. It is considered that representation of pixel values will be either represented by 1 and 2. So every white pixel is represented with 1 and every black

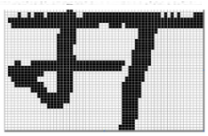

Figure6. Scaled image of Gurumukhi character saasa [1]

Pixel is represented by 0 values. Input pattern is normalized into size of 48*48 units. As shown in figure 6, an scaled

image of Gurumukhi character saasa is represented in form of matrix presentation. The process of normalization is done by dividing the all values of input pattern with largest value in the feature set. [1]

F. Classification methods

Classification process takes extracted features as an input to recognize text segment. So a training data model will help here to recognize to give unknown pattern. Different input patterns of different authors are given to produce training data. [1]

G. K – Nearest Neighbour

K nearest neighbour approached helps to compute a classification function. It examines the labelled training nodes or points in the n dimensional space. Here n represents the size of feature. K nearest neighbour algorithm calculates Euclidean distance between the given test points and rank the obtained distance [1].

H. Support Vector Machine

Proposed approach of Gurumukhi isolated handwriting character recognition used support vector machine approach which is very useful technique to classify input data. It is a learning machine which has very good generalization features. Proposed approach comes up with database of image files of handwritten Gurumukhi scripts which contain size and features of characters of Gurumukhi language. In order to store image data extensible mark-up language is used. This approach collects a total 2050 images of Gurumukhi characters. According to summarized results, there are some failure reasons of proposed approach which has different causes including

- Inputs which are not clear enough (not properly written)
- Characters which are holding very similar topological structure.
- Low quality blurred and broken images [1]

I. Primary and Secondary stroke approach for Isolated Urdu Characters

This approach follows primary and secondary stroke approach to identify the handwritten Urdu isolated characters which is drawn on the drawing panel of a PC. To achieve the goal of Urdu handwritten characters, this approaches used to analyze visual, geometrics and other auxiliary parts of characters and to identify the features of every characters which distinguishes every character. Similarly at the classification stage, it classifies every input as a separate identified character [3].

J. Primary Stroke Features

Primary stroke utilizes to train the given input pattern. It collected the following list of features from given input.

- Length and angle of bounded box (Actual Panel)
- Distance between the two points
- The result of Sum of squared of the angles.

Collect sine and cosine angles between first and last point [3]. Figure7 represents primary and secondary strokes of the characters.

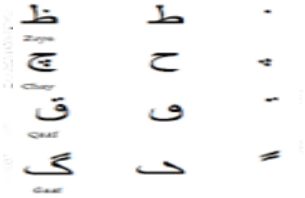

Fig-7 Characters divided into primary and secondary stroke [3]

K. Secondary Stroke Features

Secondary stroke features identify the differences between characters by holding auxiliary parts of input. The secondary stroke features are written below.

- Length and number of secondary strokes
- Position and number of dots in secondary stroke

Number of dots helps to bifurcate one character with another for example character “Bay” contains very similar geometric shapes but it differs with others by total number to dots. The position of secondary stroke is very important feature which make difference between characters [3].

L. Classification Process

For the classification of the patterns, author used weighted linear classifier approach to perform the process of recognition. This study was focused to recognize handwritten Urdu characters. The collection of samples was done from native and nonnative writers. According to results, discrepancy appears due to the cultural differences in the procedure in which characters are traditionally written. Two sample characters are collected from two different writers and tested in linear classifier which correctly classifies 92.8% of character. Similarly proposed approach also used by collecting input patterns from nonnative writers who had never seen Urdu characters. In this case 31% characters correctly identified successfully. This is because of poorly written input patterns [3].

M. Farsi/Arabic Isolated Character Recognition – Chain Code approach

Chain code algorithm includes number of locations of dots and other auxiliary parts in combination, which are used to identify the isolated character recognition. This approach is used to extract features including number of holes, number of locations. The process of chain code approach recognition is explained below.

N. Preprocessing

Preprocessing stage used to adjust input pattern to the single fixed size to one point. So that it can be used directly and efficiently by the feature extraction components. Figure 8 represents the details of image which follows thinning procedure.

Figure – 8 Farsi Character after thinning process [4]

The number of auxiliary parts is extracted by chain code approach as an identification of each character.

O. Auxiliary Parts Identification

Auxiliary parts of character are considered as core objects. An order triple is developed against any character. The procedure for creating the order triple is to read the character from top of the input image and read it row by row and assign P1 to the first object in the image. P2 and P3 are also allocated

to objects located in the lower parts of the character image. Every part of input character Integrated with 0 and 3. If the intended object is the main body of the character, the value 3 will be assigned to relevant *P*.

P. Analysis of holes in character

After identification of auxiliary parts, number is to assign to every character. In Arabic and Persian languages, script a character that has maximum one hole. For example the characters 'be', 're', 'dal' have got no holes, whilst the characters 'sad', 'fe', 'ghein' have got one hole. Several methods are available to calculate the number of holes from a particular character. *F* = [*P*1, *P*2, *P*3, *Holes*]. Using the feature vector F, the total set of Farsi characters could be separated to ten non-overlapping subsets which are shown in Table I along with their computed feature vectors.

Q. Chain Code Feature

Chain code approach focuses on the main body of character of input image. It reads from top of the input image and move row by row to collect pixel information. Character which contains no starting point will be assigned zero number. For example character "Hay" doesn't contain starting point. So each pixel of image has got eight neighbors and to each neighbor assigns one value between 1 and 8. As shown in figure 9.

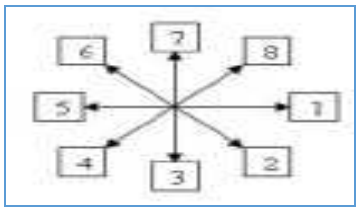

Figure -9 Neighbours of a pixel and value assign to it [4]

So Chain code computation follow following equation vector *F* = [*P*1, *P*2, *P*3, *Holes*,*Chain _Code*]

R. Chain Code Normalization

Chain code numbers are transferred into two dimensional matrixes. Matrix first row is original values of given chain code. Second row values are considered as frequency of occurrence of these values. Proposed approached followed nearest neighbor classifier in order to decrease the role of the classifier by keeping the importance of feature extraction. Metlab tool was used to implement this approach to perform and test different datasets. The average result of propose solution accuracy is near 97.4%.

IV. Propose Approach For Sindhi Isolated Handwritten Character Recognition

For constructing a neural network for any of problem, it is very necessary to select correct type of network with suitable number of hidden inputs. These parameters really depend on the nature of problem. Regarding Sindhi Isolated character recognition large number of input and output data is available which needs to be supervised. Time is the only factor which affects training process. Proposed neural network model uses feed forward back propagation neural network approach to recognize isolated Sindhi handwritten character. Back propagation neural network is very fair approach in following different cases. In case when large number of data (input and out) is available but not assured to compute the right output. In such cases back propagation is feasible solution to match inputs with outputs. Same scenario matches with Sindhi isolated handwritten character recognition where large input patterns of different native and nonnative writers are available and prescribed input is matched with proposed fifty two Sindhi alphabets. Proposed neural network model of Sindhi isolated handwritten character recognition is consisting on three layers. These layers are listed below.

1. Input Layer
2. Hidden Layer
3. Output Layer

Input Layer is the first layer of neural network model which takes the input from external surroundings. In case of Sindhi isolated handwritten character recognition, it takes input picture box of the application. User needs to give input using software pen input option by writing character on drawing box. Every neuron of input layer is also connected with all neurons of the next layer i.e. known as hidden layer. As input layer is connected with hidden layer, hidden layer passes information to the output layer and offer results. In order to simulate neural network like a human brain, values and different other parameters are propagated and suitably adjusted. Adjustment of these parameters is known as neural network training. Propose solution of neural network model follow supervise training approach. A single layer of neural network consists of more than one neuron. Every synapse which connects the nodes has some weight. Transformation of data from one neuron to the neuron calculated synapses weights by multiplying to its weights. So the synapse weight has greater role to perform correct functionality of any neural network.

To make a proof of concept neuron dot net library is used for image extraction process. Proposed approach of back propagation algorithm used supervised learning approach to train feed forward back propagation neural networks. Multiplication of weight and the source of neuron are passed as input neuron and out activation results are known as output neurons. The back propagation training approach initializes the weights at first stage and then selects different random samples for the training process. Once input neurons are assigning to the input vector, proposed method starts propagation to all neurons and push them in forward direction in order to get out from the output layer. In output layer, proposed method check error results and check the difference of output and required output. Here checking process has squared error value. Square error value denotes the efficiency of training process. At this stage delta is calculated which explain updated weights of all synapses. This updating performs variation in error retrieval process. The cycle of training epoch for all training sets perform up to reach on desired error stage. This process takes much iteration. The criterion of stopping a particular iteration is to reach on acceptable mean squared value.

A. Collection of Data from Native and none Native writers

For the training process of proposed solution, data is collected from both native and none native writers. The details of data collection are presented below.

TABLE1. DATA COLLECTION DETAILS

Writers	Patterns	Total Alphabets
Native	5	52
None Native	5	52
Total Number of patters		52* 10 = 520

To preserve the collected samples, they are serialized further in a file system. Complex structure characters like “ghay” , “jhay” and few other are really difficult to write for nonnative writers so more than one attempts of characters are saved in data dictionary. Excluding these five hundred and fifty two characters, few other characters are already entered previously before entering final data sample in the system.

B. Training Process

For the determination of good effective training process, it needs to have more training samples. Training process provides training set and adjustment of weights of network synapses. The Collection of Training set in proposed solution is Training Set = “Set of training Samples of Sindhi Isolated handwritten Characters of different authors”. So the training sample is considered as input vector and desired output vector in case of supervised learning method. Here the length of input vector is similar in size to input layer as the length of the output neuron in output layer. Supervised training process creates three layers. Input layer containing four hundred neurons. Hidden layer contains four neurons and output layer contains two neurons as shown in following figure. These layers are created to construct the network topology. To make an effective learning process for Sindhi isolated handwritten characters, four hidden layers are used. These neurons are actually desired input and output for proposed solution. Linear layer is used because user input is not further modified by adding any biased or activated function. According to figure10 the properties of any layer can be modified at any step.

```
ActivationLayer inputLayer = new LinearLayer(400);
ActivationLayer hiddenLayer = new SigmoidLayer(4);
ActivationLayer outputLayer = new SigmoidLayer(2);
new BackpropagationConnector(inputLayer, hiddenLayer);
new BackpropagationConnector(hiddenLayer, outputLayer);
```

Figure10. Neural network Layers Construction [self]

Training process needs training data to train a network. Here in the proposed solution training data is collected from native and nonnative writers. Five native and five nonnative writers comprised of fifty two characters in five different samples sets. Total five hundred twenty input patterns were collected to train. Training process iteration of back propagation network is stored and retrieved.

```
IFormatter formatter = new BinaryFormatter();
INetwork network = (INetwork)formatter.Deserialize(stream);
```

Figure11. Dieselization of Stream [self]

The above stanza of code is used in the application and will return the Back propagation Network object stored at the given path.

```
TrainingSet trainingSet = new TrainingSet(400, 2);
Alphabet ithLetter = Alphabet.GetLetter(i);
Alphabet jthLetter = Alphabet.GetLetter(j);
trainingSet.Add(new TrainingSample(instance.
GetEquivalentVector(20, 20), new double[] { 1d, 0d }));
trainingSet.Add(new TrainingSample(instance.
GetEquivalentVector(20, 20), new double[] { 0d, 1d }));
```

Figure12. Construction of Training set [self]

A class name “iserializable” is used to preserve set of training patterns into file system.

C. Learning Rate

Learning rate parameter is used in proposed solution to regulate the neural network learning process. It accelerated and can make slow down the learning process of neural network. The learning of network topology is associated with all three layers. It can be customized for each layer. The method known as “Set Learning Rate” is used to adjust learning rate of the neural network as shown in figure 13.

```
network.SetLearningRate(0.3)
```

Figure13. Change of learning Rate

The repetition process of training cycle at large number of times enables network to ultimately learn different samples very effectively. The higher learning rate (rate near to zero) is directly proportional to the large number of training epochs and due to large number of training epochs, it makes extreme slow progress during training process. So it is the only one disadvantage of back propagation algorithm that it makes slower specially in high learning rate which is very necessary for Sindhi Isolated handwritten characters because of similar geometric, calligraphic glyphs. Back propagation approach of Sindhi isolated character recognition used activation function to collect outputs from input patterns. The default linear sigmoid derivatives are computed.

D. Learning comparison

Training data were collected from two different types of writers. One of them was nonnative writer who did not know about Sindhi alphabets. Input pattern given by nonnative writers are more different in shape and size then an input given by a native writer of Sindhi language. Below is the graph which shows the average difference of learning time of two different types of writer’s inputs. Figure 14 represents the average time interval of learning process of same characters by two different writers. Vertical legends are representing time interval taking by system for the recognition and the horizontal legends are representing the input sources. According to observations, the variation in learning time appears due to different writing styles of writers, nonnative writer who cannot exactly write the characters.

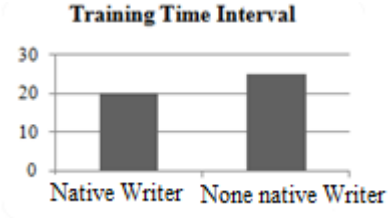

Figure14 –Interval between native and none native writer.

E. Experimental Results

Recognition process first gets input patterns from drawing panel. All numbers of recognized alphabets are loaded and iterated. Input patterns are further run by network method and provide output. According to table 2, for the testing of proposed solution the input data was collected from four different writers.

TABLE2. RESULTS OF NATIVE AND NONE NATIVE WRITERS

Native Writer	1	Correct Results	%
Number of inputs	52	48	92.30%

Native writer	2		
Total inputs	52	47	90.38%
None Native writer	3		
Total inputs	52	43	82.69%
None Native writer	4		
Total inputs	52	40	76.92%

Native writers were given total one hundred four input patterns for the recognition. Ninety five of them were successfully recognized by the model. Only nine results were in failure conditions. Similarly same number of inputs were given by two other nonnative writers. Total eighty three patterns were successfully recognized from one hundred four input patterns. The average accuracy on pattern recognition is 91% for the native writers whereas nonnative writers the average accuracy result is 79%. Similarly over accuracy performance of the propose model is 85.75% which is shown in table 3.

TABLE 3. AVERAGE ACCURACY RATE OF PROPOSE MODEL

Average Accuracy Results	
Writer 1	92.30
Writer 2	90.38
Writer 3	82.68
Writer 4	76.92
Average	342.29 /400 = 85.75%

V. CONCLUSION

Many sources of Sindhi literature including libraries, archives, museums and cultural institution are really aware of the requirement of preserve documents in digitalization form. In this study, I have complete two main objectives. One was to identify the complexities of Sindhi Handwritten recognition and second was to present an experimental work of back propagation neural network for recognizing Sindhi isolated handwritten characters. Isolated handwritten character recognition will become basic root for having complete Sindhi handwritten recognition in future. To evaluate back propagation approach inputs were collected from native and nonnative writers. Achieved results suggested that proposed approach can be more useful and accurate by increasing the number of sample sets. The proof of concept of this paper provides fifty two characters isolated character recognition for Sindhi. The training process gives small time difference in training process for the native and nonnative writer's inputs. Proposed approach is adoptable for any application which requires Sindhi isolated handwritten character recognition apart from its computation power including training time set.

The ratio of complexity to maintain the accuracy in recognition of id, depends on the form of input patterns which is required to be read. Proposed approach has problems while passing some noisy input patterns, which needs to factorize further with some good image processing techniques. As proposed model used for pattern recognition which takes character as an input, this solution can be used for words and phrases recognition by including a proper segmentation algorithm. According to experimental results the average accuracy rate of the native writes is about 91%. Similarly the average results of nonnative writers are about 79%. Total average of accuracy for both writers is 87.75%. These figures really fall in satisfactory criteria because both types of users who enter the data for the supervision and for the recognition were not used to write Sindhi language on a software system.

REFERENCES

[1] Dharamveer Sharma, Puneet Jhajj, "Recognition of Isolated Handwritten Characters in Gurumukhi Script", International Journal of Computer Applications, Volume 4-8 **, August 2010.**

[2] Rashad Al-jawfi, "Handwriting Arabic Character Recognition LeNet using Neural Network", Department of Mathematics and Computer Science, Ibb University , Yeman , **Feb 2008.**

[3] Nabeel Shahzad, Brandon Paulson, Tracy Hammond "Urdu Qaeda: Recognition System for Isolated Urdu Characters", IUI Workshop, **Feb 8 ,2009.**

[4] H.Izakian, S.A.Monadijemi, B.Tork Ladani, and K.zamanifar, "Mult-Font Farsi / Arabic Isolated Character Recognition Using Chain Codes", World Academy of Science, Engineering and Technology 43, **2008.**

[5] Stuti Asthana, Farha Haneef , Rakesh K Bhujade , "Handwritten Multiscript Numeral Recognition Using Artificial Neural Networks", Volume -1, Issue-1, **March 2011.**

[6] Amit Choudhary, Rahul Rishi,Savita Ahlawat,Vijapal Singh,"Totally Unconstrained Handwritten Character Recognition using Improved BP Neural Network", Vol.2 , **2010**.

[7] Sandhya Arora, Debotosh Bhattacharjee, Mita Nasipuri,L.Malik,M.kundu,D.K.Basu, "Performance Comparison of SVM and ANN for hand written Devnagari Character Recognition , IJCSI, Vol.7, May **2010.**

[8] Accredited, World class language solution website. Last visit:18-10-2011. (http://www.alsintl.com/resources/languages/Sindhi/)

[9] Dharamveer Sharma, Ubeeka jain, "Recognition of Isolated handwritten Characters of Gurumukhi Script Using Neocogniton", International Journal of computer Applications, Volume 10, Nov 8, **2010.**

[10] Dewi Nasien, Siti S.Yuhaniz, Habibollah Haron, "Recognition of Isolated Handwritten Latin Characters using One Continuous Route of Freeman Chain Code Representation and feed forward Neural network", WASET,67, **2010.**

[11] Hamza, Ali A, "Back propagation Neural Network Arabic Characters classification Module utilizing Microsoft Word", Journal of Computer Science 4 (0) 774-751, ISSN 1549-3636,**2008.**

[12] Vikas J Dongre, Vijay H Mankar , " A Review of Research on Devnagari character Recognition ", International Journal of computer Applications Volume 12, No.2, **Nov 2010.**

[13] Cheng-L in Liu, Fei Yin, Da-han Wang, Qiu-feng wang, "Chinese Handwriting Recognition Contest 2010", NLPR , Beijing, **2010.**

[14] Sobia T.Javed, Sarmad Hussain, Ameera Maqbool , Samia Asloob, Sehirish Jamil , Huma Moin,"Segmentation free Nastalique Urdu OCR" , WASET 70, **2010.**

[15] Sindhi Language Authority, Official Website www.sindhila.org/Reserach%20journal.htm **(Accessed 23 -08-2011).**

[16] Sindhi Language, Website http://www.sindhi language.com/script.html **(Accessed 23 -08-2011)**

[17] Alireza Alaei, P. Nagabhushan, Umapada Pal, "Fine Classification of Unconstrained Handwritten Persian/Arabic Numerals by Removing Confusion amongst Similar Classes", 10th International Conference on Document Analysis and Recognition ,**2009.**

[18] Gheith A. Abandah, Khaled S. Younis "Handwritten Arabic Character Recognition using Multiple classifier based Latter Form " , International Conference on signal processing ,pattern recognition and applications **Feb 2008.**

[19] Dharamveer Sharma, Deepika Gupta, "Isolated Handwritten Digit Recognition using Adaptive Unsupervised Incremental Learning Technique", International Journal of Computer Applications (0975 – 8887) Volume 7– No.4, **September 2010.**

[20] Reza Ebrahinpour, Mona Amini, Fateme Sharifizadehi, "Farsi Handwritten Recognition Using Combining Neural Networks Based on Stacked Generalization", International Journal on Electrical Engineering and Informatics - Volume 3, Number 2, **2011**