

Off-line Odia Handwritten Character Recognition: an Axis Constellation Model Based Research

Abhisek Sethy, Prashanta Kumar Patra

Abstract— *Handwritten Character Recognition is most challenging area of research, in which for various aspects a little enhancement can be always achieved. It is due to the irregularity of writing and shapes of different class user's orientation affects the recognition rate. In this paper we have taken the complexity of Odia handwritten character recognition and successfully resolve with Principal Component Analysis (PCA). Here we had adopted a model in which the importance of symmetric axis chords in recognition of unconstrained handwritten characters is established. This symmetric axis chords are drawn along both row-wise and column-wise among the points one end to other. In addition to we have calculated the statistical feature as Euclidian distance, Hamilton distance which drawn from the midpoint of the symmetric chord to nearest pixel of the character. Apart from it we have also reported the angular values from the centroid of the image to the character pixel. This empirical model also harnessed the PCA over the feature set and perform the dimension reduction to the feature set which later termed as the key feature set. A certain series of experiment was carried on for the proper implementation of proposed technique, henceforth we have taken the standard Handwritten Database from various research institutes. Lastly on simulation analysis Radial Basis Function Neural Network (RBFNN) has been reported as to achieve high recognition rate through Gaussian kernel and a comparison among them has also reported here with.*

Rundown phrases— *Optical Character Recognition; Principal Component Analysis (PCA); Radial Basics Function, Neural Network(NN); Euclidian Distance; Hamilton Distance.*

I. INTRODUCTION

Optical character recognition (OCR) is one of the most active challenging tasks in the field of pattern recognition. The primary goal in OCR is that it converts the physical based scripts to digital one which is quite helpful to the digital world of document digitization [1]. It has wide scope of application such as it may be considered as more productive in case of storage, indexing, and retrieval of data from larger dataset and lastly make it to the digitisation for further use [23,24]. To fulfil the need of the digital world many attempts have been made by researchers in the last decades to develop ideal OCR systems [2]. These recognition systems have not only developed for printed and handwritten characters, but also for degraded characters. The existing OCR systems of various handwritten scripts like as Roman, Chinese, Japanese and Arabic characters were listed in [3]. As comparison of other scripts, it is a tough task of

recognition multilingual and multi script country like India the work is more relevant [4]. However, there are not enough works on Odia character recognition. Odia script language is one of the regional languages of India, which is derived from Devanagari scripts and mostly speaks at the Eastern part of India [5]. In this script, it has most of the characters are identical characters in terms of similar orientation, shape for which it is very difficult to develop an ideal OCR system. Additionally, most of the characters are round in nature which again possesses extra problems in the recognition process. Recognition of handwritten Odia characters is quite more challenging than recognition of printed Odia characters. Odia is a popular script of India which is mostly spoken in the state of Odisha and some of its nearby states. It is developed from the Brahmi script of ancient India. Odia script consists of 10 numeral and 49 (35 consonants and 14 vowels) characters, apart from that there are some limited conjunct characters. It is to be noted that Odia script does not have concepts of upper and lower case.

Handwritten Character recognition is quite challenging as compared to printed ones. In every recognition system each stage has its own significance starting from image acquisition, pre-processing, feature extraction and lastly classification one [6]. But all the emphasis is put over to adopt a well-defined method for feature extraction so that it is quite easy to report high recognition rate. Here all the simulation work has been projected to the handwritten Odia characters on benchmark dataset. In that context we have drawn horizontal and vertical symmetric axis and termed them Chords along the points of orientation of handwritten characters. After that we have drawn the horizontal and vertical projection angles from the pixel to the centre of the image and noted at variation of angle at different angular values. In addition to it we had also calculated the Euclidian distance, Hamilton distance which drawn from the midpoint of the symmetric chord to nearest pixel of the character. All the empirical calculation was listed based upon the symmetric chords we proposed this recognition model and named as Axis Constellation Method. After retrieving the feature vector, we have evaluated principal component analysis (PCA) [10] to the feature which has been extracted from Axis Constellation Method to make reduction to dimension to the feature vector. Finally, Radial Basis Function Neural Network (RBNN) helps to correctly

Revised Manuscript Received on July 18, 2019.

Abhisek Sethy, Department of Computer Science and Engineering,
Koneru Lakshmaiah Education Foundation, Guntur, Andhra Pradesh,
India.

Prashanta Kumar Patra, Department of Computer Science and
Engineering, College of Engineering & Technology, BPUT, Odisha, India.

classify the character and helps to achieve a good recognition rate. After certain simulation we had also validate the proposed system with 10-fold validation schema. Hence forth the paper have partitioned into section and organised as Section 2 suggested the related work as literature review. Section 3 depicted the adopted recognition model and Section 4 speaks about the implementation about the proposed model, Section 6 and Section 7 represents the respective result analysis and conclusion along with its remark.

II. RELATED WORK DONE

In [4, 7], it has been reported that some work has been done towards the offline handwritten recognition system, it conveys the idea how effectively one must take of each segment of the recognition model. In context to Indian scripts while a very less works has been done on handwritten characters. In this paper, we only focus on recognition of one Indian script that is Odia as to the best of our knowledge only some OCR has been developed for recognition of Odia scripts [8]. Patra et al. in [9] has reported the variant and in-variant character images. For feature selection they had listed the Zernike moment and Mellin transform over printed characters and obtained 99% recognition rate using neural network. In [10] Pal et al. has introduced the curvature analysis of each individuals handwritten character later they had reported as feature values and also introduced the PCA over it. By such a new feature vector is performed having reduced dimension of feature. As an overall accuracy 94.6% rate is reported on handwritten characters using quadratic classifier. Apart from this some work has also reported to the printed Odia characters by Chaudhri et al. in [11]. Here they had suggested the water reservoir feature values are utilized and obtained 96.3% as the recognition rate. Padhi et al. in [12] have suggested a recognition model depending a two-way based approach that is in between similar and distinct characters. In addition to it they had extracted the feature vector matrix depending upon the standard deviation and zone based average centroid distance of images. At first attempt they had classified the characters of similar groups and at second stage for each individual handwritten character in the recognition system. In [13] Pujari et al. had calculated the directional features depending upon the variation zone. They had also shown interest in zoned based and make SVM as the classifier and reported 95% recognition rate. Some algorithm for classification was applied on thinned characters by Arun K. Pujari et al. in [14]. All the calculation was done to skeletonize images of character in order perform stroke preservation. They had used 10 different algorithms to maintain the structural analysis of numerals such as connectivity, topological etc. An application Tesseract OCR Engine over Odia printed document was reported by Nayak et al. in [15].

Sethy et al. in [16] had also reported the statistical feature set of the characters. In order to have that they had implemented wavelet transformation. They had also noted down along all the decomposition level at second stage reported the approximation vector. Which is then processed over the analyzed by PCA over the feature vector. And with the help of BPNN they had obtained 94.8% accuracy over a

standard database. Another type of matching was suggested by Dash et al. in [17]. They had also followed a two-way technique classification by considering histogram orientation analysis of the individual characters. For appropriate recognition they had chosen random forest tree, SVM and K-NN as the classifier and subsequently reported 89.92%, 93.77% and 95.01% recognition rate respectively. In addition to it some transformation-based approach was also reports, among the Mohapatra et al. in [18] has also introduced the orthogonal S-Transformation based features, which is discrete in nature. Here the character images variant one and performs feature reduction through PCA. As an outcome based they had listed 98.5% accuracy over the handwritten samples by implementing through BPNN as classifier. Once again Sethy et al. in [19] also reported an extended version of wavelet named as Ripplet transformation. The ripplet based feature value is quite production in terms of achieving good accuracy. For classification stage we have adopted Neural Network and HMM as the two classifier and have achieved 98.2% and 96.3% over all recognition rate for Odia handwritten characters.

III. ADOPTED RECOGNITION MODEL FOR HANDWRITTEN CHARACTER RECOGNITION

In this paper we had suggested a model which is robust in nature for recognition of Odia handwritten characters. The overall recognition model provides a unique way of approach for recognize of Odia Character and the complete model is shown in graphically in figure. 1. The phases of the proposed model includes such as Image Acquisition, Pre-processing, Feature Extraction, and Classification. The details discussion can be made in several sub-chapters in subsequent section.

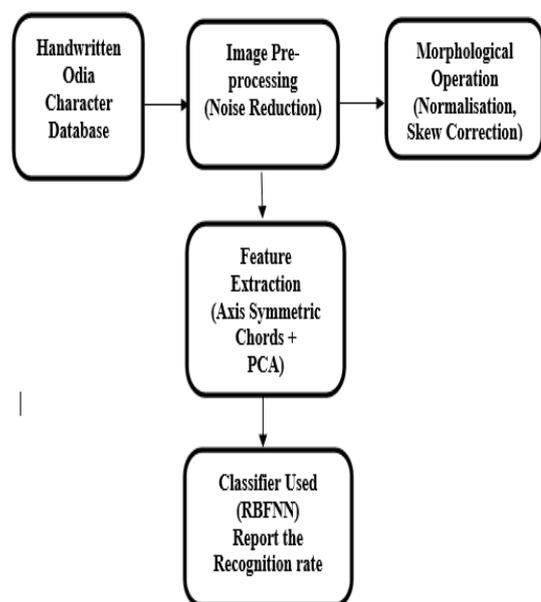


Fig. 1. Schematic Model of Odia Handwritten Recognition Model

1.1. Dataset Used

In order to have a well-defined recognition model one must have a good database. This Database is the bench mark one which is delivered by various research institutions across the globe. Here in this handwritten recognition system we had evaluated on a benchmark standard database of Odia character named as Nit Rourkela Odia database and that has been proposed by Mishra et al. [19] at NIT, Rourkela. In the above said database we have considered 350 samples in number of each 47 characters. By such the input size of handwritten character and 350*47(16,450) images respectively and sample of the handwritten characters is shown n below fig. 2.



Fig. 2. Handwritten Odia Character Sample

1.2. Image Pre-Processing

Preprocessing is way of maintaining the novelty of the picture by removing the noise and it is quite helpful in enhance the quality of the image. In this proposed work we had used the median filter and performed normalization to each pixel. After getting a normalized value we have performed the skew correction of each handwritten along with some morphological method like dilation over the handwritten image to report the boundary line of the Odia characters.

1.3. Adopted Feature Extraction Methodology

Feature extraction is way identifying the characteristics of each individual so that uniqueness can achieved from the rest character images. To achieve a good feature values, we have harnessed a new algorithm over the handwritten characters. Initially we have drawn the symmetric chords along pixel to nearest pixel of the characters which is termed as Symmetric Axis drawn and named as Axis Constellation Model. After drafting the symmetric chords, we have listed the distance between the pixel to the midpoint of the both symmetric chords. These distances are calculated in terms of Euclidian and Hamiltonian distance along both horizontal and vertical axis. The midpoint is calculated as below

$$Mid_{point(M)} = ((X_1+X_2)/2, (Y_1 + Y_2)/2) \quad (1)$$

Apart from this we have also added one more that is from horizontal vertical projection we have noted down what are the projection angle that are encountered from center of the image to nearest pixel. One basic advantage is that we have performed the operation on skeletonized images. Here all the feature values have evaluated while the images passes through at various angle such as 45°,135°,180°,225°,270°,315°,360° respectively from the

center of the image.All the empirical calculation was calculated and clearly mentioned the following Algorithm. I and Algorithm. II.

1.4. Algorithm I: Horizontal and Vertical Symmetry axis

Input: Report the boundary points (N) of handwritten character images

Output: Report the Horizontal Symmetric Axes (HE_i, i=1, 2, .., m)

Report the Vertical Symmetric Axes (VE_i, i=1, 2,, n)

1. For every i = 1 to **Nperform**

While Until horizontal symmetric chords is satisfied.

Map the row chords accordingly so that sets of parallel row chords are grouped.

End While.

2. **While** Until vertical symmetric chords is satisfied

Map the vertical chords accordingly so that sets of vertical column chords are grouped.

End While.

3. **End For.**

4. Report each Midpoint of Horizontal and Vertical Symmetric Axes.

Resulting axes consist of midpoints of horizontal Symmetric Axes.

Provide potent Horizontal symmetry axes (HE_i)

Resulting axes consist of midpoints of Vertical Symmetric Axes.

Provide potent column symmetry axes (VE_i)

5. Return HE_i and VE_i

1.5. Algorithm II: Feature Extraction

Input: Report the boundary points (N) of handwritten character images

Report the Horizontal Symmetric Axes (HE_i, i=1, 2, ..., m)

Report the Vertical Symmetric Axes (VE_i, i=1, 2,, n)

Output:Symmetric Axis Cancellation Model (SACM) Feature (FV).

1. Start **For** i = 1 to m **do**

2. Report each Midpoint of Horizontal Symmetric Chords.

Calculate projection angle centre of image to midpoints of horizontal symmetry axis.

Calculate Euclidian and Hamilton distance which drawn from the midpoint of the symmetric chord to boundaries of row symmetry axis.

3. **End For**

4. **For** i=1 to n **do**

5. Report each Midpoint of Vertical Symmetric Chords

Calculate projection angle centre of image to midpoints of vertical symmetric axis

Calculate the Euclidian and Hamilton distance which drawn from the midpoint of the symmetric chord to boundaries of row symmetry axis.

6. End For

7. Calculate mean projection angle centre of image to midpoints of Horizontal and Vertical symmetric axis

Report the mean Euclidian and Hamilton distance which drawn from the midpoint of the both symmetric chord to boundary pixels.

8. Return FV.

On evaluating both the above-mentioned algorithm we can produced a good no feature values of the handwritten images are reported. In this work we have focused on axis symmetry chord which are drawn from one pixel to other boundary pixel and calculate the distance from the midpoint of the axis along both horizontal and vertical wise respectively and take a mean of the distances. In subsequent step we have also report the angle of projection of the pixel to the center of the image, All these are depicted in the fig. 3 and fig.4 and fig. 5.

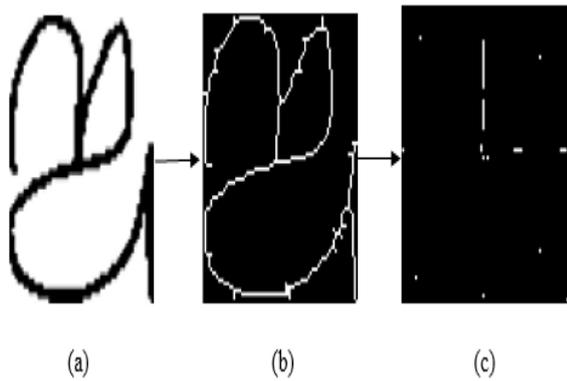


Fig.3. (a) Character Image ,(b) Respective Skeletonize Image ,(c) Angel pixel Point

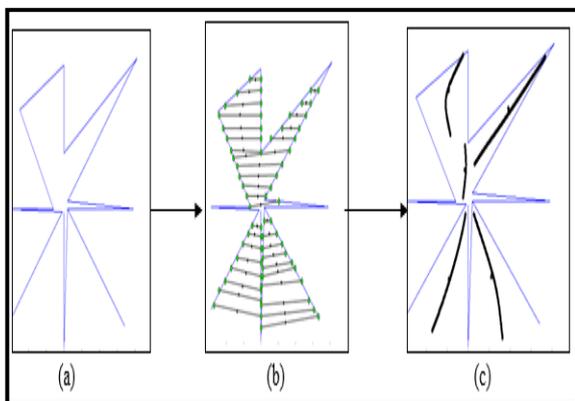


Fig.4. (a) Draw the Pixel Point, (b) Draw the Horizontal Symmetric Axis, (c) Midpoint of Horizontal Symmetric Axis

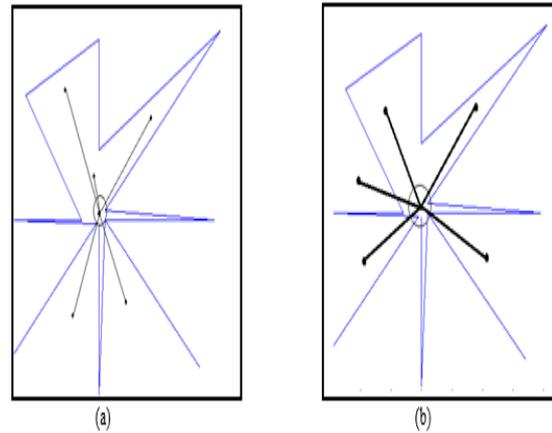


Fig.5. (a) Angle & Distance pixel to row mid of Horizontal Symmetric (b) Angle & Distance pixel to row mid of Vertical Symmetric

1.6. Principal Component Analysis (PCA) over Feature Vector

After extracting all the respective distance and angular values along both the symmetric axis are process to next step. In the following Step Principal Component Analysis [16] is harnessed over the feature set. Already we known that less the no feature it will take less computation time, and this can achieve by mapping the higher dimensional value to the lesser dimensional one. PCA [20] basically evaluates depending upon the orthogonal parameters of transform from which variance and Eigen values are reported. Subsequently a covariance matrix is formed, and the highest Eigen values are treated as the principal component and shown in Algorithm III. These Pc scores have listed down upon the algorithm 1. We have calculated 30, 40, 50 pc score for reduction in the dimension of the feature vector

1.7. Algorithm. III: Implementation of PCA

- 1.Process the higher dimension Feature vector ($M \times N$)
 M : Required Number of features, N : The overall number of images.
- 2.Calculate the mean vector($M \times 1$).
- 3.Subsequently performs the subtraction with respect to the feature along with the mean values.
- 4.Report the Co-variance matrix($M \times M$).
- 5.Note down the η Eigen vector having η high Eigen values.

Map the size of ($n \times N$) such that $n < N$.

1.8. Classification Using Radial Basic Function Neural Network

After successful completion all the above step the primary feature set is drawn, and all these are forwarded to the Radial Basic Function Neural Network (RBFNN) [21] classifier for classification. The RBFNN is basically takes a real valued function which calculates the distance from some specific point and it must satisfy the property of radial function [22] in the classification problem. In this multilayer perceptron we have maintained three layers one for input,

one for hidden and one for output. In context to this classifier the RBF neuron s takes the Gaussian function which is mostly used in the classification problem. The respective equation of the function as

$$\rho(\|X - C_i\|) = \exp[-\beta\|X - C_i\|^2] \quad (2)$$

Where ρ is the norm, C_i is the center vector of neuron.

Here all the Odia handwritten Images that is 350*47 images are labeled with 47 categories that means the output having 47 categories depending upon each character must be classified. All the performance was listed depending upon the value of the mean square error.

IV. IMPLEMENTATION OF THE PROPOSED MODEL

To evaluate the proposed model, we have taken each step of the system precisely. For successful implementation of recognition handwritten Odia Character, we have focused in various aspect in terms of pre-processing, feature technique used and at last classifier used. All the performed steps for implementing the proposed work is listed in the Algorithm. IV. And again we had classified into two stage one is offline and online procedure and mentioned below.

1.9. Algorithm IV. Implementation Offline Procedure

1. Perform the Image Acquisition of the Handwritten Odia characters from database.
2. Perform the pre-processing stage along with normalisation of the pixel values.
3. Apply the Axis Constellation Model and Perform for both horizontal and vertical symmetric axis.
4. Extract the primary features through feature descriptors.
5. Perform the dimension reduction using PCA and retain the principal components (PCs)
6. Apply Radial Basic Function Neural Network for Classification.
7. Evaluate the model based on testing data

1.10. On-line Procedure

1. Load the input query image of the User.
2. Forwarded the query image to preprocessing section.
3. Report the primary features the Axis Constellation Model
4. Retained PCs by harnessing the PCA over the primary features.
5. Apply the trained classifiers separately and predict the class label

V. ANALYSIS OVER THE SIMULATION RESULTS

Here in this paper the proposed algorithm was evaluated one by one. And all the simulation was done with a system which is of windows 8, 64-bit operating system, and Intel (R) i7 – 4770 CPU @ 3.40 GHz as system specification. To have proper simulation set up that has been depicted by through matlab14 (a) over a standard database. Here we have proposed a robust technique and name as Axis

Constellation Model, where we have made horizontal and vertical symmetric chords and calculate the distance and angle respectively which is mentioned in Algorithm I & II, from which feature are extracted. To make it less complex we have also reported the PCA over it and reported the primary features in Algorithm. III. For the novelty of the proposed system we have listed 30,40,50 pc score or primary feature and noted down the recognition rates. To make the system robust we classified the whole dataset into some data to train the system and used some for testing purpose as 70, 30 ratios. Here Gaussian Kernel function of RBFNN convergences the overlapping decomposition of problem domain. After simulation we have noted 98.8% as overall recognition rate for handwritten Odia Characters and respective confusion matrix is shown in Table. I.

Table 1. Successive recognition rate achieved for Handwritten Odia Character through RBFNN Classifier

SL. No	CC	MC	Accuracy (%)	SL.No	CC	MC	Accuracy (%)
1	346	4	98.8	25	346	4	98.8
2	346	4	98.8	26	343	7	98
3	346	4	98.8	27	343	7	98
4	346	4	98.8	28	343	7	98
5	346	4	98.8	29	343	7	98
6	346	4	98.8	30	343	7	98
7	346	4	98.8	31	341	9	97.4
8	346	4	98.8	32	341	9	97.4
9	346	4	98.8	33	341	9	97.4
10	346	4	98.8	34	341	9	97.4
11	343	7	98	35	341	9	97.4
12	343	7	98	36	341	9	97.4
13	341	9	97.4	37	341	9	97.4
14	341	9	97.4	38	341	9	97.4
15	343	7	98	39	341	9	97.4
16	343	7	98	40	341	9	97.4
17	346	4	98.8	41	346	4	98.8
18	346	4	98.8	42	346	4	98.8
19	343	7	98	43	346	4	98.8
20	343	7	98	44	346	4	98.8
21	346	4	98.8	45	341	9	97.4
22	346	4	98.8	46	341	9	97.4
23	346	4	98.8	47	346	4	98.8
24	341	9	97.4	Average Recognition Rate = 98.8%			

VI. CONCLUSION & FUTURE WORK

Here in this proposed Axis Constellation method which is meant for offline recognition of Handwritten Odia characters. Where initially we had drawn horizontal and vertical symmetric axis along pixel of boundary points and reported the midpoint of each symmetric Chords. From that we have also calculated the Euclidian Distance and Hamming distance from the point to the boundary pixel, simultaneously we have also noted the various projection angle from the center of the image to the pixel points. In context of recognition Gaussian based RBFNN model is used. We have achieved 98.8% as average recognition rate over benchmark handwritten data set. Apart from its various invented machine learning approaches such as direction



based, extreme learning based can be applied to such problem domain. Further, other techniques are to be explored for better recognition accuracy.

VII. ACKNOWLEDGEMENTS

The authors are sincerely thankful to the Department of Computer Science and Engineering, KL University Vijaywada. And we are also thankful to all the authors of the references

VII. REFERENCES

1. J. Mantas, An overview of character recognition methodologies, *Pattern recognition* 19, no. 6 (1986) 425-430.
2. V.K. Govindan, and A. P. Shivaprasad. Character recognition—a review. *Pattern recognition* 23, no. 7 (1990) 671-683.
3. R. Plamondon and S. N. Srihari, On-Line and Off-line Handwritten Recognition: A Comprehensive Survey, *IEEE Trans on PAMI*, vol. 22, (2000)62-84.
4. U. Pal and B. B. Chaudhuri, Indian script character recognition: a survey, *Pattern Recognition*, vol. 37 (2004) 1887-1899.
5. B. B. Chaudhuri, U. Pal, and Mandar Mitra, Automatic recognition of printed Oriya script., *Sadhana* 27, no. 1 (2002) 23-34.
6. W.K. Pratt, *Digital image processing* John Wiley & sons. Inc., New York. 1991.
7. Arica, Nafiz, and Fatos T. Yarman-Vural, An overview of character recognition focused on off-line handwriting. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 31, no. 2 (2001)216-233.
8. K.S. Dash, N. B. Puhana, and G. Panda, Odia character recognition: a directional review. *Artificial Intelligence Review* 48, no. 4 (2017): 473-497.
9. P. K. Patra, M. Nayak M, and S. K. Nayak and N.K Gobbak, Probabilistic neural network for pattern classification, In *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290) (2002) 2*, pp. 1200-1205.
10. U. Pal, T. Wakabayashi, and F. Kimura, A system for off-line Oriya handwritten character recognition using curvature feature. In *10th international conference on information technology (ICIT 2007)* (pp. 227-229). IEEE.
11. B. B. Chaudhuri, U. Pal, and M. Mitra, Automatic recognition of printed Oriya script. *Sadhana* 27.1 (2002): 23-34.
12. D. Padhi and D. Senapati, Zone centroid distance and standard deviation based feature matrix for odia handwritten character recognition. In *Proceedings of the International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2013* (pp. 649-658). Springer, Berlin, Heidelberg.
13. Mitra, Chandana, and Arun K. Pujari., Directional decomposition for odia character recognition, In *Mining Intelligence and Knowledge Exploration*, pp. 270-278. Springer, Cham, 2013.
14. Pujari, Arun K., C. Mitra, and S. Mishra. ,A new parallel thinning algorithm with stroke correction for odia characters, In *Advanced Computing, Networking and Informatics-Volume 1*, pp. 413-419. Springer, Cham, 2014.
15. M. Nayak and A. K. Nayak., Odia characters recognition by training tesseract OCR engine, *International Journal of Computer Applications* 975 (2013): 8887.
16. A. Sethy, P.K. Patra, and D. R. Nayak., Off-line Odia Handwritten Character Recognition: A Hybrid Approach, In *Computational Signal Processing and Analysis*, pp. 247-257. Springer, Singapore, 2018.
17. K. S. Dash, N. B. Puhana, and G. Panda., BESAC: Binary External Symmetry Axis Constellation for unconstrained handwritten character recognition., *Pattern Recognition Letters* 83 (2016)413-422.
18. R. K. Mohapatra, B. Majhi and S. K. Jena., Classification of handwritten Odia basic character using Stockwell transform, *International Journal of Applied Pattern Recognition* 2.3 (2015)235-254.
19. T. K. Mishra, B. Majhi, P. K. Sa and S. Panda, Model Based Odia Numeral Recognition using Fuzzy Aggregated Features, *Front. Comput.Sci. Springer*, (2014) 916–922.
20. D. R. Nayak, R. Dash, and B. Majhi., Brain MR image classification using two-dimensional discrete wavelet transform and AdaBoost with random forests, *Neurocomputing* 177 (2016)188-197.
21. D. Singh D, J.P. Saini and D. S Chauhan, Hindi character recognition using RBF neural network and directional group feature extraction technique, In *2015 International Conference on Cognitive Computing and Information Processing (CCIP) 2015 Mar 3* (pp. 1-4). IEEE.
22. Er, Meng Joo, Shiqian Wu, Juwei Lu, and Hock Lye Toh. ,Face recognition with radial basis function (RBF) neural networks., *IEEE transactions on neural networks* 13, no. 3 (2002)697-710.
23. P. Ajitha, Dr. G. Gunasekaran.,” Effective Feature Extraction For Document Clustering To Enhance Search Engine Using Xml”, *Journal of Theoretical and Applied Information Technology*, Little Lion Scientific. PP 20-26, 10th October 2014. Vol. 68 No.1. EISSN 18173195 ISSN 19928645.
24. Dr.R.Subhashini and Niveditha.P.R, "ANALYZING AND DETECTING EMPLOYEE'S EMOTION FOR AMELIORATION OF ORGANIZATIONS", *International Conference on Intelligent Computing, Communication & Convergence (ICCC-2014)*, Bhubaneswar, Odisha on 27th-28th December, 2014.