

Angular Symmetric Axis Constellation Model for Off-line Odia Handwritten Characters Recognition

Pyari Mohan Jena¹, Soumya Ranjan Nayak²

¹Departement of Computer Science and Engineering, College of Engineering and Technology, Bhubaneswar, India

²Departement of Computer Science and Engineering, K L University, Andhra Pradesh, India

Article Info

Article history:

Received Oct 12, 2017

Revised Feb 20, 2018

Accepted Mar 26, 2018

Keyword:

Feature Vector

Optical Character Recognition

Pattern Recognition

Random Forest

SVM

ABSTRACT

Optical character recognition is one of the emerging research topics in the field of image processing, and it has extensive area of application in pattern recognition. Odia handwritten script is the most research concern area because it has eldest and most likable language in the state of odisha, India. Odia character is a usually handwritten, which was generally occupied by scanner into machine readable form. In this regard several recognition technique have been evolved for variance kind of languages but writing pattern of odia character is just like as curve appearance; Hence it is more difficult for recognition. In this article we have presented the novel approach for Odia character recognition based on the different angle based symmetric axis feature extraction technique which gives high accuracy of recognition pattern. This empirical model generates a unique angle based boundary points on every skeletonised character images. These points are interconnected with each other in order to extract row and column symmetry axis. We extracted feature matrix having mean distance of row, mean angle of row, mean distance of column and mean angle of column from centre of the image to midpoint of the symmetric axis respectively. The system uses a 10 fold validation to the random forest (RF) classifier and SVM for feature matrix. We have considered the standard database on 200 images having each of 47 Odia character and 10 Odia numeric for simulation. As we have noted outcome of simulation of SVM and RF yields 96.3% and 98.2% accuracy rate on NIT Rourkela Odia character database and 88.9% and 93.6% from ISI Kolkata Odia numerical database.

*Copyright © 2018 Institute of Advanced Engineering and Science.
All rights reserved.*

Corresponding Author:

Soumya Ranjan Nayak,

Departement of Computer Science and Engineering, K L University,

Greenfields, Vaddeswaram, Guntur, Andhra Pradesh, 522502-India.

Email: nayak.soumya@kluniversity.in

1. INTRODUCTION

In the era of digital image processing, the character recognition is one of the significant and useful emerging research topics is the area of pattern recognition. The main intend of character recognition is to translate human readable character to machine readable code so that machine can efficiently recognize the character. There are mainly two broad category of character recognition system are found such as offline and online recognition process. In case of online character recognition process, it represents the two dimensional co-ordinates of successive points of the handwriting as a function of time are stored in particular order described by [1]. where as in case of the offline handwriting, only the completed writing is available as an image describe by [2]. In this paper, our research intend confined with offline handwritten character recognition. Our recognition stage comprises of three broad stages including acquisition, feature extraction and classification step. Beside that a recognition system mostly depends upon a well-defined feature extraction procedure along with a good classifier, in order to achieve high success rate [3]. In order to achieve a good recognition system for handwritten format is quite still challenging because of variation in

writing skills, shapes and orientation. Various approaches are followed up by different researcher to various scripts like Arabic, Chinese, and English etc are reported [4]. Basically Odia Script language is one the language which is derived from Devangiri scripts. It is one of the regional languages of India, mostly spoken at eastern part (Odisha) and some south, north part of India. To achieve a good accuracy of recognition for handwritten characters of Odia character is quite impressive. Though a good number of works has done for Indian regional languages but a less in number related to Odia script. In these past recent years different authors make an attempt for analysis with respect to Odia scripts are reported in [5]. The feature extraction technique for recognition of handwritten character is a challenging task in the research field of pattern recognition. In this regard a large number of feature extraction technique and classification algorithm have been presented in recent year described by [6]. Several character recognition technique of different language is found in many literatures [7-9]. In line to character recognition the extensive survey has been reported based on different kinds of feature extraction technique [10]. In this survey paper, author reported different feature extraction technique applied on Template matching, Projection histograms, Deformable templates, Contour profiles, Unitary image transforms, Zoning, Graph description, Zernike moments, Spline curve approximation and Fourier descriptors has been applied on gray level character, Binary character, Character Contour, character skeleton and character graph image representation form in the pre processing steps. As the Indian language is concerned, the optical character recognition plays a vital role now days. In this paper we have made an attempt to design a novel approach that efficiently recognize the odia character by implementing angular measurement and Euclidian distance by taking the midpoint from the axis, which was generated by taking the midpoint of two boundary edge of row symmetric axis as well as column symmetric axis to the centre of the images. Odisha state, so far has been able to uphold the pride of having the largest number of palm leaf manuscripts (over 20,000 manuscripts) in the world. [11]. Million books would have been printed from starting where “New Testament” that got printed in 1809 was first published.[12]. Odia got classical status except 5 other Indian languages on the basis of its literary heritage following approval of the Union cabinet.

2. RELATED BACKGROUND WORK

Odia script has been extracted using Bhrami scripts and one of the most ancient languages among Indian regional language most spoken eastern part of India basically in state Odisha, West-Bengal, Gujarat etc. The most important scenario of this language that it has no lower and upper case format. Here in the script is has no upper case lower structure. A certain well-defined approaches are adopted by different researchers to achieve high recognition rate. Recognition is the process of accepting the unknown samples of handwritten character image or words and then proceeds into a pattern recognition problem for testing. Recognition process can be achieved either in three important way, which is described as template matching, statistical technique and neural network techniques. These character recognition approaches uses either top down approaches or analytical strategies for recognition. Template matching is the simplest form of training and recognition. Here is the idea is to match the stored predefined prototype with the unknown handwritten characters. In this matching technique only selected pixel are compared with data samples and ruled based decision tree analysis. Rule based decision technique were used by chaudhuri et. at in 2002 [13]. Statistical technique considered as more effective while recognition of Odia characters. In this regard obaidullah et al [14] in 2014 uses the linear logistic regression model by using higher order statistical decision model to provide better performance rather than the linear model in performance. In 2007 pal et al [15] used quadratic function for classification is based on Bayesian estimation. In 2009 and 2005 a similar techniques of pseudo Bayesian estimation technique was adopted by waxabyashi et al [15], and roy et al [16] for odia handwritten numerical recognition. They used conventional quadratic discriminant function. In 2006 Hidden Markov Model (HMM) was purposed by Bhowmik et al [17]. This is used non homogeneous quadratic method for training and recognition of handwritten numerical. In 2014 Dash et al [18]-[19] have adopted a discriminative learning based quadratic discriminant classifier (DLQDF) and Non-redundant Stockwell transform based feature extraction for handwritten digit recognition. Neural network is the parallel processing method having interconnection of neurons inside this technique. It perform computation at higher speed in comparison with statistical and template matching. Neural network can be performed either in two ways like feed forward network (FFNA) and back propagation network (BPNN). In 2013 mishra et al [20] perform the classification with BPNN and got a high accuracy of 90.44 percentage. In 2011 Majhi et al [21] authors have proposed a nonlinear neural network classifier it is an analogy of functional link artificial neural network (FLANN) classifier. In 2012 Chanda et al [22] propose a method for writer identification from Odia handwritings which uses the SVM for classification. In 2015 kalyan et al [23] purposed BESAC symmetric axis constellation model using classifier SVM, Nearest neighbour and Random forest having accuracy 98.90, 99.48, 96.76 percentage respectfully. The details comparison of recognition accuracies are described in below Table. 1.

Table 1. List of All Features in Recognition of Odia Characters with Accuracy

Method	Features	Classifier	Database	Accuracy (%)
Pal et al. (2007a)	Gradient + curvature	MQDF	Odia basic characters on IITBBS	94.60
Wakabayashi et al. (2009)	Weighted gradient	MQDF	Odia basic characters on IITBBS	95.14
Roy et al. (2005)	Directional	Quadratic	Odia numerals on IITBBS	94.12
Bhowmik et al. (2006)	Scalar	HMM	Odia numerals on ISI Kolkata	90.50
Bhowmik et al. (2006)	Scalar	HMM	Odia numerals on IITBBS	91.26
Mishra et al. (2013)	DCT	BP NN	Odia numerals on ISI Kolkata	92.00
Mishra et al. (2013)	DCT	BP NN	Odia numerals on IITBBS	90.44
Dash et al. (2014b)	Hybrid topology	DLQDF	Odia numerals on ISI Kolkata	98.50
Dash et al. (2014a)	Stockwell transform	k-Nearest neighbor	Odia numerals on ISI Kolkata	98.80
Dash et al. (2014b)	Hybrid topology	DLQDF	Odia numerals on IITBBS	98.28
Dash et al. (2014a)	Stockwell transform	k-NN	Odia numerals on IITBBS	99.00
Kalyan S Dash et al. (2016)	BESAC	Random forest	Odia numerals on ISI Kolkata	98.44
		SVM	Kolkata	99.02
		Nearest neighbor		99.35
Kalyan S Dash et al. (2016)	BESAC	Random forest	Odia numerals on IITBBS	97.30
		SVM		98.56
		Nearest neighbor		98.90

3. PROPOSED HANDWRITTEN CHARACTER RECOGNITION SYSTEM

In this section, we have made a novel technique that efficiently recognizes the Odia character. The complete proposed method is described graphically in Figure 1. These proposed systems are carried out by including the certain steps like Image like Image Acquisition, Pre-processing, Feature Extraction, and Classification. The details discussion can be made in several sub-chapters in subsequent section.

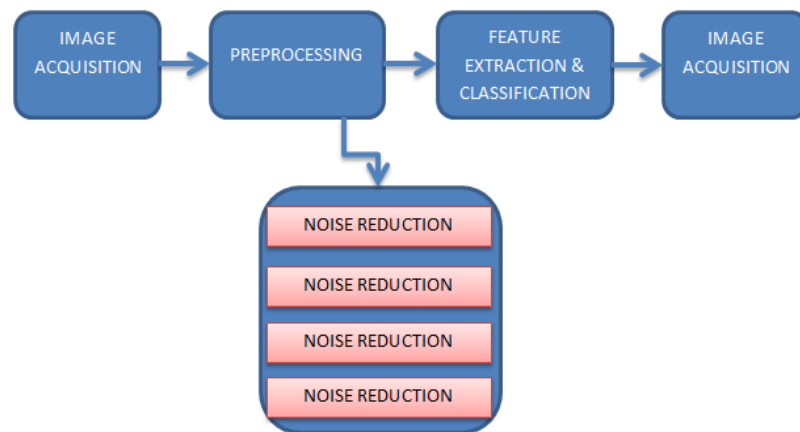


Figure 1. Schematic Model of Recognition Model

3.1. Image Acquisition

As per our proposed methodology described above we have consider the standard database of odiya character named as Nit Rourkela Odia database, which was developed at NIT, Rourkela by Mishra et al. [20]. In this database they had composed of various 15040 numbers of images of both character and numerals. In this research analysis, we have considered 47 characters having 200 numbers of samples for our experimental study. The modern Odia script consists of 12 vowels, 3 vowel modifiers, 37 simple consonants, 10 numerical digits and about 159 composite characters (juktas). Odia script is a curved appearance of writing patterns on

Palm leaves which have been secure from tear if writer uses too many straight lines. Table. 2 describes about Odia characters with their Phonetics.

Table 2. List of All Features in Recognition of Odia Characters with Accuracy

Phonetics	Letter	Phonetics	letter	Phonetics	Letter	Phonetics	letter	Phonetics	letter
a		Ka		Da		Ma		Ek	
Aa		Kha		Dha		Ja		Due	
E		Ga		Na		Ra		Tini	
Ee		Gha		Ta		La		Chari	
U		Una		Tha		Laa		Paanch	
Uu		Cha		Da		Ya		Chha	
Ru		Chha		Dha		Sha		Saath	
Ruu		Ja		Na		Sha		Aath	
A		Jha		Pa		Sa		Na	
Ae		Nya		Pha		Ha		Shun	
O		Ta		Ba		Khya			
Ou		Tha		Bha					

3.1. Image Pre Processing

Pre-processing is an important step during the image acquisition process in order to get higher accuracy result by means producing noise free images as well as free of skewnes. In this analysis step, our pre-processing steps are done by using different phases like noise reduction, Normalization, skew or slant adjustment and segmentation. The details description of these pre-processing steps are summarised in the following sub sections.

3.2. Noise Reduction

Noise is the unwanted output comes with the pixel intensity value in the scanned document whereas reduction of noise is the process of eliminating spurious points due to the poor sampling rate of the scanner.

3.3. Normalization

Normalization is the process of separating what data we get and what data we required. We adopt binarization as the intensity normalization in the pre-processing step. Then we adjust the size of each sample as 81*81 dimensions for size normalization.

3.4. Skew or Slant Adjustment

Skewness in the image undergoes some rotation of scanned image. This is very important to eliminate rotation in the pre-processing step. Rotation can be eliminated by implementing the elimination of degree of tilt angle and rotation of opposite direction.

3.5. Segmentation

Segmentation is the process of separation of text and non text area in the scanned handwritten document. It is the challenging part for pre-processing there are 2 types of segmentations can have in the pre-processing steps, External segmentation perform separation of paragraph, words or sentence from scanned documents whereas internal segmentation is the process of separation character from each word.

3.6. Feature Extraction

Feature extraction techniques are used to evaluate the uniqueness of each character image by which they differs from the rest character images. In this section we have implemented a unique algorithm for evaluation of feature vector by considering the mean distance of row, mean angle of row, mean distance of column and mean angle of column from centre of the image to midpoint of the symmetric axis respectively. All the operations were performed over skeletonized image of handwritten characters. Our feature extraction

implementation is mainly focuses on five unique steps, and this is considered as the key feature values of our proposed system. The details of the five steps are described as follows:

1. This unique technique extracted the feature of image according to binary image point passing through the angle of 45° , 135° , 180° , 225° , 270° , 315° , 360° respectively from the centre of the image represented in Figure 3.
2. Extract the point positions of the sample image having passes through these angles represented in Figure 4.
3. Plotting all the points create a unique polygon shaped image for each sample represented in Figure 5.
4. Find row symmetry axis and column symmetry axis based on Figure 6 & 9 respectively.
5. Estimate mean angle and distance of both the symmetry axis based on Figure 8 & 11 respectively.

For the above all empirical calculation of our implementation methodology, we have developed two algorithms which were depicted in Algorithm I and Algorithm II respectively.

Algorithm 1 row and column symmetry axis

Input: Extract the boundary points $\{N\}$ of character image

Output: Extract row Symmetry axes ($RE_i, i=1, 2, 3 \dots j$)

Extract column Symmetry axes ($CE_i, i=1, 2, 3 \dots K$)

```

Begin
  For i=1 to N do
    While row chords creation is satisfied
      Draw chords from each boundary point to other boundary point in same row
      Map the row chords accordingly so that sets of parallel row chords are grouped.
    End while
    While column chords creation is satisfied
      Draw chords from each boundary point to other boundary point in same column
      Map the column chords accordingly so that sets of vertical column chords are grouped.
    End while
  End for
  Find midpoint parallel row chords
  Resulting axes consist of midpoints of row external chords
  Provide potent row symmetry axes ( $RE_i$ )
  Find midpoint vertical column chords
  Resulting axes consist of midpoints of column external chords
  Provide potent column symmetry axes ( $CE_i$ )
  Return  $RE_i, CE_i$ 

```

Algorithm 2 feature extraction

Input: Extract row Symmetry axis ($RE_i, i=1, 2, 3 \dots j$)

Extract column Symmetry axis ($CE_i, i=1, 2, 3 \dots k$)

Output: mean angle and mean distance of row and column symmetry axis feature (f)

```

Begin
  For i=1 to j do
    Find the midpoints of row symmetry axis
    Compute angle from centre of image to midpoints of row symmetry axis
    Compute distance from centre of image to boundaries of row symmetry axis
  End for
  For i=1 to k do
    Find the midpoints of column symmetry axis
    Compute angle from centre of image to midpoints of column symmetry axis
    Compute distance from centre of image to boundaries of column symmetry axis
  End for
  Compute mean angle from centre of image to midpoints of row symmetry axis
  Compute mean distance from centre of image to boundaries of row symmetry axis
  Compute mean angle from centre of image to midpoints of column symmetry axis
  Compute mean distance from centre of image to boundaries of column symmetry axis
Return f

```

The proposed character recognition method are divided the images into two parts of operation and the first part operation included a chord that is drawn from each boundary pixel to straight of boundary pixel in row wise and the second part consisting a chord that is drawn from each boundary to straight of its boundary pixels in column wise and the complete description of these two steps are discussed in Algorithm 1 and Algorithm 2 respectively. For N no of boundary pixel and K no of boundaries; the number of available cord is $(N/2)*k$ in row wise and column wise. However, we discard those boundary chords which having less than 3 pixels in that cords. The remaining cord is called row chords and column chords because these chords are present in the same row and same column. These chords are parallel present in row chords, which is

presented in Figure 6 and the cord are vertically present in column chords presented in Figure 9 respectively. In our subsequent step we have group the row chords and column chords, in order to find symmetry axis from parallel row chords and vertically column chords. The midpoint of the parallel row chords and vertical column chords could generate a number of row symmetry axes as well as column symmetry axes which are presented in Figure 7 and Figure 10 respectively. In order to find the accurate symmetry axes to represent the perceptual parts, we propose midpoint criteria of the respective chords to be verified in the following method.

$$\text{Midpoint of row chords} = \frac{Y_r - X_r}{2}$$

Where $r=(1,2,3,\dots,n)$ No of boundaries
 $Y_r = Y$ point of r 'th row boundary point
 $X_r = X$ point of r 'th row boundary point

$$\text{Midpoint of column chords} = \frac{Y_c - X_c}{2}$$

Where $c=(1,2,3,\dots,n)$ No of boundaries
 $Y_c = Y$ point of c 'th column boundary point
 $X_c = X$ point of c 'th column boundary point

After successfully analysed of above implementation model, we have obtained the final set of row and column symmetry axis. Then we have developed the constellation model according to their relative symmetric axis pixel position and midpoint pixel angle from centre of the image. This constellation model generates two set of parameter for each row symmetry axis and column symmetry axis. Where one parameter show mean value of relative distance of every symmetry axis pixel position to centre of the image and other parameter shows the angle between the midpoints of the symmetry axis to the centre of the image. Thereafter we found four parameter of each image having two parameter each for row symmetry axis and column symmetry axis presented in Figure 8 and Figure 11 respectively.



Figure 2. Sample image



Figure 3. Skeletonized images



Figure 4. Angle Pixel point extraction

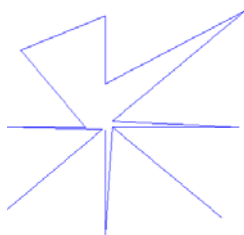


Figure 5. Plotting pixel point

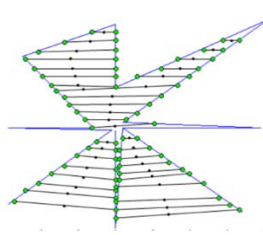


Figure 6. Row symmetry axis

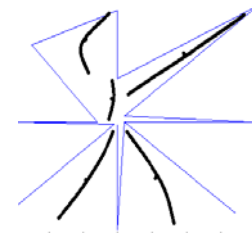


Figure 7. Midpoint of row symmetry axis

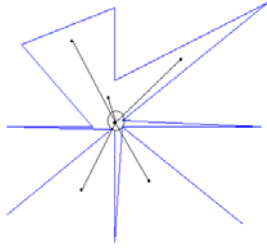


Figure 8. Angle and distance from centre

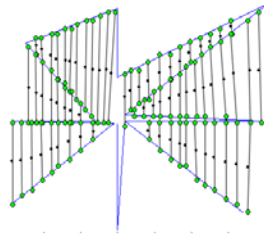


Figure 9. Column symmetry axis

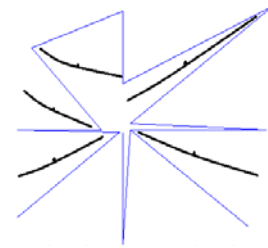


Figure 10. Midpoint of column to midpoint in row symmetry

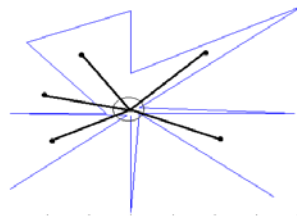


Figure 11. Angle and distance from centre to midpoint in column symmetry axis

3.7. Classification

Classification is one of the important phases of any recognition model. According to our implementation model we have adopted a two way strategy for recognition. In this regard we have chosen two well liked classifier namely support vector machine (SVM) [24] and random forest tree (RFT) [25] for recognition of handwritten characters. After evaluating the desired key feature values we process these vector to classifier separately and noted down the overall recognition accuracy. We have first evaluated the SVM [16] classifiers which are multi-class classifier and supervised one. Secondly random forest tree [25] which is work based on the idea of bagging and random selection of features. All the performance was listed depending upon the value of the mean square error. And tells about which classifier is the best one.

4. RESULT AND DISCUSSION

All the implementation of our proposed method were carried out with the system having specification with windows 8, 64 bit operating system, and Intel (R) i7 – 4770 CPU @ 3.40 GHz, and all the simulation is done through matlab14 (a) over a standard database. As per standard Database containing 200 samples from each of the 47 categories named as NIT Rourkela Odia database and considering numeric database from ISI Kolkata having 16 samples from each of the 10 categorised. After getting the four key feature vector values from each database as mean distance of row, mean angle of row, mean distance of column and mean angle of column from centre of the image to midpoint of the symmetric axis from each image. Hence total size of input for Odia character becomes 4×9400 and numeric character becomes 4×9400 and makes these as input to well defined classifier such as SVM and random forest and also performed the validation by implementing 10 fold-cross validations to the system. Consequently all the observation was counted to certain as 75, 25 ratio as training and testing. At first SVM classifier is implemented followed up by random forest classifier. We have also made a comparison analyses among these two classifiers, and listed 93.6% as the recognition rate for SVM and 98.2% for the random forest for NIT Odia character, similarly for ISI numeric character the recognition rate for both SVM and random forest as 88.91% and 96.3% respectively.

5. CONCLUSION

In this paper, we have presented an angular symmetric constellation technique for offline Odia characters recognition. This system uses row and column symmetric axis for generating four key feature vector values from each database as mean distance of row, mean angle of row, mean distance of column and mean angle of column from centre of the image to midpoint of the symmetric axis from each image. For classification purpose, SVM and RF model is used. An experimental result from this research gives

satisfactory recognition result over the standard dataset, but still the development is in its infancy. Further, other techniques are to be explored for better recognition accuracy.

REFERENCES

- [1] H. Swethalakshmi, *et al.*, "Online handwritten character recognition of devanagari and telugu characters using support vector machines," *2006 International Workshop on Frontiers in Handwriting Recognition*. 2006.
- [2] U. Pal, *et al.*, "A system for off-line Oriya handwritten character recognition using curvature feature," *2007 international conference on information technology, India*, pp. 227-229, 2007.
- [3] V. K. Govindan, and A. P. Shivaprasad, "Character recognition a review," *Pattern Recognition*, vol. 23, pp. 671-683, 1990.
- [4] J. Mantas, "An overview of character recognition methodologies," *Pattern Recognition*, vol. 19, pp. 425-430, 1986.
- [5] U. Pal and B. B. Chaudhuri, "Indian script character recognition: a survey," *Pattern Recognition*, vol. 37, pp. 1887-1899, 2004.
- [6] R. Plamondon, S. N. Srihari, "On-Line and off-line handwritten recognition: A comprehensive survey," *IEEE Trans on PAMI*, vol.22, pp.62-84, 2000.
- [7] N. Arica and F. Vural, "An Overview of Character Recognition Focused on Off-Line Handwriting," *IEEE Transactions on System Man Cybernetics-Part C: Applications and Reviews*, vol.31, pp. 216-233, 2001.
- [8] L. M. Lorigo and V. Govindaraju, "Offline Arabic Handwriting Recognition: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 712-724, 2006.
- [9] G. Nagy, "Chinese Character Recognition, A twenty five years retrospective," *1998 IEEE international conference on pattern recognition*, Italy, pp.109-114, 1988.
- [10] A. K. Jain and T. Taxt, "Feature Extraction Methods for Character Recognition-A Survey," *Pattern Recognition*, vol. 29, pp. 641-662, 1996.
- [11] <http://www.odishamuseum.nic.in/?q=dept/manuscripts>
- [12] P. Pattnaik, "Presentation on digitization of Odia books" in Utkal University (21 February 2014).
- [13] B. B. Chaudhuri, *et al.*, "Automatic recognition of printed Oriya script," *Sadhana*, vol. 27, pp. 23-34, 2002.
- [14] M. Obaidullah, *et al.*, "Structural feature based approach or script identification from printed indian document," *2014 IEEE international conference on signal processing and integrated networks, India*, pp. 120-124, 2014.
- [15] T. Wakabayashi, *et al.*, "F-ratio based weighted feature extraction for similar shape character recognition," *2009 IEEE international conference on document analysis and recognition, Spain*, pp. 196-200, 2009.
- [16] K. Roy, *et al.*, "Oriya handwritten numeral recognition system," *2005 IEEE 8th international conference on document analysis and recognition, South Korea*, pp. 770-774, 2005.
- [17] T. K. Bhowmik, *et al.*, "An HMM based recognition scheme for handwritten Oriya numerals," *2006 IEEE international conference on information technology, India*, pp. 105-110, 2006.
- [18] K. S. Dash, *et al.*, "A hybrid feature and discriminant classifier for high accuracy Odia handwritten numeral recognition," *2014 IEEE technical symposium, Malaysia*, pp. 531-535, 2014.
- [19] K. S. Dash, *et al.*, "Non-redundant Stockwell transform based feature extraction for handwritten digit recognition," *2014 IEEE signal processing and communication, India*, pp. 1-4, 2014.
- [20] T. K. Mishra, *et al.*, "A comparative analysis of image transformations for handwritten Odia numeral recognition," *2013 IEEE international conference on advances in computing, communications and informatics, India*, pp. 790-793, 2013.
- [21] B. Majhi, *et al.*, "Efficient recognition of Odiya numerals using low complexity neural classifier," *2011 IEEE international conference on energy, automation, and signal, India*, pp. 1-4, 2011
- [22] S. Chanda, *et al.*, "Text independent writer identification for Oriya script," *2012 IEEE international workshop on document analysis systems, Australia*, pp. 369-373, 2012.
- [23] K. S. Dash, *et al.*, "BESAC: Binary External Symmetry Axis Constellation for unconstrained handwritten character recognition," *Pattern Recognition Letters*, vol. 38, pp. 413-422, 2016.
- [24] B. Xu, *et al.*, "An improved random forest classifier for image classification," *2012 IEEE International Conference on Information and Automation, China*, pp. 795-800, 2012.
- [25] C. Mitra and A. K. Pujari, "Mining Intelligence and Knowledge Exploration. Editor. Mining Intelligence and Knowledge Exploration," *2009 Springer Lecture Notes in Computer Science, India*, pp. 82-84, 2013.
- [26] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, pp. 5-32, 2001.