

# Identification and Classification of Historical Kannada Handwritten Scripts based on their Age-Type using Line Segmentation with GLCM features

**Parashuram Bannigidad<sup>1</sup>, Chandrashekar Gudada<sup>2\*</sup>**

<sup>1,2</sup>Dept. of Computer Science, Rani Channamma University, Belagavi, Karnataka, India

\*Corresponding Author: [chandrugudada@gmail.com](mailto:chandrugudada@gmail.com)

DOI: <https://doi.org/10.26438/ijcse/v7i3.754763> | Available online at: [www.ijcseonline.org](http://www.ijcseonline.org)

Accepted: 20/Mar/2019, Published: 31/Mar/2019

**Abstract**— The glorious history of the dynasties is recorded within the variety of inscriptions/epigraphic records. The department of ancient history and Archaeology is exhuming new inscriptions and need for the automatic explanation of such inscriptions is increasing, that minimizes the work or eliminates the necessity of partner epigrapher in translating antiquated epigraphs. The ancient inscription on the rock, metal plates, cloth and other writing materials are the main sources to recreate the culture and history of Karnataka in India. The offline handwritten text recognition is one of the most challenging tasks in document image analysis; our aim is to recreate the cultural importance of the Kannada Language writing tradition through the historical degraded manuscripts. In the present digital era, we need to protect and digitize the resources of our Indian culture and heritage by digitizing those manuscripts which are losing its status; the degraded manuscripts are influenced by weather condition. In this paper, we have attempted to identify and recognise the historical Kannada handwritten scripts of various dynasties; namely, Vijayanagara dynasty (1460 AD), Mysore Wodeyar dynasty (1936 AD), Vijayanagara dynasty (1400 AD) and Hoysala dynasty (1340 AD) by using the improved seam carving text line segmentation method with GLCM features. The average classification accuracy for different dynasties is computed. The LDA classifier has yielded 86.5%, K-NN classifier has yielded 85.3% and SVM classifier has 85.6%. Based on the experimentation, the LDA classifier has recorded good classification performance comparatively K-NN and SVM classifiers for historical Kannada handwritten scripts.

**Keywords**—Restoration, Seam carving, Line segmentation, Kannada, LDA, K-NN, SVM, Recognition, GLCM, handwritten script, historical documents, document image analysis.

## I. INTRODUCTION

Archaeological resources like inscriptions play an important role in the building of well cultured civilization by knowing the history, spiritual and astrological knowledge of ancient periods. Due to the negligence in maintaining, these inscriptions are in the state of degradedness. Hence, the digitization of these degraded documents is a challenging task to restore the deciphering inscriptions. Particularly, in the state of Karnataka, many dynasties have ruled and contributed their knowledge to the Indian civilization. In this work, we are trying to experiment with the available historical Kannada handwritten manuscripts (inscription) written on paper from various dynasties, namely; Vijayanagara dynasty (1460 AD), Mysore Wadiyar dynasty (1936 AD), Vijayanagara dynasty (1400 AD) and Hoysala dynasty (1340 AD) collected from various institutions or universities for identification and recognition of historical Kannada handwritten manuscripts.

Very few researchers have contributed to this area in the literature; Seam carving for text line extraction on colour and grayscale historical manuscript was proposed by Nikolaos et.al. [4]. Seam carving for content-aware image resizing has been investigated by Avidan et.al. [5]. Influence of text line segmentation in handwritten text recognition was presented by Romero et.al. [6]. Texture feature evaluation for segmentation of historical document images was proposed by Mehri et.al.[7]. The texture feature benchmarking and evaluation for historical document image analysis is proposed by Mehri et.al. [8]. the script recognition by statistical analysis of the image texture was investigated by Brodic et.al. [9]. Barakat et.al. [10] proposed algorithm for Arabic historical documents to binarization free layout analysis using fully convolution networks. Laurence et.al. [11] has done a survey on text line segmentation of historical documents. Text line segmentation for gray scale historical document images was proposed by Asi et.al. [12]. Parashuram and Chandrashekar [13, 14] have proposed an image enhancement method for degraded historical Kannada

handwritten document images. Identification and recognition of historical Kannada handwritten document images using GLCM features were proposed by Parashuram and Chandrashekar [15]. The main objective of the proposed work is to design an algorithm for recognize and classify the epigraphic documents from different age-types. The proposed method involves the various stages, namely, pre-processing and segmentation of epigraphic documents, recognition and classification of epigraphs into the respective age-types.

The paper is organized into various sections; Section I contains the introduction and literature survey of historical Kannada handwritten scripts using LBP features. The section II contains the details of proposed method along with algorithm. Similarly, section III describes the results and discussion of the proposed method and finally, section IV concludes research work with future directions.

## II. PROPOSED METHOD

The objective of the proposed method is to digitize and recognize the historical Kannada handwritten manuscripts based on their age types using improved seam carving text line segmentation approach by extracting GLCM features using LDA, K-NN and SVM classifiers. The proposed method mainly consists of data collection, segmentation, pre-processing, feature extraction and classification; which are discussed as below:

### i. Data collection

The availability of standard datasets of historical Kannada handwritten manuscripts are rarely found in the literature. Hence these documents are collected individually by visiting many resource centres, like; Department of P. G. Studies and Research in Kannada, Gulbarga University, Kalaburgi and Department of Hasatapatri, Kannada University, Hampi. These historical Kannada handwritten manuscripts are captured through Canon 1300D, 18 megapixels DSLR Camera at 5184×3456 resolutions in the JPEG format. There are 121 manuscripts of different dynasties are captured and stored them in a separate file.

### ii. Segmentation

Image segmentation is a process of dividing an image into multiple regions. This is typically used to identify objects or other relevant information in digital images. For historical manuscript analysis, there are many different ways to perform image segmentation to extract words or lines. In this work, we have considered the seam carving text line extraction method with the improved algorithm. Seam carving [5] algorithm is originally developed by Shai Avidan for content aware image resizing, it is also called as liquid rescaling. It works by building up various seams in the image

and consequently expels seams to diminish picture size or embeds seams to expand it. Seams cutting additionally permits manually characterizing regions in which pixels may not be adjusted and include the capacity to expel entire articles from an image. Many of the authors have used this algorithm for image retargeting, but in this work, we tried to improve the algorithm to extract text line in historical Kannada handwritten manuscripts.

### iii. Pre-processing

It is more important that the historical manuscript documents are badly affected by many factors [1, 2] namely; it contains smear, uneven background illumination, and spot due to age or marks resulting from the ink bleed-through. Apart from this, the style of writing varies from manuscript to manuscript [3], which leads to the confusion and complexity to recognize the historical manuscript documents. The pre-processing steps which include image enhancement and restoration, the enhancement method improves the quality of the image but also removes the unwanted objects, debris, uneven background illumination and noise, etc. In the previous papers, we have proposed a novel image enhancement and restoration technique for degraded historical Kannada handwritten manuscript images [13, 14].

### iv. Feature Extraction

The feature extraction process used to obtain the feature vector sequence of individual images which describes the properties of the individual objects. The main contribution of the present work is the application of GLCM features for recognition of the historical Kannada handwritten manuscript document images. In this paper, we have used GLCM features [15]; and these are extracted based on their texture features namely; correlation, contrast, shades, etc., for recognition of the Kannada handwritten documents.

### v. Classification

The image classification mainly works with the numerical properties of an assorted image features with organized classes, the image classification contains two phases i.e., training and testing phase. In the initial training phase, characteristic properties of typical image features are isolated and based on these, unique description of each classification category, i.e. training class, is created. In the subsequent testing phase, these feature-space partitions are used to classify image features. In this experiment, we have used Linear discriminant analysis (LDA), K-nearest neighbor (KNN) and Support Vector Machine (SVM), with k-fold experimentation based on the GLCM features with 24 and 20 features for classification of historical Kannada handwritten documents.

The detailed approach of the proposed method is discussed in the form of algorithm, which is described below:

### Algorithm for recognition of Historical Kannada Handwritten manuscript

1. Input Camera capture historical Kannada handwritten manuscript of different age-types: namely Hoysala, Vijayanagara and Mysore dynasties.
2. Apply Improved seam carving text line segmentation method for line Extraction from historical manuscript:
  - 2.1. convert the given original colour image to grayscale image
  - 2.2. compute edge image using the Sobel edge detector
  - 2.3. medial seam computation with a projection profile matching
    - 2.3.1. Compute horizontal projection profiles of all edge image slices and find their local maxima
    - 2.3.2. Match local maxima of the projection profiles between two consecutive image slices
    - 2.3.3. Remove lines that start from some intermediate column of the image
    - 2.3.4. Extend the small lines towards the end column of the image
  - 2.4. Separating seam computation with constrained seam carving
    - 2.5.1. apply constrained seam carving for each pair of text lines
    - 2.5.2. compute minimum energy separating seam using dynamic programming
    - 2.5.3. overlay separating seams on the original image
      - 2.5.3.1. Compute the coordinate values of the overlay separating seam and store them in temp
      - 2.5.3.2. Concatenate the present coordinate values with temp
    - 2.5.4. Using coordinate value, extract the region of interest by roipoly() function
    - 2.5.5. Apply the image enhancement technique for binarization to the extracted region of interest (text line)
  - 2.5. Original image overlaid with both types of seams
  - 2.6. Apply the skew correction to the segmented text line
3. Combination of Local Otsu and Global Otsu method is applied to each individual text line for binarizing the images on step2
4. Apply size normalization to each individual text line on step3
5. Extract GLCM features of size normalized individual text line of different dynasties, namely; Hoysala, Vijayanagara and Mysore dynasties and store them as a knowledge base

6. Apply the classification techniques, namely; LDA classifier, K-nearest neighbour classifier and SVM classifier to classify and recognize the historical Kannada handwritten manuscripts, whether they belong to the Hoysala dynasty or Vijayanagara dynasty or Mysore dynasty?

The detailed approach of the proposed method is given in the Figure 1.

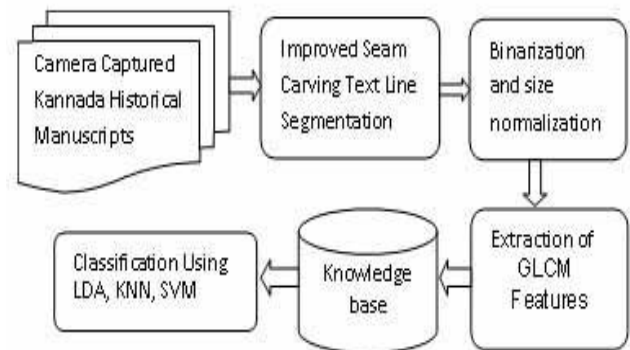


Figure1. The detailed approach of the proposed method

### III. EXPERIMENTAL RESULTS AND DISCUSSION

We have considered the datasets of different dynasties, namely; Vijayanagara (1400 AD and 1460 AD ), Hoysala (1340 AD) and Mysore Wodeyar (1936 AD) (described in Sect. II, Data collection) for experimentation. The experimentation is done with Intel Core i5 system using Matlab R2018b. Input the camera captured historical Kannada handwritten manuscript document images (Figure 2a) for extraction of text line segmentation. To extract text line segmentation, we applied the improved seam carving method which includes the computation of medial seam (Figure 2b), Separating seam computation with constrained seam carving based on medial seam (Figure 2c), overlaid with both type of seam (Figure 2d), region of interest i.e., text line is extracted based on the overlaid seam using roipoly() function (Figure 2e). And then apply the image enhancement method for binarization, restoration and size normalization to each individual text line (Figure 2f). Extracted the GLCM features for all the text lines and store them as a knowledge base. Finally, apply classification techniques; i.e., LDA classifier, K-NN classifier and SVM classifier for classification and recognition of the historical Kannada handwritten manuscripts based on their age-type. The other sample images of the proposed algorithm used for other dynasties namely, Mysore Wodeyar(1936 AD) dynasty, Vijayanagara(1400 AD) dynasty and Hoysala(1340 AD) dynasty, which are shown in Figure 3, Figure 4 and Figure 5, respectively.

The details of the individual text lines extracted from the historical Kannada handwritten manuscript based on their age-type are given in the Table 1. In that Vijayanagar (1460 AD) dynasty has 24 digitized manuscripts and we have 250 individual text lines extracted correctly out of 252; the Mysore Wodeyar(1936 AD) dynasty has 28 digital manuscripts and 224 individual text line are correctly extracted out of 235; the Vijayanagara (1400 AD) dynasty has 39 digital and 552 individual text lines are correctly extracted out of 674. In this dynasty, we have more number of text lines identified wrongly i.e. 122 text lines due to maximum degrades in manuscripts. Similarly, the Hoysala (1340 AD) dynasty has 30 digital manuscripts and 335 individual text lines are extracted out of 336 text lines in the manuscripts. Further, we have done experimentation on correctly segmented and extracted individual text line of various dynasties by eliminating wrongly identified text lines. The reason for wrongly identified text lines is due to poor quality, distortion, ink bleed through etc. which will be taken care of these text lines in near future work.

The average classification accuracy of the proposed method is given in the Table 4. Initially, we have extracted the GLCM features with 24 features using k-fold experiment and these results are given in the Table 2. Further, to improve the results we have reduced the GLCM features with 20 features and certainly results are improved. The results of these reduced features with k-fold experiments using LDA, K-NN and SVM classifiers are given in the Table 3. As per the results, it is observed that GLCM features with 20 features are given better results comparatively 24 features. Hence, we propose only 20 features by reduced features from 24 and overall accuracy is calculated and represented only based on 20 GLCM features. The classification accuracy for different dynasties represents that the LDA classifier has yielded 86.5%, K-NN classifier has 85.3% and SVM classifier has 85.6%. Based on the experimentation, it seems that the LDA classifier has got a good classification performance comparatively K-NN classifier and SVM classifier for historical Kannada handwritten manuscript document images.

The results of the confusion matrix of LDA classifier is given in the Table 5, K-NN classifier is given in Table 6 and SVM classifier is represented in Table 7, which indicates better recognition rates towards historical Kannada handwritten manuscript document images.

**Table 1.**The results of the text lines extracted from the manuscripts of various dynasties based on their age-type

Dynasties	No. of Manuscript	Total No. of Text	Correctly Identified	Unidentified Text lines
-----------	-------------------	-------------------	----------------------	-------------------------

	Images	lines in the Manuscript images	and Segmented text lines	
Vijayanagara (1460)	24	252	250	2
Mysore Wodeyar (1936)	28	235	224	9
Vijayanagara (1400)	39	674	552	122
Hoysala (1340)	30	336	335	1

**Table 2.**Classification accuracy of GLCM with 24 Features for different k-fold experimentation

Classifiers	5 Fold	4 Fold	3 Fold	2 Fold
LDA	85.9	86.1	86.1	85.9
KNN	84.4	85.1	83.5	84.7
SVM	85.6	84.4	84.2	85.1

**Table 3.**Classification accuracy of GLCM with 20 Features for different k-fold experimentation

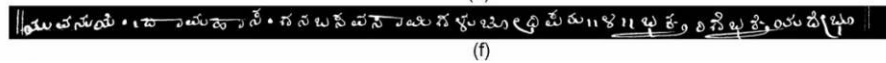
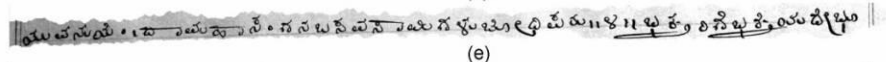
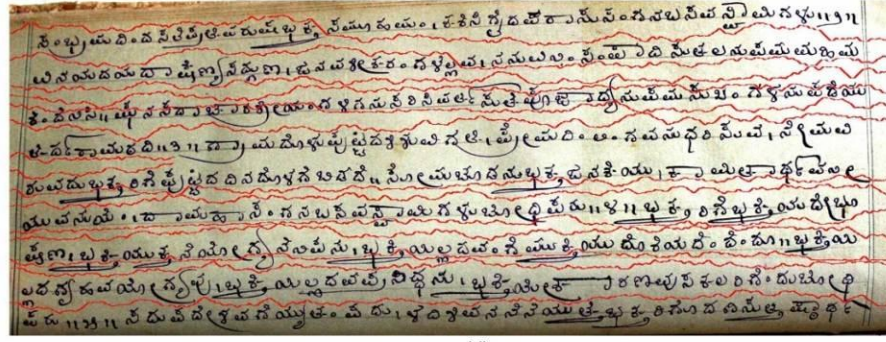
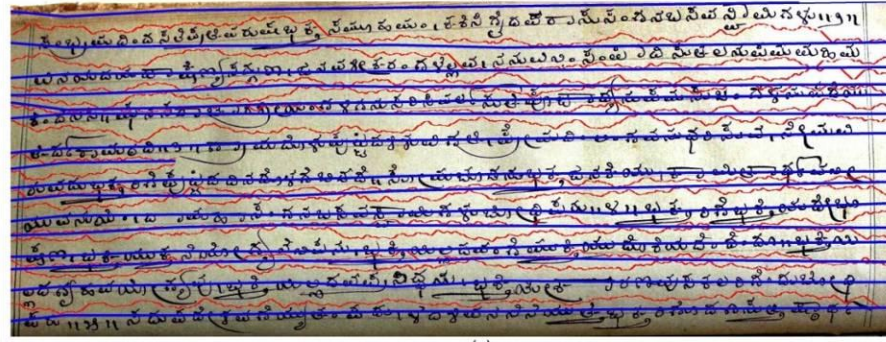
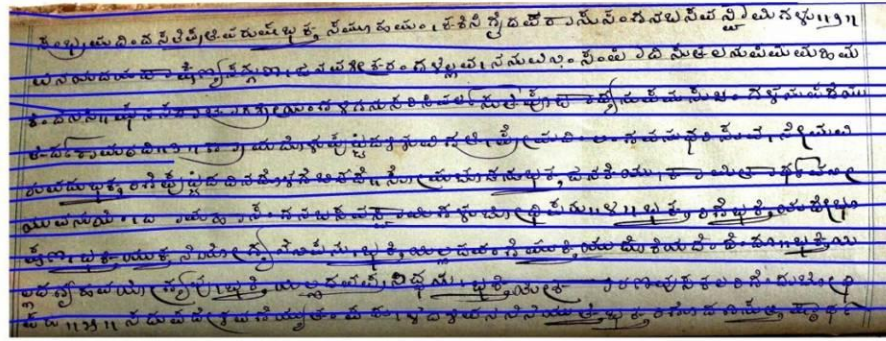
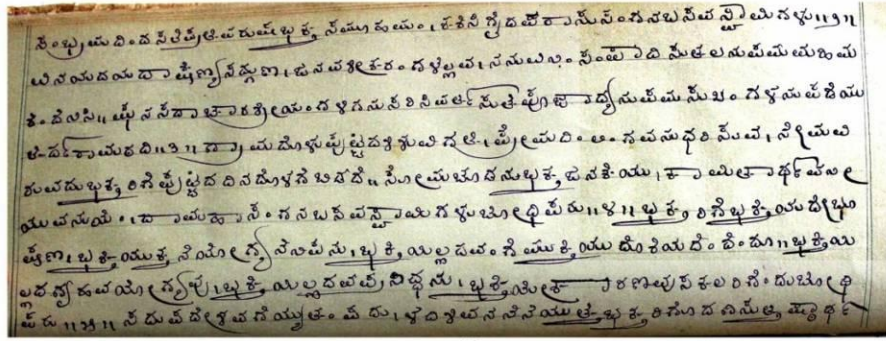
Classifiers	5 Fold	4 Fold	3 Fold	2 Fold
LDA	86.5	86	85.8	85.7
KNN	85.3	85.4	84.7	85.1
SVM	85.6	85.4	85.1	85.4

**Table 4.**The average classification accuracy of proposed method with LDA, K-NN and SVM classifiers

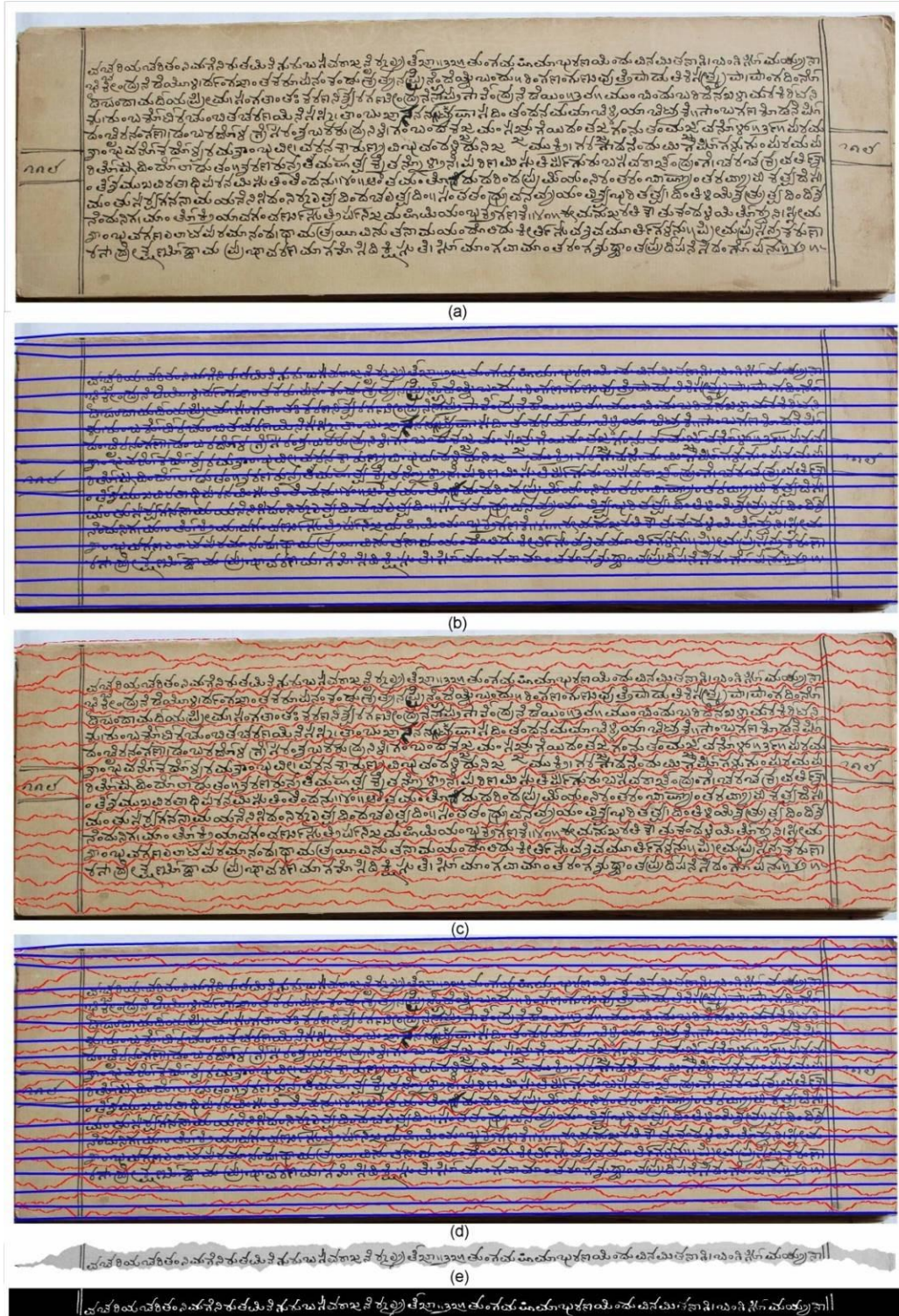
Dynasties	Accuracy (%)					
	LDA		K-NN		SVM	
	Recog. Rate	Error Rate	Recog. Rate	Error Rate	Recog. Rate	Error Rate
Vijayanagara (1460)	89	11	93	7	90	10
Mysore Wodeyar (1936)	61	39	50	50	56	44
Vijayanagara (1400)	99	1	96	4	95	5
Hoysala (1340)	85	15	90	10	91	9
Average accuracy	86.5%		85.3%		85.6%	



Figure 2. Sample image of the proposed algorithm (a) Original camera captured historical Kannada handwritten manuscript document image of Vijayanagara dynasty (1460AD) (b) medial seam computed image (c) Separating seam computation image with constrained seam carving based on medial seam (d) overlaid image with both type of seam (e) text line is extracted based on the overlaid seam (f) Enhanced and size normalized image



**Figure 3.** Sample image of the proposed algorithm (a) Original camera captured historical Kannada handwritten manuscript document image of Mysore wodeyar dynasty (1936AD) (b) medial seam computed image (c) Separating seam computation image with constrained seam carving based on medial seam (d) overlaid image with both type of seam (e) text line is extracted based on the overlaid seam (f) Enhanced and size normalized image



**Figure4.** Sample image of the proposed algorithm (a) Original camera captured historical Kannada handwritten manuscript document image of Vijayanagara dynasty (1400AD) (b) medial seam computed image (c) Separating seam computation image with constrained seam carving based on medial seam (d) overlaid image with both type of seam (e) text line is extracted based on the overlaid seam (f) Enhanced and size normalized image



**Figure5.** Sample image of the proposed algorithm (a) Original camera captured historical Kannada handwritten manuscript document image of Hoysala dynasty (1340AD) (b) medial seam computed image (c) Separating seam computation image with constrained seam carving based on medial seam (d) overlaid image with both type of seam (e) text line is extracted based on the overlaid seam (f) Enhanced and size normalized image

**Table 5.**Confusion matrix for LDA classifier

Dynasties	Vijayanagara (1460)	Mysoure Wodeyar (1936)	Vijayanagara (1400)	Hoysala (1340)	Total No. of Text line
Vijayanagara (1460)	223	26	0	1	250
Mysoure Wodeyar (1936)	34	133	0	52	219
Vijayanagara (1400)	0	0	405	1	406
Hoysala (1340)	2	47	0	285	334

**Table 6.**Confusion matrix for K-NN classifier

Dynasties	Vijayanagara (1460)	Mysoure Wodeyar (1936)	Vijayanagara (1400)	Hoysala (1340)	Total No. of Text line
Vijayanagara (1460)	232	10	6	2	250
Mysoure Wodeyar (1936)	36	109	22	52	219
Vijayanagara (1400)	4	8	389	5	406
Hoysala (1340)	1	27	5	301	334



Table 7. Confusion matrix for SVM classifier

Dynasties	Vijayanagara (1460)	Mysore Wodeyar (1936)	Vijayanagara (1400)	Hoysala (1340)	Total No. of Text line
Vijayanagara (1460)	224	18	7	1	250
Mysore Wodeyar (1936)	24	122	14	59	219
Vijayanagara (1400)	0	16	385	5	406
Hoysala (1340)	0	28	2	304	334

#### IV. CONCLUSION

In this paper, we have proposed an algorithm to identify and recognise the historical Kannada handwritten manuscript documents based on their age-type by using the improved seam carving text line segmentation method with GLCM features and LDA, K-NN, SVM classifiers are used to classify and recognize the script. The average classification accuracy for different dynasties is computed. The LDA classifier has yielded 86.5%, K-NN classifier has yielded 85.3% and SVM classifier has 85.6%. Based on the experimentation, the LDA classifier has recorded good classification performance comparatively K-NN and SVM classifiers for historical Kannada handwritten scripts with GLCM features. The experimental outcomes are tested with manual results, which show the robustness of the proposed method. The same algorithm may be used for other dynasties of historical Kannada handwritten manuscripts with different feature sets, which will be done in the future work.

#### ACKNOWLEDGMENT

The authors are grateful to The Chairman, Department of P. G. Studies and Research in Kannada, Gulbarga University, Kalaburgi and Dept. of Hastapatri, Kannada University, Hampi for providing the historical Kannada handwritten manuscripts and visualization of manual results.

#### REFERENCES

- [1] Manjunath, M.G., Devarajaswamy G.K., "Kannada Lipiya Vikasa", Published by Jagadhguru Sri Madhvacharya Trust, Sri Raghavendra Swami Matta, Mantralaya
- [2] Narasimha Murthy, A.V., "Kannada Lipiya Ugama Mattu Vikasa", Kannada Adhyayana Samsthe, Mysore University, Mysore (1968)
- [3] Reddy D., "Lipiya HuttuMattu Belavanige—Origin and Evolution of Script", Kannada Pustaka Pradhikara (Kannada Book Authority), Bangalore
- [4] Nikolaos Arvanitopoulos and Sabine Susstrunk, "Seam Carving for Text Line Extraction on Color and Grayscale Historical Manuscripts" 14th IEEE International Conference on Frontiers in Handwriting Recognition, DOI: 10.1109/ICFHR.2014.127, pp.726-731, 2014.
- [5] S. Avidan, A. Shamir, "Seam Carving for Content-Aware Image Resizing", ACM Transactions on Graphics, vol. 26, no. 3, p. 10, 2007.
- [6] Veronica Romero, Joan Andreu Sanchez, Vicente Bosch, Katrien Depuydt and Jesse de Does, "Influence of Text Line Segmentation in Handwritten Text Recognition" 13th IEEE International Conference on Document Analysis and Recognition (ICDAR), 978-1-4799-1805-8/15, pp.536-540, 2015.
- [7] Maroua Mehri, Petra Gomez-Kramer, Pierre Héroux, Alain Boucher and Rémy Mullot, "Texture feature evaluation for segmentation of historical document images", 2nd International Workshop on Historical Document Imaging and Processing, 2013, Washington, United States, pp.102-109, 2013.
- [8] Mehri M., Héroux P., Gomez-Krämer and Remy Mullot, "Texture feature benchmarking and evaluation for historical document image analysis" IJDAR 20: 1. DOI: 10.1007/s10032-016-0278-y, pp. 1-35, 2017.
- [9] Darko Brodić, "Script Recognition by Statistical Analysis of the Image Texture", X International Symposium on Industrial Electronics INDEL 2014, Banja Luka, November 06 to 08, 2014, pp.168-174.
- [10] Berat Kurar Barakat and Jihad El-Sana, "Binarization Free Layout Analysis for Arabic Historical Documents Using Fully Convolutional Networks", IEEE 2nd International Workshop on Arabic and Derived Script Analysis and Recognition, {ASAR}, London, UK, March 12-14, 2018, pp.151-155, 2018.
- [11] Laurence Likforman-Sulem, Abderrazak Zahour and Bruno Taconet, "Text line segmentation of historical documents: a survey" IJDAR (2007) pp:123-138, DOI 10.1007/s10032-006-0023-z
- [12] Abedelkadir Asi, Raid Saabni and Jihad El-Sana, "Text Line Segmentation for Gray Scale Historical Document Images", HIP '11 Proceedings of the 2011 Workshop on Historical Document Imaging and Processing, DOI: 10.1145/2037342.2037362, pp. 120-126, 2011.
- [13] Parashuram Bannigidad and Chandrashekar Gudada, "Restoration of Degraded Historical Kannada Handwritten Document Images using Image Enhancement Techniques", International Conference on Soft Computing and Pattern Recognition (SoCPaR 2016), 2016, pp. 498-508.
- [14] Parashuram Bannigidad and Chandrashekar Gudada, "Restoration of Degraded Kannada Handwritten Paper Inscriptions (Hastapatri) using Image Enhancement Techniques", IEEE International Conference on Computer Communication and Informatics (ICCCI -2017), 2017, pp.1-6.
- [15] Parashuram Bannigidad and Chandrashekar Gudada, "Identification and Recognition of Historical Kannada Handwritten Document Images Using GLCM Features", International Journal of Advanced Research in Computer Science, Vol.9, No.1, 2018, pp.686-690.

- [16] Robert M Haralick, K Shanmugam and Its'hak Dinstein, "Textural Features for Image Classification", IEEE Transactions on Systems, Man, and Cybernetics. SMC-3 (6), 1973, pp. 610-621.

### Authors Profile

---

*Dr. Parashuram Bannigidad*, Assistant Professor, Department of Computer Science, Rani Channamma University, Belagavi, Karnataka, India. He has specialized in Image Processing and Pattern Recognition, Medical and Bio-Medical, Document Image Analysis and Nanotechnology. He has published more than 50 research papers in international journals and Conference proceedings.



He has been awarded as Young Scientist from VGST, Dept. of IT, BT and DST, Govt. of Karnataka. He has Completed 04 research projects funded by RCUB, VGST, Govt. of Karnataka and UGC, New Delhi, INDIA.

*Mr. Chandrashekar Gudada* received the B.Sc. and M.Sc. degrees in Computer Science from Gulbarga University, Kalaburgi, India, in 2009 and 2012 respectively, and presently pursuing Ph.D. Degree in computer science from Rani Channamma University, Belgavi, India. Specialized in Image Processing and Pattern Recognition and Document Image Analysis

---

