



# A Comprehensive study on Historical palm leaf manuscript preservation, digitalization and image enhancement

Mahaveer <sup>1</sup>, and Basavanna M<sup>1</sup>

<sup>1</sup>Department of Studies in Computer Science, Davangere University, Davanagere, Karnataka, India,

**Abstract :** In the past, knowledge and information were saved by writing on stone, wood, metal, and leaves. However, most information is recorded in the form of palm leaf handwriting since the leaves of palm tree are easily obtained. The preservation of manuscripts related to ancient civilization is a highly challenging task because of drawbacks such as stains, noise, lack of standard alphabets, degradation, and other issues. A proposition is made for preserving these manuscripts, as well as traditional image enhancement methodologies in this paper.

**Keywords:** Ancient documents, Palm leaf, Image enhancement, Generative Adversarial Networks, Discriminator, Generator, Tamil, Deep Learning, Convolution Neural Networks

## I. INTRODUCTION

India is well known for having a vast knowledge base and rich cultural past. Throughout history, knowledge and information have been preserved in a vast array of ways, including writing on wood, metal, stone, and leaf surfaces. However, because of palm leaf's simple availability (1-3), the vast majority of knowledge and information from earlier times has been preserved as handwriting on palm leaf. In addition, palm leaf has an average lifespan of 300–400 years. Even for the present generation, the foundation for sophisticated advancements is still the leaf of the palm writings on subjects like astronomy, literature, architecture, and ayurveda.



Fig 1: Manuscript of a plam leaf

The most challenging task in digital image processing is text recognition from historical texts. The primary stage is to enhance the clarity, quality, and visibility of the digital images because the ancient Palm Leaf Manuscripts suffer from a wide range of issues, including deterioration, a lack of standard alphabets, stains, noise, thick and arbitrary layout, low print contract, and different types of typeface present.

**Digitalization:** The process of converting a palm leaf text from its physical form into a digital version is known as digitalization. This can be done by taking a photograph of the manuscript in its physical form or by scanning the document.

The technique of extracting text from handwritten digital document files utilizing a multifariousness of technologies is known as the Digitalized Palm Leaf Images Documents Analysis System. The challenge is identifying the Indian Classical Languages Script (kannada, telugu, tamil, and so on), notwithstanding many proposed approaches in the literary texts on popular languages and writings from developed regions such as American and European Countries focus on character and text recognition.

### **Indian Classical Languages Palm Leaf Manuscripts**

There are many native tongues spoken in India, majority of the languages being from the Indo-Aryan and Dravidian groups. The official and frequent language spoken in the nation is Hindi. Other widely spoken languages are Tamil, Telugu, Bengali, and Marathi. Due to their rich history and independence, India's constitution acknowledges six other languages in addition to the 22 scheduled languages that have official status. These native tongues are frequently referred as classical languages. A language has to be original and undiluted from any other culture, have a strong historical legacy, and have records that go back up to 1500–2000 years in order to qualify as classical.

#### **Kannada**

The majority of kannadigas converse kannada, a Dravidian tongue, are found in Karnataka in southwest India. Along with Kannadigas living overseas linguistic minorities in the states of Maharashtra, Andhra Pradesh, Tamil Nadu, Telangana, Kerala, and Goa also converse in kannada. By 2011, there were about 43 million kannadigas. Over 12.9 million non-native speakers of Kannada speak it as a second or third language in Karnataka, making a total of 56.9 million speakers.

#### **Telugu**

A large number of Telugu speakers are found in Andhra Pradesh and Telangana, where Telugu is the official language. Like kannada Telugu is also a Dravidian language. In Odisha, Karnataka, Tamil Nadu, Kerala, Punjab, Chhattisgarh, Maharashtra, and the Andaman & Nicobar Islands, Telugu is a less frequently conversed language. It is also an official language in West Bengal and the Yanam district of Puducherry. There are 6 languages recognized as classical languages and Telugu is also a part of it.



Fig 3: Palm Leaf Manuscript in Telugu

Since majority of historical information of India available in Dravidian languages written on palm leaf, so it's a major concern to enhance and extract the information.

## **2. Literature Survey**

In this section, the various contribution made by different authors has been revealed.

When it comes to Palm Leaf script identification, there are several noticeable variations. Numerous techniques for identifying characters from manuscripts, written documents, and inscriptions have been developed by researchers. While there are methods for enhancing pictures, character identification, and recognition from historical records, Indus documents, and inscriptions in the referenced papers [8–12], there aren't any for recognizing writing on palm leaves.

The cited studies [13–18] do not provide any strategies for recognizing and recognising the palm leaf text. Rather, techniques for character recognition, identification, and preprocessing images from old handwritten paper texts and deteriorating palm leaf manuscripts have been suggested.

Dona Valy et al. [8] provided methods for two historical document analysis tasks using digitized Khmer palm leaf manuscripts. Dewa The study conducted by Made Sri Arsa et al. [9] focuses on the identification of individual handwritten Balinese characters from palm leaf manuscripts. A suitable and trustworthy combination of feature extraction strategies was proposed by Made Antara Kesiman et al. [10] in order to increase the recognition rate of Balinese text.

R. S. Sabeenia et al. [11] showed a CNN based approach for character identification, with a focus on Tamil palm-leaf characters especially.

Sunny Singh is one of the others [12] The first part of this paper offers a thorough synopsis of a review of many research studies. The second section offers a number of feature extraction and classification techniques. Narahari Sastry Panyama, et al., [13], The depth of indentation, a special three-dimensional feature exploited in this study, is associated with the pressure applied by the scribe at that particular place. This 3D property is present in every pixel point of a Telugu palm leaf character.

Dr. I. Manimozhi et al. This study presents a software that can accurately read handwritten Tulu characters and generate Kannada characters as output.

Although the listed papers [15–18] have offered approaches based on line segmentation and text identification in historical and palm leaf literature, no techniques for Kannada/Telugu palm leaf text recognition have been proposed.

The survey specifies that there is currently no accurate methodology for identifying and classifying palm leaf manuscript texts of Indian classical languages, particularly Telugu and Kannada. This is an open challenge. This thereby created a path for recognising Indian classical languages.

### 3. Challenges of Palm Leaf Manuscripts

1. The palm leaf manuscript's physical state presents the first difficulty because it has a significant impact on the standard of the document photos that are taken. When capturing images for digital image data study, optical scanning is typically utilized to acquire data from paper documents. However, photographic methods are frequently employed when taking images from documents on alternative media, such as microfilm, fabric, or palm leaves. These days, the improvement in image processing and enhancement techniques for palm leaf manuscripts to extract pertinent information is seen to be a new research topic in handwritten document analysis because of the unique qualities of the physical support of the manuscripts.
2. The second difficulty is identifying the text; the palm leaf manuscript's handwriting is hard to detect because of the distinct style.  
Diversity: A wide range of scripts, including Manipravalam, Telugu, Tamil, Malayalam, Grantha, and Modi, are utilized in the palm-leaf manuscripts. Other packages, like Telugu and Malayalam combined with Grantha, come with numerous scripts.  
Languages: A number of languages, including Pali, Tamil, Telugu, Sanskrit, and Medieval; forms and shapes of palm leaves; different kinds of leaves, including birch bark, etc.  
Time period: The ancient literatures come from a variety of historical periods, which broadens the diversity immensely.
3. Arrangement: There is no numbering on the leaves of the manuscript. A thread connects them, and if this thread breaks, the sequence is damaged. The only thing that can be used to rearrange them is the contents. Numerous times, pens have been used by library staff to write numbers directly onto the original palm leaves, altering their natural form.
4. Resource Persons: There is a deficiency in the knowledge about the numerous facets of manuscripts, such as how to handle them, digitize them, and comprehend the different scripts and languages needed to catalog them.

#### 4. Various Historical Documents Digital Images Quality Improvements Methodologies Below are the various traditional image quality improvement techniques

##### 4.1 Image Enhancement

Technology for image processing that provide a choice of approaches to improve image quality and make them more readable.

In the traditional method, they were split into three groups, which are as follows:

- (a) Thresholding/binarization
- (b) Thresholding
- (c) Approaches not reliant on thresholds

##### 4.2 Image Segmentation

- (a) Background Elimination: Otsu Formula separates text from extraneous noise.
- (b) Line segmentation, which is the division of lines into segments.
- (c) Character segmentation assesses the image's histogram.

We analyze and find the best method of palm leaf manuscript preservation and various enhancement methodologies.

Second section stands above literature survey on various palm leaf manuscript preservation, image enhancement & Text recognition technique.



In section three, we can see the traditional method of image enhancement techniques. Section four machine learning based methods. Conclusion is framed at the end.

## CONCLUSION:

Ancient information mostly comes from manuscripts on palm leaves. To gather and disseminate the knowledge contained in the manuscripts and make them easily available to domain experts and the general public, a comprehensive digitization process that considers the unique qualities of palm-leaf manuscripts is required. To do this, a range of instruments and technologies—including those pertaining to the challenges involved in the identification, recognition, and improvement of palm leaf manuscripts—need to be created.

## REFERENCES:

- [1] "Institute Of Asian Studies". Instituteofasianstudies.com. N.p., 2016. Web.
- [2] "National Mission For Manuscripts". Namami.org. N.p., 2016. Web. 11 Oct.
- [3] "Kritisampada: The National Database Of Manuscripts". Namami.org. N.p., 2016. Web. 11 Oct. 2016.
- [4] Kumar, Neethu S., Dinesh S. Kumar, S. Swathikiran, and Alex Pappachen James. "Ancient indian document analysis using cognitive memory network." Presented In IEEE International Conference on Advances in Computing, Communications and Informatics, ICACCI 2014, pp. 2665-2668
- [5] Shi, Zhixin, and Venu Govindaraju. "Historical document image segmentation using background light intensity normalization." In Proc. SPIE, vol. 5676, 2005, pp. 167-174.
- [6] Cherala, Sridhar, and Priti Rege. "Palm leaf manuscript/color document image enhancement by using improved adaptive binarization method." presented in Sixth Indian Conference on Computer Vision, Graphics & Image Processing, ICVGIP, 2008.
- [7] Chamchong, Rapeeporn, and Chun Che Fung. "Character segmentation from ancient palm leaf manuscripts in Thailand In "Proceedings of the Workshop on Historical Document Imaging and Processing", ACM, 2011.
- [8] Dona Valya,b, Michel Verleysena, Sophea Chhunb, and Jean-Christophe Buriec, **Character and Text Recognition of Khmer Historical Palm Leaf Manuscripts**, 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR).
- [9] Dewa Made Sri Arsa, Gusti Agung Ayu Putri, Remmy Zen and Stephane Bressan, **Isolated Handwritten Balinese Character Recognition from Palm Leaf Manuscripts with Residual Convolutional Neural Networks**, The 2020 12<sup>th</sup> International Conference on Knowledge and Systems Engineering(KSE).
- [10] Made Windu Antara Kesiman, Sophea Prum, Jean-Christophe Burie1, Jean-Marc Ogier, **Study on Feature Extraction Methods for Character Recognition of Balinese Script on Palm Leaf Manuscript Images**, 2016 23rd International Conference on Pattern Recognition (ICPR)Cancún Center, Cancún, México, December 4-8, 2016.
- [11] R. S. Sabeenian, M. E. Paramasivam, R. Anand and P. M. Dinesh, **Palm-Leaf Manuscript Character Recognition and Classification Using Convolutional Neural Networks** © Springer Nature Singapore Pte Ltd. 2019.
- [12] Sunny Singh a, Pradeepta Kumar Sarangi a, Chaitanya Singla a, Ashok Kumar Sahoo, **Odia character recognition system: A study on feature extraction and classification techniques**, S. Singh et al. / Materials Today: Proceedings 34 (2021) 742–747.
- [13] Narahari Sastry Panyama,\* , Vijaya Lakshmi T.R. b, RamaKrishnan Krishnanc, Koteswara Rao N.V, **Modeling of palm leaf character recognition system using transform based techniques**, N.S. Panyam et al. / Pattern Recognition Letters 84 (2016) 29–34.
- [14] Dr.I.Manimozhi Dr. Manoj challa, **An Efficient Translation of Tulu to Kannada South Indian Scripts using Optical Character Recognition**, Proceedings of the Fifth International Conference on Computing Methodologies and Communication (ICCMC 2021) IEEE Xplore Part Number: CFP21K25-ART.

- [15] Suman Kumar Bera a,\* , Soumyadeep Kundu a , Neeraj Kumar b,c,d , Ram Sarkar a , **Distance transform based text-line extraction from unconstrained handwritten document images**, Expert Systems With Applications 186 (2021) 115666.
- [16] R. Spurgen Ratheash1\* , M. Mohamed Sathik2, **A Detailed Survey of Text Line Segmentation Methods in Handwritten Historical Documents and Palm Leaf Manuscripts**, Vol. 7(8), Apr 2019, E-ISSN: 2347-2693.
- [17] Reya Sharma, Baijnath Kaushik, **Offline recognition of handwritten Indic scripts: A state-of-the-art ,survey and future perspectives**, . Sharma and B. Kaushik / Computer Science Review 38 (2020) 100302.
- [18] Zhuo Chena,b,\* , Fei Yina,b , Xu-Yao Zhanga,b , Qing Yanga,b , Cheng-Lin Liua,b,c, **MuLTReNets: Multilingual text recognition networks for simultaneous script identification and handwriting recognition**, Z. Chen, F. Yin and X.-Y. Zhang et al. / Pattern Recognition 108 (2020) 107555.
- [19] Manigandan T ,Dr. V.Vidhya et.al ,Tamil Character Recognition from Ancient Epigraphical Inscription using OCR and NLP, International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS-2017).
- [20] R. Kiruthika,Mrs. S. Kalyani, Multispectral Image Enhancement of Historical Document Images, International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Published by, www.ijert.org NCSPC - 2016 Conference Proceedings.
- [21] A. S. Kavitha, P. Shivakumara, G. Hemantha Kumar, Skewness and Nearest Neighbour based Approach for Historical Document Classification, 2013 International Conference on Communication Systems and Network Technologies.
- [22] Xusheng Liang a, Abbas Cheddad , Johan Hall, Comparative Study of Layout Analysis of Tabulated Historical Documents , 2214-5796/© 2021 The Authors. Published by Elsevier Inc.
- [23] Huseyin Kusetogullari a,\* , Amir Yavariabdi b, Johan Hall c, Niklas Lavesson , DIGITNET: A Deep Handwritten Digit Detection and Recognition Method Using a New Historical Handwritten Digit Dataset, 2214-5796/© 2020 The Author(s). Published by Elsevier Inc.
- [24] Aladhahalli Shivegowda Kavitha a , Palaiahnakote Shivakumara b,† , Govindaraj Hemantha Kumar a , Tong Lu, A new watershed model based system for character segmentation in degraded text lines, .S. Kavitha et al. / Int. J. Electron. Commun. (AEÜ) 71 (2017) 45–52.
- [25] YOUNES AKBARI , SOMAYA AL-MAADEED , (Senior Member, IEEE), AND KALTHOUM ADAM, (Member, IEEE), Binarization of Degraded Document Images Using Convolutional Neural Networks and Wavelet-Based Multichannel Images.
- [26] Alaa Sulaiman \* , Khairuddin Omar and Mohammad F. Nasrudin, Degraded Historical Document Binarization: A Review on Issues, Challenges, Techniques, and Future Directions.
- [27] Radhika Patel and Suman K. Mitra, Extracting Text from Degraded Document Image.
- [28] A.S. Kavitha a , P. Shivakumara b ,\*, G.H. Kumar a , Tong Lu, Text segmentation in degraded historical document images.
- [29] Parashuram Bannigidad, Chandrashekar Gudada, Restoration of Degraded Kannada Handwritten Paper Inscriptions (Hastapatti) using Image Enhancement Techniques.
- [30] MARTIN ESAKSSON , FIVE GANs FOR BETTER IMAGE PROCESSING,
- [31] HANDWRITTEN ANCIENT TAMIL CHARACTER RECOGNITION USING GENERATIVE ADVERSARIAL NETWORK, N SASIPRIYA
- [32] Patel, C.I., Patel, R. and Patel, P., 2011. Handwritten character recognition using neural network. *International Journal of Scientific & Engineering Research*, 2(5),
- [33] Patil, V. and Shimpi, S., 2011. Handwritten English character recognition using neural network. *Elixir Comput Sci Eng*, 41, pp.5587-5591.
- [34] Denker, J.S., Henderson, D., Howard, R.E. and Hubbard, W., 1990. Handwritten character recognition using neural network architectures. In *Proceedings of the 4th USPS Advanced Technology Conference, Washington DC* (pp. 1003-101).
- [35] Obaid, A.M., El Bakry, H.M., Eldosuky, M.A. and Shehab, A.I., 2016. Handwritten text recognition system based on neural network. *Int. J. Adv. Res. Comput. Sci. Technol. (IJARCST)*, 4(1), pp.72-77.