

# An Analysis of Challenges in Character Recognition on Palm Leaf Manuscripts, and Their Solutions



**Nagendra Panini Challa, Lakshmi Kalyani, Vasanth Kumar Mehta, Lakshmi Mounika**

**Abstract:** Information was recorded and transmitted through the use of many traditional methods in ancient times using naturally available material like palm leaves, tree barks etc. which were found suitable for long-term storage. With passage of time, these materials are increasingly prone to deterioration from many internal and external agents, and we are at a risk of losing the valuable information inscribed on them. Digitization is a feasible and available solution preservation of such content. Scanning and digitizing these manuscripts facilitates making it visible and readable for the users. The process of Knowledge retrieval from these manuscripts begins with the first step of Character Recognition. This primary step itself is a very complex task due to the variety of scripts and handwritings on the palm leaf manuscripts. This paper provides an overview about different challenges faced by global researchers pertaining to palm leaf scripts and the solutions proposed by them.

**Keywords :** Digitization, Hand-Written Text, Optical Character Recognition, Palm Leaf Manuscripts.

## I. INTRODUCTION

Knowledge from different scholars was preserved in both oral and written forms for future generations[4]. India passed through the ages of speaking, drawing, and painting. By series of natural steps, writing grew out of drawing. With the increase in the use of writing, different types of writing materials such as leaves, barks, stones, metals wood, etc., were discovered. Of all the available different types of materials, palm leaves were widely used due to its enormous availability and capability to withstand rigorous conditioning.

Palm leaf manuscripts are an invaluable part of India's documented heritage. Palm leaf manuscripts are scattered all over the country in temples, libraries, museums, etc., in different Indian languages and written scripts[4]. Despite the conditioning capability, different processing, preservation, and conservation techniques of palm leaves, palm leaves have an average life of 300-400 years. Many of the palm leaf manuscripts that were written 3 centuries ago are therefore in a stage of decay and we stare at the possibility of losing the valuable content recorded on them.

These priceless manuscripts give deep insight into Literature, Astronomy, Ayurveda, Architecture, Fine Arts written in different languages and scripts of various periods. In ancient days, the only way to restore and transfer this valuable information to future generations was rewriting the information again. With the advent of Technology, digitization is being used to create digital images of the manuscripts for preserving their content. The species of palm tree which are used for writing were brought to India from Africa. The species are the Palmyra palm (*Borassus flabellifer* Linn) which is named as 'Tala', and the second one is Fan palm (*Corypha umbraculifera* Linn). It is also called as 'Sritala' or 'Talipot' [3]. Writing on palm leaf manuscripts was done in two ways. One way was to engrave on the leaves with a pointed metal 'stylus' or 'salaka' - a pointed tip was used which is made of iron, steel, brass, bone etc. While writing for the movement of the tip left hand thumb support is required. For clarity of letters engraved leaves were treated with turmeric powder and a mixture of vegetable juice with lamp shoot which is mixed with some volatile oil like citronella oil for clarity of letters. The other way was to write with ink on leaves using 'bamboo pen' or 'brush' [28]. In general, the length of leaves varied from 15cm-60cm and width between 3cm-12cm. Before writing on palm leaves the leaves were processed and made ready for writing.

Manuscript received on February 10, 2020.

Revised Manuscript received on February 20, 2020.

Manuscript published on March 30, 2020.

\* Correspondence Author

**Dr. Nagendra Panini Challa**, Assistant Professor, Department of IT, Sri Vishnu Engineering College for Women, Bhimavaram-534202, Andhra Pradesh, Email: [cnpanini@svcew.edu.in](mailto:cnpanini@svcew.edu.in)

**Ms. Adurthi Lakshmi Kalyani**, Undergraduate Student, Department of CSE, SCSVMV, Kanchipuram, Tamil Nadu, India. Email: [kalyaniadurthi99@gmail.com](mailto:kalyaniadurthi99@gmail.com)

**Dr. E. Vasanth Kumar Mehta**, Associate Professor, Department of CSE, SCSVMV, Kanchipuram, Tamil Nadu, India. Email: [vasanth.mehta@kanchiuniv.ac.in](mailto:vasanth.mehta@kanchiuniv.ac.in)

**Ms. Aalavala Lakshmi Mounika**, Undergraduate Student, Department of CSE, SCSVMV, Kanchipuram, Tamil Nadu, India. Email: [sweetyreddy59@gmail.com](mailto:sweetyreddy59@gmail.com)

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)





Fig.1 Types of metal stylus used for engraving [1]



Fig.2 Palm leaf written using pointed metal stylus [1]



Fig.3 Palm-leaf written using ink [1]

Palm leaf manuscripts will be processed in different ways [28]. Steps involved in processing palm leaves are:

1. Palm leaves should be processed in special manner to make them fit for writing.
2. Palm leaves should be cut from the trees before they become dry up and brittle.
3. Only half opened young shoot of palm leaves are suitable for making manuscripts.
4. The different methods used for processing palm leaves are:
  - a. Palm leaves were dried, soaked in water, again the leaves should be dried and polished with stones and cut to proper size.
  - b. Palm leaves were dried in shade and gingili oil was applied for smoothing the leaf.
  - c. Palm leaves were hung in kitchen where smoke emits, leaves should be cleaned and turmeric paste will be applied and used for writing.
5. The unwanted middle ridge is removed from main leaf. The desired portion is then trimmed to size. Holes were made on either side of leaves with a red hot wire. Thread is then passed through the holes to hold the leaves together.
6. Two wooden planks which measures same as leaf dimension will be placed above and below the leaf bundle to protect the leaves as shown in Fig-4. All the bundles were kept together and will be covered with a red cloth which is significant that is high repellent for insects and protects from dust [27].



Fig.4 Wooden Plank with holds which is used to bind the manuscripts [1]

Despite the care taken for processing and storing palm leaves, these leaves may get deteriorated because of their organic nature. There are various factors for deterioration of palm leaves[27]. Of all the various factors the most important factors are:

I. Physical Deterioration Factors

- a. With time leaves get physically weak and brittle.
- b. The main reasons for physical deterioration are light, heat, moisture, humidity, mechanical damage by humans.
- c. These Physical changes may bring Oxidative changes in palm leaves.

II. Chemical Deterioration Factors

- a. It is very hard to categorize between physical and chemical factors of deterioration.
- b. Different reasons for chemical deterioration of palm leaves include Acidity, Dis-colouration of palm leaves, Action of inks.

III. Biological Deterioration Factors

- a. Due to the fault in storage of palm leaf manuscripts biological deterioration factors takes place.
- b. The main factors of biological deterioration are Fungi, Insects like silverfish, cockroaches, etc., and rodents[27].

As these palm leaf manuscripts are the invaluable property which we got from our ancestors, it is our responsibility to preserve them from being deteriorated. The different techniques used for preserving palm leaf manuscripts [28] are listed below:

1. Preservation techniques are designed to minimize the physical and chemical deterioration of manuscripts to prevent them from loss of informational content.
  1. Preservation encompasses the activities which prolong the usable life of archival records.
  2. Each region has their own ways to preserve palm leaves. Few methods include
    - a. The use of natural herbs like sweet flag or margosa leaves with the manuscripts to keep insects away.
    - b. Application of citronella oil, camphor oil, lemon grass oil on the surface of the leaves to keep it flexible. This prevents physical damage due to brittleness.
    - c. Fumigation with thymol vapors helps to prevent fungus.
    - d. Fading of ink is restored by applying carbon black mixed with oil to the leaf. As we all know that “Prevention is better than Cure”, it is better to conserve the manuscripts rather than preserving them after deterioration.

1. Conservation is to stabilize an object and to maintain it as close as possible to its original condition.
2. Conservation is the physical and chemical treatment of material to retard future deteriorate.

India possess more than 5 million manuscripts. Even though our ancestors tried preserving these manuscripts, many of the manuscripts were fragmented in different countries and some are no longer accessible to anyone. The invaluable and unique pools of knowledge are under threat and manuscripts are disappearing at an alarming rate. Increase in number of users and their different needs have made the manuscripts libraries apply a new technology. As a result of this there is a worldwide need and trend to plan, collect, organize, manage, protect and distribute information in digital form.

#### A. DIGITIZATION:

Digitization is defined as conversion of analog format for purpose of extending access and where appropriate to assist with preservation [2]. It can't be seen in isolation, it is linked to all aspects like acquiring, converting, storing, retaining information in standardized and organized manner with technology support. Digitization process basically involves making the "electronic photograph" of the required document. The scanned image of the document or image will be converted to digital form and can be stored electronically and accessed via computer. For creating high quality copy of original item Digital image technology can be used.

Optical character recognition (OCR) programs allow scanning a document and converting it to a word processing document in three steps - Image Acquisition and pre-processing, feature extraction and classification. First step cleans up and enhances the quality of the image by noise removal, binarization, color adjustment and text segmentation. Second step extract and capture data from the acquired text image which can be used for classification. Final step, the part of the segmented text in the document image is mapped towards the equal textual image [22]. Even though OCR technology was originally designed to recognize printed text, it can be used to recognize and verify handwritten text as well. It is an offline process i.e, the recognition starts after writing or printing has been completed. Character recognition system extensively uses the methodologies of pattern recognition which allots an unknown sample to a predefined class. Many technologies for character recognition are available and are classified as classical techniques and soft computing techniques.

1. Classical Techniques:
  - a. Template Matching
  - b. Statistical Techniques
  - c. Structural Techniques
2. Soft Computing Techniques:
  - a. Artificial Neural Networks
  - b. Fuzzy Logic Technique
  - c. Evolutionary Computing Techniques [22].

There are many reasons for digitizing palm-leaf manuscripts. Some of the reasons are:

1. Digital materials occupy less space and large volumes can be stored.
2. Users are more convenient in in digital form of documents.

3. Digital documents are easy to access.
4. Digital documents are easy to access via internet from anywhere and anytime [2].

#### II. LITERATURE SURVEY

The main objective of this paper is about the entire process involved in palm-leaf manuscripts - Gathering of leaves, Processing of leaves, Storing of manuscripts, factors for deterioration of palm leaves, various preservation techniques, need for digitization and various techniques for digitization of palm leaf manuscripts. Many researchers mentioned about the enormous information related to Palm-Leaf manuscripts - importance, necessity of these manuscripts for future generations, need and techniques used for digitization, advantages and disadvantages of digitization of manuscripts and many other thing related to Palm-leaf manuscripts. As the technology is improving day-by-day the techniques for preserving and digitization of handwritten and printed text palm-leaf manuscripts are also attaining more efficiency, accuracy with improvement in Algorithms.

**Narenthiran R, Saravanan G, Ramanujam K**[2] mentioned about the factors for deterioration of palm leaves manuscripts, need for digitization and various techniques for digitization of palm leaf manuscripts and different ways for storing the digitized information. They also mentioned various advantages and disadvantages of digitization of palm leaf manuscripts. Stressed that "Digitization is necessary to give this invaluable wealth to our future generations".

**Mohamed Sageer T.K & Dr.A.T.Francis**[3] stated that "Due to the absence of a proper solution for the bulk storage of data, over the past years many palm leaf manuscripts were decayed even without retaining a copy or image for future use". Mentioned that if there was a technology to store the information in bulk, it would have save crores of valuable documents and data and the future generations would have benefited. Stresses that compared to natural and unavoidable threats on the life of palm leaves, the only remedy for preserving the data on palm leaves is to convert them into digital format as it ensure global access from the users premise and facility for providing translation and transliteration along with original documents. Also surveyed that 85% of the Government Institutions in Kerala are ready for the digitization process provided they are equipped with sufficient infrastructure and trained staff.

**Deepti Ganapathy**[4] mentioned about the lack of awareness among the people about the Preservation, Conservation and Digitization of palm leaf manuscripts for future generations. Informed about the "KRITISAMPADA" which was launched by "National Mission for Manuscripts(NMM)" on 14<sup>th</sup> February of 2007, which contains more than a million Indian Manuscripts. Also informed that NMM works with 57 Manuscript Resource Centers, 34 Manuscript Conservation centers with a team of trained conservators with a laboratory equipped to undertake manuscript conservation and 34 Manuscript partner centers across the country. Stated that "The life of a palm leaf manuscript is far longer than a modern-day device like CD or microfilm", but the increasing popularity of printed books reduced the interest for Collecting, Processing and Preservation of palm leaves.

# An Analysis of Challenges in Character Recognition on Palm Leaf Manuscripts, and Their Solutions

Finally, highlights that “Unless the present generation is drawn into supporting the heritage structures that protect these collections of manuscripts, the mission to digitize them will end up on the shelf”.

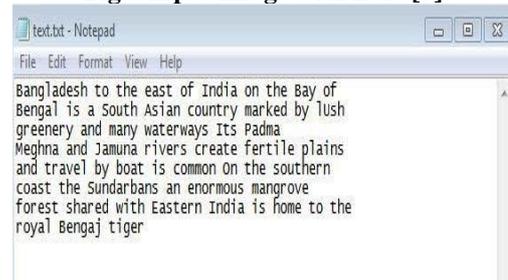
**Jabir Ali V, Jose T Joseph[5]** proposed a Convolution Neural Network model for processing real time input images containing Malayalam handwritten characters and the task of segmentation of words and characters from a document image and to make character prediction from the CNN model created. In this model the feature extraction task is implicitly done in CNN by using the gradient descent algorithm which was proposed by Yann Le Cunn. This system is effectively used for digitization of Malayalam handwritten script which contains 13 vowels and 36 consonants and which is carried out through various phases and achieved an accuracy of 97.26% for trained dataset.

**P.K.SandhyaBalakrishnan, L.Pavithira[6]** proposed a method to optimize Convolution Neural Networks using Stimulated Anneling which proves the efficiency of character recognition. They mentioned about various deep learning methods, functionality of CNN which is a deep supervised learning and methods for training CNNs. Stated that “Character recognition is procedure of identifying and recognizing characters from input image and converts them into ASCII or other equivalent machine editable form”. The proposed system evaluated the OCR accuracy different language texts from various books which shows that percentage accuracy of CNN by SA is better than the original CNN.

**Md. Anwar Hossian, SadiaAfrin[7]** served a technique for extracting features from images and identification of exact characters to produce text document called **TEMPLATE MATCHING**. They stated “OCR functions like human ability of reading. As in natural process how an image seen by the human eyes is the input for brain, the same way the image scanned by the scanner will be the input for the OCR system”. They mentioned that this technique works by comparing derived image features of the image and the template for each possible displacement and the approaches for text recognition and detection in images and videos can classify as connected component based method and texture based method. Informed various concepts about OCR, Template Matching and different steps involved in implementation of Template Matching. They evaluated the system by comparing the results of trained datasets and unprepared datasets where trained datasets gave good results with 100% accuracy.

Bangladesh to the east of India on the Bay of Bengal is a South Asian country marked by lush greenery and many waterways Its Padma Meghna and Jamuna rivers create fertile plains and travel by boat is common On the southern coast the Sundarbans an enormous mangrove forest shared with Eastern India is home to the royal Bengal tiger

**Fig.5 Input Image with noise[7]**



**Fig.6 Output text document [7]**

**Chandra SekharV,SrinivasaChakravarthy, Viswanath P[8]** proposed an efficient OCR model based on Convolution Neural Networks for recognizing Bharati characters which can serve as a common script for most Indian languages. They stated that “Bharati Script is a unified Indic Script that can be used to express most major Indian languages. The simplicity of its glyphs and lucid and logical compositionality makes it an ideal candidate for OCR development”. As other Indic characters Bharati character also has a three-tier structure where three tiers are disconnected and are clearly segmentable in OCR by connected component analysis. The proposed model achieves 76.70% with test documents consists of 50% noise and 99.98% with test documents of 0% noise.

Vowels

अ	आ	इ	ई	उ	ऊ	ऋ	ॠ	ऌ	ॡ	ए	ऐ	ऑ	ओ	औ	अं	अँ	अः
८	८	८	८	८	८	८	८	८	८	८	८	८	८	८	८	८	८

Consonants

क	ख	ग	घ	ङ
ॠ	ॡ	ॢ	ॣ	।
च	छ	ज	झ	ञ
॥	०	१	२	३
ट	ठ	ड	ढ	ण

**Fig-7.Bharati Script For Devanagarscript [8]**

**Table- I: Survey on various research papers using OCR**

S.NO	AUTHOR NAME	PAPER NAME	RESEARCH FINDING	EVALUATION PROCEDURE USED	SUMMARY
------	-------------	------------	------------------	---------------------------	---------

1.	Md.Anwar Hossain & Sadia Afrin [7]	Optical Character Recognition based on Template Matching.	Grey scale images and bitmap image format were used and alphabets and numbers were recognized by comparing both the images using matlab R2018a software.	Uses Template Matching technique to convert text image into editable text documents.	Correctly and incorrectly clarified letters are counted and accuracy is calculated and is 100% for trained set of font.
2.	Yu weng, chunlei xia [9]	A new deep learning-Based handwritten character recognition system on mobile computing devices.	By considering the computing environment and data characteristics of mobile devices, a lightweight network structure for optical character recognition on specific data sets was proposed.	A convolution neural network is applied to handwritten characters to recognize characters.	With the increase in the network scale i.e. dataset, the results of convolutional neural networks are getting better, but training time is bit longer.
3.	Shreshtha garg, kapil kumar gupta, nikhil prabhakar, amulya ratan garg, aayush trivedi. [10]	Optical character recognition using artificial intelligence.	OCR system will train itself and help in extracting text from any image by using neural networks and back propagation techniques. ANN approach is used for classification.	OCR with neural networks is used.	This method and approach was able to recognize the character upto 100% when image was noiseless and about 95% in noisy image.
4.	Gauri katiyar, shabana mehfuz [11]	A hybrid recognition system for off-line handwritten characters.	This approach achieved better accuracy and reduced computational time for recognition using genetic algorithm.	A genetic algorithm approach along with an adaptive multi layer perception is used.	It improved overall performance on real world recognition problems.
5.	Israr Ud Din, Imran Siddiqi, Shehzad, Tahir Azam. [12]	Segmentation-free optical character recognition for printed urdu text	A segmentation-free optical character recognition system for printed Urdu Nastaliq font using ligatures as units of recognition.	It used segmentation-free optical character recognition system is used.	An overall recognition rate of 94.55% is achieved on ligatures database.
6.	Ali Farhat, omar hommos, ALi Al-Zawqari, Abdulhadi Al-aahtani, Faycal Bensaali, Abbas Amira, Xiaojun Zhai [13]	Optical Character Recognition on heterogeneous SoC for HD Automatic Number Plate Recognition System.	A comparison based on the recognition rate, time performance, hardware utilization of the proposed algorithm with already existing ones is conducted.	Vector crossing, zoning, The 1 <sup>st</sup> approach zoning and template matching were used.	It is found that the heterogeneous SoC saves 80% of the energy consumed by the PC.
7.	P.K.Sandhya Balakrishnana, L.Pavithira [6]	Multi font Optical Character Recognition using deep learning.	To optimize convolution neural networks using simulate Annealing for optical character recognition.	Stimulated Annealing algorithm is used.	A number of scanned books in different languages are downloaded and their OCR accuracy is evaluated.
8.	Ondrej Bostik, Jan Klecka [14]	Recognition of CAPTCHA characters by supervised machine learning algorithms.	To compare several common machine learning algorithms for Optical Character Recognition of CAPTCHA codes.	Different Supervised methods such as Decision trees, K-Nearest Neighbours, Support vector machines, Artificial Neural Networks were used.	All the used algorithms can classify the objects into right class with success rate around 99%.
9.	Savitha Attigeri [15]	Neural network based Handwritten Character Recognition system.	A neural network based off line handwritten character recognition system without feature extraction has been introduced for classifying and recognizing the 26 english alphabets.	A multilayer Feed Forward Technique is used.	This system is less complex compared to the offline methods using feature extraction techniques.

# An Analysis of Challenges in Character Recognition on Palm Leaf Manuscripts, and Their Solutions

10.	Sandhya Balakrishnan P K, Dr.L.Pavithira [16]	Optical Character Recognition using improved Rootsift features.	A Neural network is designed to model the way in which the brain performs a particular task or function of interest and improved RootSIFT feature extraction technique is applied for calculating the features of pixels.	An improved RootSIFT descriptor for feature extraction along with the artificial neural network algorithm is used.	This system has achieved a maximum accuracy of 97.3077% compared with the neural network algorithm, which has a 90.8022% accuracy.
-----	-----------------------------------------------	-----------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------

### III. ALGORITHMIC ANALYSIS:

#### A.ConvolutionNeuralNetworks:

Convolution Neural Network is a Deep Learning Algorithm which teaches the computers to learn by examples, in which the input image will be taken and importance will be assigned to various objects in the image and is able to differentiate one object from other.

**Table-II: Analysis of CNN based approach for recognizing Malayalam hand written characters.**

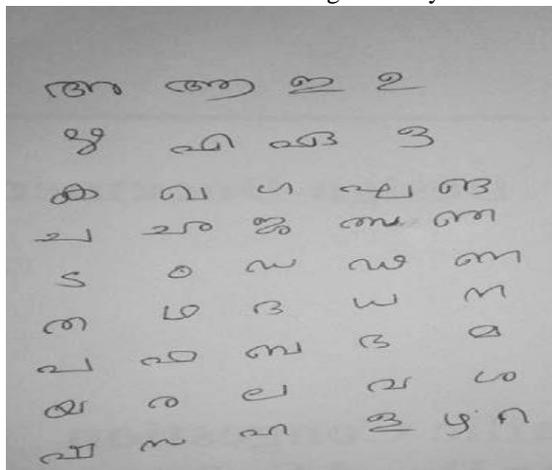
The proposed model for recognizing handwritten Malayalam

S. No	Challenge	Solution	Efficiency
1.	A CNN based model for recognizing Malayalam handwritten Characters.	To accurately recognize the Malayalam handwritten script which contains 13 vowel letters and 36 consonant letters and the letters formed by the combination of vowels and consonants.	The proposed system is successfully implemented in real time with 97.26% accuracy.

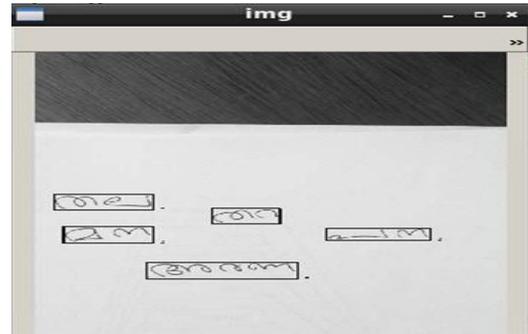
characters uses CNN for classification task, and the system architecture consists of four phases:

1. Data set creation and Augmentation
2. Defining CNN Architecture
3. Training the CNN model.
4. Deployment of the model.

After creating CNN model hyper parameters are tuned to get the increased accuracy. The model gave a training accuracy of 97.14 with loss 0.378, validation accuracy of 96.04 with loss 0.3242 and a testing accuracy of 97.26.



**Fig.8 Handwritten Malayalam Characters [5]**



**Fig.9 Recognition of handwritten Malayalam characters by system [5]**

#### B. Pattern Recognition:

Pattern Recognizing is the process of classifying input data into objects based on Key Features and yields information about given data set. It has applications in computer vision, radar processing, speech recognition and text classification.

**Table-III: Analysis of printed Bangla OCR system.**

S. No	Challenge	Solution	Efficiency
1.	A system to read printed Bangla script using pattern recognition.	To read the printed Bangla script which has 11 vowels and 39 consonant characters.	The overall accuracy of the system is about 95.05% in word level and 99.10% in character level.

In proposed system, at first the document is to be scanned for digital image, Noise and image skew if any, should be corrected. The lines, words and characters in the scanned image are to be segmented and different categories of characters are to be grouped accordingly. The errors if any while scanning the image are to be corrected by calling the pre defined error correction algorithm. The results showed that 74.22% of single character errors generated by the OCR were corrected by error correction module and the overall accuracy of the system is 95.50% in word level which is equivalent to 99.10% in character level.

#### C. Textual Processing Analysis:

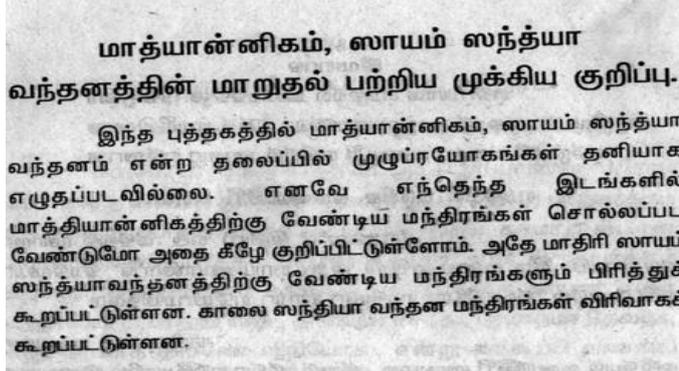
Text processing is focused on textual characters at the highest computing level which is concerned with automatic transmission of information. It involves computer commands which invoke content, content changes and cursor movement and generate a processed report of the content of text file.

**Table-IV: Analysis of Tamil OCR system**

S. No	Challenge	Solution	Efficiency
1.	A precise system to recognize printed Tamil text using skew detection approach.	To recognize the printed Tamil script which has 12 vowels and 23 consonants and the letters formed by the combination of vowels and consonants.	The overall recognition accuracy of the system is around 98%.

மாத்யான்னிகம், ஸாயம் ஸந்த்யா வந்தனத்தின் மாறுதல் பற்றிய முக்கிய குறிப்பு. இந்த புத்தகத்தில் மாத்யான்னிகம், ஸாயம் ஸந்த்யா வந்தனம் என்ற தலைப்பில் முழுப்பிரயோகங்கள் தனியாக எழுதப்படவில்லை. எனவே எந்தெந்த இடங்களில் மாத்தியான்னிகத்திற்கு வேண்டிய மந்திரங்கள் சொல்லப்பட வேண்டுமோ அதை கீழே குறிப்பிட்டுள்ளோம். அதே மாதிரி ஸாயம் ஸந்த்யாவந்தனத்திற்கு வேண்டிய மந்திரங்களும் பிரித்துக் கூறப்பட்டுள்ளன. காலை ஸந்தியா வந்தன மந்திரங்கள் விரிவாகக் கூறப்பட்டுள்ளன.

**Fig.11 Recognized Document [18]**



**Fig.10 Original Document [18]**

**Table-V: Performance comparison of some of the existing OCR systems**

S. No	OCR System	Accuracy (IN %)	TEST Used
1.	A complete Tamil optical character recognition system.	98%	Efficiency
2.	A complete Bangla OCR system.	95.05% in word level, 99.10% in character level.	Efficiency.
3.	A convolution neural network based approach for recognizing Malayalam handwritten characters.	97.26%	Efficiency.
4.	Neural network based Handwritten Character Recognition system.	90.19%	Efficiency
5.	Recognition of CAPTCHA characters by supervised machine learning algorithms.	99%	Efficiency
6.	Multi font Optical Character Recognition using deep learning.	97%	Efficiency
7.	Segmentation-free optical character recognition for printed urdu text.	94.55%	Efficiency
8.	A hybrid recognition system for off-line handwritten characters.	91.56% for capital letters and 87.49% for small letters.	Efficiency
9.	An efficient Multi Lingual Optical Character Recognition system for Indian languages through use of Bharati script	76.70% for 50% noise and 99.98% for 0% noise.	Efficiency
10.	A complete OCR for printed Hindi text in Devanagari script.	93% at character level	Efficiency

**IV. CONCLUSION:**

In this paper, different works on character recognition are addressed along with their algorithms and evaluation. This paper highlights about the identified research problems, their solutions and efficiency proposed by different researchers. These findings can be applied on palm leaf manuscript of different languages as a kick start to the complex character and text recognition problems. The more the required features

are extracted the more accurate will be the detection and recognition of the characters.

**REFERENCES**

1. D Uday Kumar, G. V. Sreekumar, U. A. Athvankar.: Traditional writing system in Southern India -- Palm Leaf Manuscripts. Design Thoughts (2009)



# An Analysis of Challenges in Character Recognition on Palm Leaf Manuscripts, and Their Solutions

2. Narenthiran R, Saravanan G, Ramanujam K.: The digitization of palm leaf Manuscripts.(2012)
3. MohamadSageer T.K , Dr.A.T.Francis.: Palm Leaves Manuscripts in Kerala and their Preservation:Factors necessitating digital archiving (IJSRP, volume 4, Issue 2, 2014)
4. DeeptiGanapathy.: Preserving India's palm leaf manuscripts for the future.(2016)
5. Jabir ali V, Jose T Joseph.: A Convolution Neural Network based Approach for Recognizing Malayalam handwritten charactes.(IJSER volume 9,issue 12, 2018)
6. P.K.SandhyaBalakrishnan, L.Pavithira.: Multi-font optical character recognition using Deep Learning.(2019)
7. Md.AnwarHossain ,SadiaAfrin.:Optical Character Recognition based on Template Matching(2019)
8. Chandra SekharVoruganti, SrinivasaChakravarthy, ViswanathPulabaigari.: An efficient Multi Lingual Optical Character Recognition system for Indian languages through use of Bharati script (ICVGIP, 2018).
9. Yu weng,ChunleiXia.:A new deep learning-Based handwritten character recognition system on mobile computing devices.(2019)
10. Shreshthagarg, kapilkumargupta, nikhilprabhakar, amulyaratangarg, aayushtrivedi.: Optical Character Recognition using Artificial Intelligence (IJCA, Volume 179 - N0.31, 2018)
11. Gaurikatiyar, shabanamehfuz.: A hybrid recognition system for off-line handwritten characters.(2016)
12. IsrarUd Din, ImaranSiddiqi, Shezad Khalid, TahirAzam.: Segmentation-free optical character recognition for printed urdu text.(2017)
13. Ali Farhat, Omar Hommos, Ali Al-Zawqari, Abdulhadi Al-Qahtani, FaycalBensaali, Abbas Amira, XiaojunZhai.: Optical character recognition on heterogeneous SoC for HD automatic number plate Recognition system(2018)
14. OndrejBostik, Jan klecka.: Recognition of CAPTCHA characters by supervised machine learning Algorithms (IFAC-Papers Online, volume 51, issue 6, 208-213, 2018)
15. SavithaAttigeri.: Neural network based Handwritten Character Recognition system (IJECs, vol 7 No.3, 23761-23768, 2018)
16. SandhyaBalakrishnan P K, Dr.L.Pavithira.: Optical character recognition using improved rootsift Features(2018).
17. B.B. Chaudhuri, U Pal al.: A complete printed BANGLA OCR system (Pattern Recognition, Volume 31, issue 5, 531-549, 1998)
18. Aparna K G, A G Ramakrishnan.: A complete Tamil optical character recognition system (5<sup>th</sup> International workshop, DAS 2002)
19. Nagendra Panini Challa, R Vasanth Kumar Mehata.: Evaluation of Automatic Metadata Schema for Indian palm leaf manuscripts(IJITEE, volume 8, issue 5, 2019)
20. S.S.Devika,K.Vijayakumar.: Digitization of palm leaf manuscripts in Tamil Nadu(2016)
21. Palm-leaf manuscript, en.wikipedia.org
22. Prof. BambKalpesh.K.: A literature survey on character recognition of Indian scripts for new researchers.(2016)
23. PrzemyslawKudlacik,Piotrporwik.: A new approach to signature recognition using fuzzy method(2012)
24. Suryaparakashkompalli, Srirangarajsetlur,venugovindaraju.: Devanagari OCR using a recognition driven segmentation framework and stochastic language models(2009)
25. Jayashree Rajesh prasad.: Handwritten character recognition:A review(2014)
26. VeenaBansai, R.M.K Sinha.: A complete OCR for Printed Hindi text in Devanagari Script (ICDAR, 2001)
27. Causes of deterioration of the manuscripts, ways and means for proper preservation.
28. Soumenghosh, ArnabMahajan, Dr. SwapnaBenerjee.: Palmleaf manuscript conservation, the process of seasoning with special reference to saraswathimahall library, tamilnadu in India : some techniques (IJIM, volume 2 issue II, 122-128, 2017).

**Lakshmi Mounika Alavala** is an undergraduate student from Department of CSE , SCSVMV. Her area of research includes Image Processing, computer vision, ontologies.

**Nagendra Panini Challa** is an Assistant Professor from Department of IT, SVECW. He has done his Ph.D. in CSE from SCSVMV and M. Tech in Department of CSE from JNTU Kakinada and B.E from SCSVMV. His area of research includes digital image and video processing, computer vision related approaches.

**Lakshmi Kalyani Adurthi** is an under graduate student from Department of CSE, SCSVMV . Her area of research includes Image Processing, Computer Vision , Ontologies.

## AUTHORS PROFILE

**Dr. R. Vasanth Kumar Mehta** is currently working as an Associate Professor and Head, Department of CSE at SCSVMV. He obtained his Ph.D. in CSE from SCSVMV and M.Sc (Tech) and B.Sc in CSE from BITS Pilani His research interests are computer vision, machine processing and image processing.

Retrieval Number: F8339038620/2020©BEIESP  
DOI:10.35940/ijrte.F8339.038620  
Journal Website: [www.ijrte.org](http://www.ijrte.org)

Published By:  
Blue Eyes Intelligence Engineering  
& Sciences Publication