iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VIII **Month of publication:** August 2022

DOI: https://doi.org/10.22214/ijraset.2022.41764

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VIII Aug 2022- Available at www.ijraset.com

Handwritten Steno Character Recognition in Gujarati

Prof. Shweta Joshi[1], Prof. Kishita Doshi[2], Prof. Pooja Pandya[3]
Department of Computer Science, Dr. Subhash University, India

Abstract: ***Steno is symbol of shorthand character which is used to record spoken word. court reporters use the steno language for quick recording of each moment in court room. Some stenographer can write 300 word per minute. many implementations has been done in recognition of different steno symbol into particular language. In this paper we recognize steno symbol (small script) into Indian language which is gujarati. HSCRG stands for Handwritten Steno Character Recognition In Gujarati. HSCRG Algorithm is a process of recognition steno symbol and extract it into equivalent gujarati character based on PCA and Logistic Regression. HSCRG is based on PCA and Logistic Regression accordingly unsupervised and supervised learning. These system mainly has five stages Preprocessing, Feature Selection, classification, Logistic Regression, Extract into gujarati character.***

Keywords: PCA, Logistic Regression, Machine Learning, Steno Symbol

I. INTRODUCTION

The process of writing in shorthand is called stenography, from the Greek stenos (narrow) and graphein (to write). It has also been called brachygraphy, from Greek brachys (short), and tachygraphy, from Greek tachys (swift, speedy), depending on whether compression or speed of writing is the goal.[1] The word stenography is of Greek origin (στενός 'narrow, compact' and γράφω 'to write'). This is the name of fast handwriting made by means of special signs and acronyms, permitting a person to listen to oral speech and record it simultaneously, and also to improve writing techniques. All over the world, stenography was primarily used at meetings, sessions of state and public organizations, legislative, judicial, and other institutions. [9] Stenography provides short forms for words and characters. As technologies are growing all the things are connected with computer. In the past stenography is used to record the data in mainly in journalism, or secret data transformation. To record the data in digital devices we have to convert stenographic character into normal character. Our proposed work is mainly based to recognize the Gujarati steno character and convert it into normal Gujarati readable form. There are so many character recognition techniques to fetch the character but in this proposed work we have used PCA technique to recognize the Gujarati steno character. In recent trends technology and artificial intelligence are mostly used to describe handwriting in many languages. Handwriting recognition is made possible because of Artificial Neural Networks (ANN), a widely used concept to imitate the workings of the human brain. [17] Character recognition methods are listed under two main headlines. The "Offline" methods use the static image properties. The offline methods are further divided into four methods, which are Clustering, Feature Extraction, Pattern Matching and Artificial Neural Network. The online methods are subdivided into k-NN classifier and direction based algorithm.[15] Offline handwritten character recognition is the process of converting offline handwritten characters into a machine process-able format. In this paper, character recognition system using principal component analysis (PCA). A handwritten character recognition system consists of several phases, namely digitization, preprocessing, feature extraction and classification. The feature extraction stage analyzes a handwritten character image and selects a set of features that can uniquely be used for recognition of that character. Different feature extraction methods have been proposed for representation of characters, such as projection histograms, contour profile, zoning, Zernike moments, gradient features and Gabor features, etc. Singhet al. [17] presented a study of different feature extractors and classifiers for handwritten Devanagari character recognition. Aradhya et al. [5] presented a system based on PCA for online handwritten character recognition. Sundaram and Ramakarishnan [18] presented 2D-PCA for online Tamil character recognition. Bhattacharya et al. [3] presented an efficient two-stage approach for handwritten Bangla character recognition. Kumar et al.[7] presented an offline handwritten Gurmukhi character recognition system based on support vector machines (SVM). In that work, they performed recognition without using PCA and used only an SVM classifier for classification purpose. They also provided an offline handwritten Gurmukhi character recognition system using a k-nearest neighbor (k- NN) classifier [8]. Sharma et al.[16] presented an online handwritten Gurmukhi Script recognition system.

©IJRASET: All Rights are Reserved

They used an elastic matching method in which the character is recognized in two stages. The first stage recognizes the strokes and, in the second stage, the character is constructed on the basis of recognized strokes. In the present work, a PCA-based offline handwritten Gurmukhi character recognition system is proposed from experimenting with different recognition methods, namely, k-NN, Linear-SVM, Polynomial-SVM, RBF-SVM and combinations of these recognition methods

II. RELATED LITERATURE

P NAGABHUSHAN and BASAVARAJ S ANAMI have proposed a knowledge-based approach for recognition of handwritten Pitman shorthand language strokes. where English is the transaction medium. It involves discipline like Pitman shorthand language, character recognition, English text, production and primitives.[2] Sadanand A. Kulkarni, Prashant L. Borde, Ramesh R. Manza, Pravin L. Yannawar have proposed recognition of different handwritten characters from a digital image of documents called HOCR(Handwritten Optical Character Recognition). It involves discipline like Handwritten Optical Character Recognition.[3]. Noman Islam, Zeeshan Islam,Nazia Noor have proposed digitizing a document image into its constituent characters and converts printed text and images into digitized form such that it can be manipulated by machine using different technique and OCR.[4] Rajasekaran. R and Dr. K. Ramar have proposed hand written Gregg shorthand character or word is converted into an appropriate Symbolic Character in a computer (It may be a scanned character or online written character) using Hand writtenGregg Shorthand,Competitive and Artificial Neural Network.[5] Leedham, Graham, Ma, Yang, Blumenstein, Michael have proposed Progress and critical research issues for a Pitman/Renqun Shorthand Online Recognition System.). It involves discipline like Hausdorff distance, Pen computing, Bayesian networks.[6] Francis Jesmar P. Montalbo and Davood Pour Yousefian Barfeh have present 2000 common court stenography words and phrases as the core dataset for training,validation, and testing in English It involves Convolutional Neural Networks, Stenography,Canny Edge Detection for classification of stenography[9].] Sang-Keun, Kim have presented Korean stenographic recognition algorithms, which recognize the characters by using some methods, have a quantitative problem that despite the simplicity of the structure, a lot of basic characters are impossible to classify into a type using Error Back Propagation Algorithm.[10]. R.Ethindra Kumar , A.Agarwal , Arun K. Pujari have proposed the possibility of fast man-structural characteristics which is the most machine communication through printed desired features for automatic recognition using SEGMENTOR.[12] Jayashree Rajesh Prasad and Uday Kulkarni have present a handwritten character recognition system for Gujrati, an Indian language spoken by 40 million people.In this paper a unique pattern descriptor and Gabor phase XNOR pattern are the two features that are newly proposed for isolated handwritten character set of
Gujrati. It contributes is a novel mean X2 distance measure with the help of KNN and it gives better accuracycompare to K-NN classifier. [13] Khuat Thanh Tung and Le Thi My Hanh uses Principal component analysis with multilayer network to recognize text from image.[21] Vanita Mane and Lena Ragha present a new elastic image matching (EM) technique based on an eigen-deformation for recognition of offline isolated English uppercase handwritten characters and offline isolated handwritten characters of Devnagari ,the most popular script in India.[23] C. Vasantha Lakshmi, Ritu Jain and C.Patvardhan implements novel way ideas regarding edge directions histograms and splines along with PCA for enabling recognition accuracies[25]

A. PCA

The handwritten characters can change with every person. So, it is complex to recognize the text and convert it to any specific language. Here we have applied the concept of PCA for handwritten character recognition in Gujarati language. There are application areas where the concept of PCA can be used which are to find pattern in data, to recognize face and so many more. It is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. With this method we can identify data patterns and also it becomes easy to find similar and different data. To analyze the data PCA is robust tool. PCA is one of the best techniques whenever there is a need of finding data pattern and compress data without losing so much information.

B. Logistic Regression

It is used for predicting the categorical dependent variable using a given set of independent variables. We can use this method to predict the output of a categorical dependent variable. The dependent variable should be categorical. The independent variable cannot be co linear. There are mainly three types of LR can be there. We can categorize logistic regression as binomial, multinomial and ordinal etc.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

III. METHODOLOGY MATERIAL

To achieve the research goal provides some methodology with the used of experiments. Fig-2 illustrates the methodology for the research. The dataset consists stenography symbols. Dataset contained preprocessing through feature extraction for each image. PCA or feature selection. After preprocessing, the data enters for classification. PCA then determines the classification task.

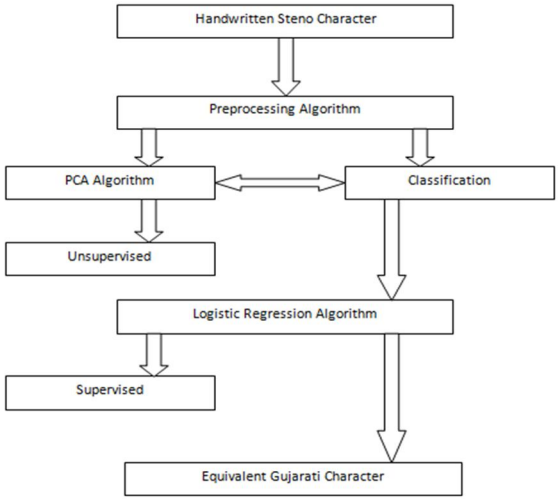


Fig 2 HSCRG Methodology

Methodology diagram of Handwritten Steno Shorthand Recognition In Gujarati is as above. Handwritten steno shorthand image was taken as an input(offline or online). Preprocessing algorithm performs some filter like gray image, binary image, edge image resize and filter image. PCA algorithm perform mathematical task like subtract mean, covariance, and Eigen value. Classification algorithm used some activation function that convert all real value into binary value before prepare for classification. Logistic regression is multinomial logistic regression in which the target variable can have 3 or more possible unordered types in supervised learning. HSSRG algorithm compares the input steno character image with already trained set of steno character image and produced equivalent the result in gujarati character. This HGSR algorithm is a combination of all the above sub algorithms.

A. Preprocessing Algorithm

Steno shorthand Image is given as an input in this algorithm. It performs filtering operations, binary operation, resizing operations etc. The output image can be saved.

Begin preprocess

1) *Step1:* The steno character is taken as an input(Image)
2) *Step2:* Image size is determined and resize if it necessary.
3) *Step3:* Filtering can be applied based on the need of the image either it can be applied for Binerization,Noise Elimination, Skew Correction and so on.
4) *Step4:* Save Image

End preprocess

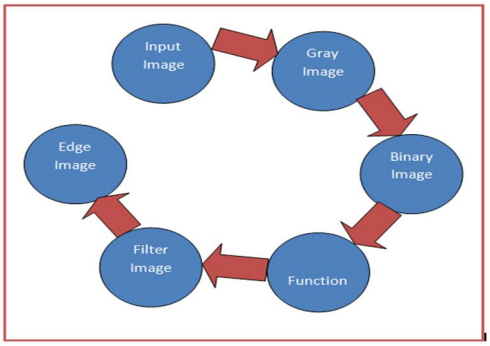


Fig.3. Preprocessing Image

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VIII Aug 2022- Available at www.ijraset.com

Begin

a) Step1: Get the data from image

b) Step2: Subtract the mean

$$\bar{X} = \frac{1}{N}\sum_{I=1}^{N} X_i$$

$$\emptyset i = X_i - \bar{X}$$

c) Step3: Calculate the covariance

$$C = \frac{1}{N}\sum_{i=1}^{N} \emptyset * \emptyset_i{}^T$$

d) Step4: Calculate Eigen vector and Eigen estimation

$$(A - \lambda I) = 0$$

Where λ is eigen value and I is matrix

e) Step5: Select part and shaping of vector.

f) Step6: Derive new Dataset.

Result Data=Row Feature*Row Adjust

End

B. PCA Algorithm

Principal Component Analysis is an unsupervised learning algorithm. This algorithm applies some sort of transformation for large set of variable into fewer variables. It performs to determine whether they were reduce redundancy and check the extent to which data compression was achieved.

PCA algorithm perform two phases for this research

1) Training Phase

2) Recognition Phase

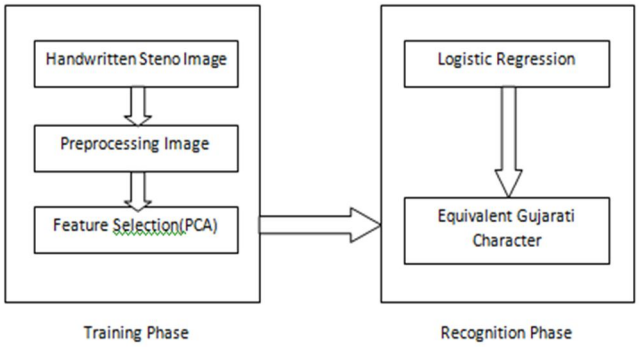


Fig.3. PCA Algorithm

C. Classification Algorithm

HSSRG will classify stenography symbol into gujarati character . HSSRG will be trained through multiple procedures to create classifications using the preprocessed images as the training set.

Begin classification

*1) Step1:*Create label accordingly category

2) Step2: Classify image as per label in directory

3) Step3: Covert image and label into numeric

4) Step4: Save new derived data for image and label

End classification

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

D. Logistic Regression Algorithm

This research used Multinomial Logistic Regression algorithm for supervised learning.

Begin Learning

1) *Step1:* Data Preprocessing
2) *Step2:* Fitting Logistic Regression for training set data.
3) *Step3:* Prediction of the test result from data
4) *Step4:* Creation of Confusion Matrix for test accuracy
5) *Step5:* Visualizing the test result

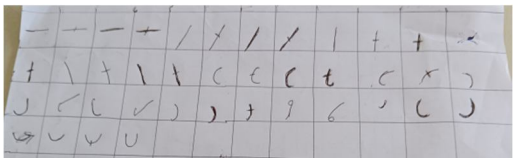
End Learning

IV. EXPERIMENT AND DISCUSSION

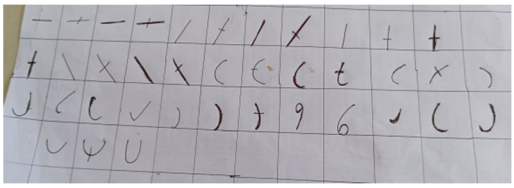
A. The Objective of Experiment

Using steno symbols as a input and extract it into equivalent Gujarati. For 34 Gujarati letter, we have used 6 steno sample symbol for recognition and extract it in Gujarati. In this research, proposed approaches which are PCA and Logistic Regression.

B. Experiment for the Case of Standard input

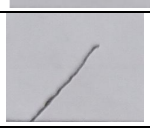


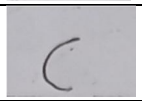
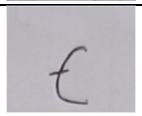
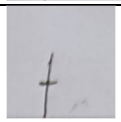
Above figure shows handwritten steno symbols which we used for sample in our research. And further we have also classified steno symbols using some algorithm like PCA and Logistic Regression.



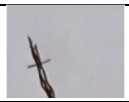
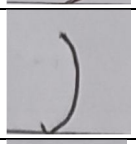
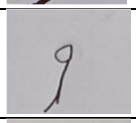
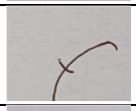
C. Evaluation they Obtained Result

Following table shows steno symbol with its stroke of nature and also display its equivalent Gujarati characters.

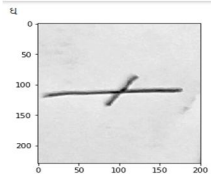
Sr. no	Steno Symbol	Stroke Nature	Equivalent Gujarati Character
1		Thin	ક
2		Thin	ખ
3		Thick	ગ
4		Thick	ઘ
5		Thin	ચ

6		Thin	૭
7		Thick	૪
8		Thick	૯
9		Thin	૮
10		Thin	૬
11		Thick	૫
12		Thick	૨
13		Thin	૧
14		Thin	૦
15		Thin	ૂ
16		Thick	૬
17		Thick	૫
18		Thin	ૈ
19		Thin	૫
20		Thin	૬

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

21		Thick	બ
22		Thick	ભ
23		Thick	મ
24		Thin	ય
25		Thin	ર
26		Thin	લ
27		Thick	વ
28		Thick	શ
29		Thin	સ
30		Thick	ષ
31		Thin	હ
32		Thin	ળ
33		Thick	ક્ષ
34		Thick	જ્ઞ

Following figure depict the outcome of our evaluation using HSCRG algorithm.



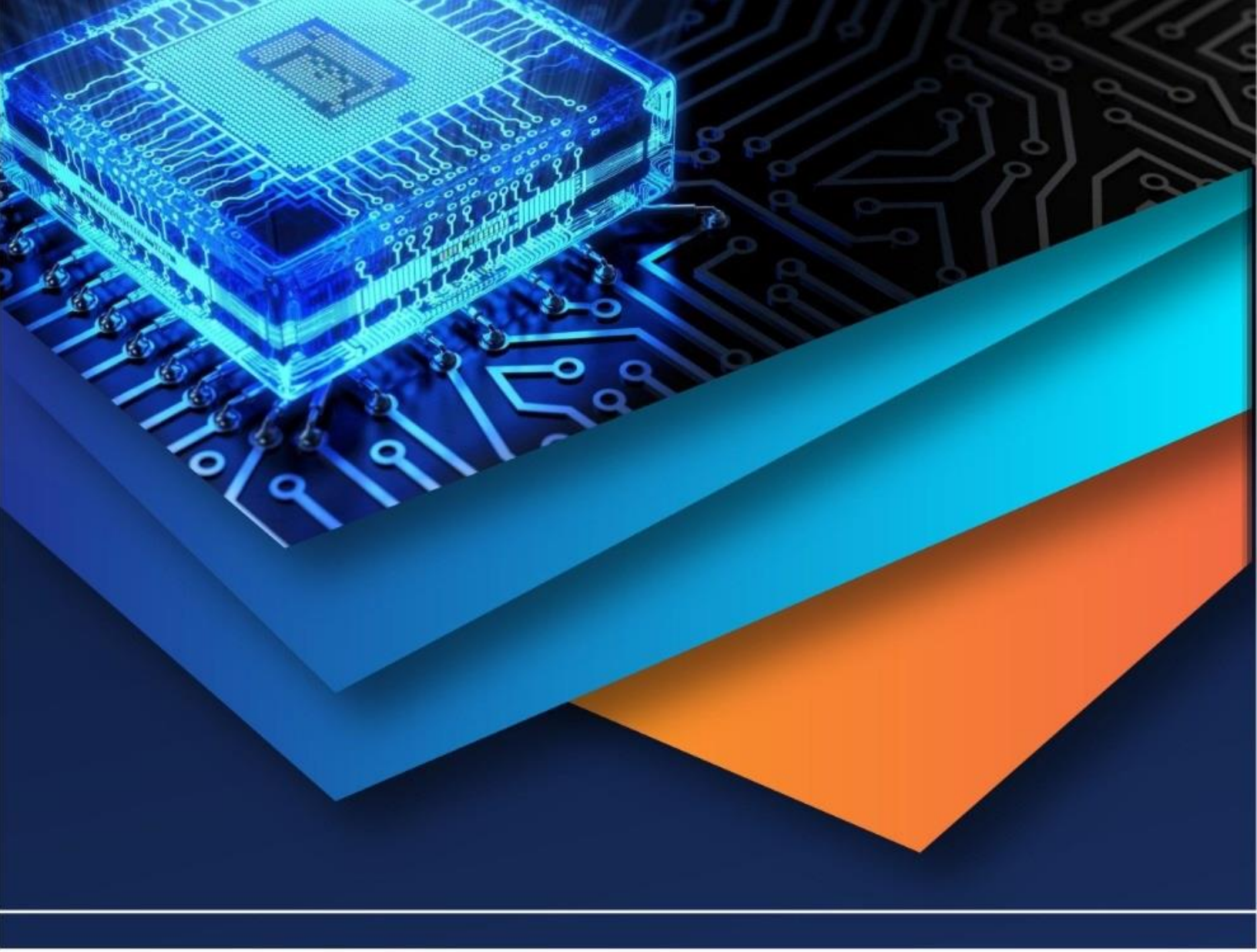
V. CONCLUSION AND FUTURE WORK

The Proposed method in this paper is applied for recognizing steno symbol which is extracted in gujarati language based on PCA and Logistic Regression. This Work is useful in different area like secret communication, legal trancription, compression etc. The success rate for recognition is 99 to 100% for ideal strokes. and 65 to 70% for somewhat lower symbol. This approach is performed only on single character recognition. Our future work will be extract vowels, word, segmentation in gujarati language.

REFERENCES

[1] Rajasekaran, R., and K. Ramar. "Handwritten Gregg Shorthand Recognition." International Journal of Computer Applications 41.9 (2012): 31-38.
[2] Naiemi, Fatemeh, Vahid Ghods, and Hassan Khalesi. "An efficient character recognition method using enhanced HOG for spam image detection." Soft Computing 23.22 (2019): 11759-11774.
[3] Chatterjee, Agneet, Sudipta Kumar Ghosal, and Ram Sarkar. "LSB based steganography with OCR: an intelligent amalgamation." Multimedia tools and applications 79.17 (2020): 11747-11765.
[4] Nagabhushan, P., and Basavaraj S. Anami. "A knowledge-based approach for recognition of handwritten Pitman shorthand language strokes." Sadhana 27.6 (2002): 685-698.
[5] Kulkarni, Sadanand A., et al. "Offline handwritten MODI character recognition using HU, Zernike moments and Zoning." arXiv preprint arXiv:1406.6140 (2014).
[6] Islam, Noman, Zeeshan Islam, and Nazia Noor. "A survey on optical character recognition system." arXiv preprint arXiv:1710.05703 (2017).
[7] Leedham, Graham, Yang Ma, and Michael Blumenstein. "Handwritten shorthand and its future potential for fast mobile text entry." International Journal of Pattern Recognition and Artificial Intelligence 23.05 (2009): 1031-1051.
[8] Andrianova, Irina Svyatoslavovna. "Stenography and Literature: What did Western European and Russian Writers Master the Art of Shorthand Writing For?." Studia Slavica Academiae Scientiarum Hungaricae 64.1 (2019): 1-11.
[9] Nagabhushan, P., and Basavaraj S. Anami. "A knowledge-based approach for recognition of handwritten Pitman shorthand language strokes." Sadhana 27.6 (2002): 685-698.
[10] Sang-Keun, Kim. "Implementation of the recognition system of the Korean stenographic characters by error back propagation algorithm." FUZZ-IEEE'99. 1999 IEEE International Fuzzy Systems. Conference Proceedings (Cat. No. 99CH36315). Vol. 2. IEEE, 1999.
[11] Dedgaonkar, Suruchi G., Anjali A. Chandavale, and Ashok M. Sapkal. "Survey of methods for character recognition." International Journal of Engineering and Innovative Technology (IJEIT) 1.5 (2012): 180-189.
[12] Kumar, R. Ethindra, Arun Agarwal, and Arun K. Pujari. "Off-line Shorthand Recognition System." MVA. 1990.
[13] Montalbo, Francis Jesmar P., and Davood Pour Yousefian Barfeh. "Classification of Stenography using Convolutional Neural Networks and Canny Edge Detection Algorithm." 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE). IEEE, 2019.
[14] Prasad, Jayashree Rajesh, and Uday Kulkarni. "Gujrati character recognition using weighted k-NN and mean χ 2 distance measure." International Journal of Machine Learning and Cybernetics 6.1 (2015): 69-82.
[15] Rajavelu, A., Mohamad T. Musavi, and Mukul Vassant Shirvaikar. "A neural network approach to character recognition." Neural networks 2.5 (1989): 387-393.
[16] Tung, Khuat Thanh. "Applying Principal Component Analysis, Multilayer Perceptron and Self-organizing Maps for Optical Character Recognition." ICTACT Journal on Image and Video Processing 6.2 (2015): 1115-1121.
[17] Rijal, Aayush. "Machine learning implementation to recognize manuscripts." International Journal for Advance Research and Development 4.7 (2019): 33-36.
[18] Mane, Vanita, and Lena Ragha. "Handwritten character recognition using elastic matching and PCA." Proceedings of the International Conference on Advances in Computing, Communication and Control. 2009.
[19] Agrawal, Mimansha, Bhanu Chauhan, and Tanisha Agrawal. "Machine Learning Algorithms for Handwritten Devanagari Character Recognition: A Systematic Review." (2022).
[20] Lakshmi, C. Vasantha, Ritu Jain, and C. Patvardhan. "Handwritten Devnagari numerals recognition with higher accuracy." International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007). Vol. 3. IEEE, 2007.
[21] Fate, Shubham Arun, Nikhilesh R. Lingayat, and Snehal Golait. "Recognition of Marathi Characters using PCA Algorithm." International Journal of Computer Applications 975: 8887.
[22] Du, Jun, et al. "Writer adaptation using bottleneck features and discriminative linear regression for online handwritten Chinese character recognition." 2014 14th International Conference on Frontiers in Handwriting Recognition. IEEE, 2014.
[23] Izumi, Tetsuya, et al. "Characters recognition method based on vector field and simple linear regression model." IEEE International Symposium on Communications and Information Technology, 2004. ISCIT 2004.. Vol. 1. IEEE, 2004.
[24] Ho, Tin Kam, Jonathan J. Hull, and Sargur N. Srihari. "Regression approach to combination of decisions by multiple character recognition algorithms." Machine Vision Applications in Character Recognition and Industrial Inspection. Vol. 1661. International Society for Optics and Photonics, 1992.
[25] Al Smadi, Takialddin A., and Mohammed Maitah. "An efficiency and algorithm detection for stenography in digital symbols." International Journal of Computer Network and Information Security 6.1 (2014): 34-38.

★★★



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)