International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 07 | July - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Character Recognition of Malayalam Palm Leaf Manuscripts Using Convolutional Neural Networks

Sruthi S Kumar

Department of Computer Science
University of Kerala

Abstract - A wealth of knowledge, tradition, and ancient culture can be found in the palm leaf manuscripts. These scriptures must be preserved to the greatest extent feasible because they are an essential component of our rich culture. To determine the optimal way for digitalizing the Malayalam language, an extensive literature survey was done on the various Character recognition techniques used in Malayalam as well as other languages. The proposed method transforms characters from ancient Malayalam palm leaf manuscripts into a computer recognizable text file format. The input of these systems should be images of palm leaf documents and the obtained output will be recognized characters in current version of Malayalam text. In this Malayalam character recognition system we integrated Ostu's method based binarization and Horizontal, Vertical projection profile for segmentation. And for training the model, we used Convolutional Neural Networks (CNN) which simplifies the entire process by avoiding feature extraction. And we obtained 95.30% accuracy for the developed model.

Key Words: Palm leaf manuscripts, Ostu's method, Binarization, Convolutional Neural Network (CNN)

1. INTRODUCTION

Historical palm leaf manuscripts are abundant sources of knowledge. It is one of the oldest medium of writing that helps to understand humanity with various fields such as medicine, astronomy, astrology, mathematics etc. [2]. The Palm leaf character recognition has many challenges. Palm leaf manuscripts are different from handwritten characters on papers or other medium. Many of the Palm Leaf images have varying contrast and illumination, stains, smear, smudges and contaminations [3]. Valuable and immeasurable information could be passed to the next generation if we preserve the palm leaf manuscripts in a proper transform. Palm leaf manuscripts are an inevitable part of our rich culture. It is our responsibility to preserve this bundle of information for future generations. Digitalization is the best way to preserve palm leaf manuscripts. But the data collection from the palm leaf is a tiring piece of work because of various challenges like character set consisting of large amount of noise and facing difficulty in reading and understanding old Malayalam scripts.

Recognition of character is one of the vital areas in the field of image processing and pattern recognition [4]. Optical Character Recognition (OCR) is the technique that we used to recognize characters from Palm leaf manuscripts. Actually OCR is used to recognize text in a digital image. It can transform a physical document, palm leaf manuscripts, or an image into machine-readable text. The goal of OCR is to convert legal or historical documents into a text-based digital form. This character recognition technique involves scanning a text document with an image scanner, which converts the written document into an image, and then translating the image into character codes.

2. LITERATURE REVIEW

A review of the available literature in the field of palm leaf manuscript character recognition was carried out.

A system for extracting information from old palm leaves and converted the ancient Malayalam scripts to their current version based on contrast-based adaptive binarization and Convolutional Neural Networks algorithm is implemented in article [1]. This system has an accuracy of 96.7%. In paper [2], a study on image segmentation of Malayalam palm leaf manuscripts was explored. Ostu's algorithm using binarization is carried out and then segmentation processes are done in both thinned and binary palm leaf document images. And get into the conclusion that the segmentation of thinned images is better than binary palm leaf images.

Two different methods are introduced for character recognition in Khmer manuscripts [3]. The first one is isolated character recognition by different types of algorithms like CNN, LSTM-RNN, and a mix of both. And the second task is for identifying text image patches of variable length and at the same time localizing glyphs in the character image. And from the result they concluded that both CNN and RNN based model perform well on this task. For word recognition task LSTM network is used. In paper [4] for the Tamil palm leaf character recognition, the performance of CNN model compared with other machine learning algorithms they are, K-Nearest Neighbour (K-NN), Support Vector Machine (SVM) and Fast Artificial Neural Network (FANN). Finally get in to the conclusion that CNN model has higher accuracy of 96.21%.

An effective Cnvolutional neural network for character recognition is carried out. Starting with image acquisition step then passed through a series of phases like pre-processing, binarization, segmentation, classification and mapping. And this system results an accuracy of 96.46% [5]. Deep Convolution Neural Network and Probabilistic Neural Network (PNN) models are implemented for the character recognition of Tulu Manuscripts. This system shows a maximum accuracy of 88.07% while using Deep CNN architecture [6]. The palm leaf characters have an additional property like depth. This method explored how to extract these 3D features and how they can be used effectively in the

identification and classification tasks using Random Transform and Nearest Neighbourhood Classifier (NNC). The accuracy obtained for this method is 93% [7].

Segmentation method for connected characters of Thai-Noi script on palm leaf manuscripts is introduced in article [8]. The input to the system is the binary image of handwriting and the output is segmented characters. By using the contour tracing algorithm, the accuracy obtained for this model is 85.57%. In article [9] an improved Viterbi algorithm based on Hidden Markov Model (HMM) is developed to find all possible segmentation paths. And then, a path filtering method is used to find the optimal paths of segmented text blocks. The average segmentation accuracy obtained is 89.91%. In article [10] five different binarization methods are implemented using Otsu's, Niblack, Sauvola, Bernsen, Howe algorithm. And get in to the conclusion that Niblack method shows high accuracy of 85%.

A binarization free approach for line and character segmentation of palm leaf document images are introduced [11]. There are four sub tasks for this method; they are brushing character area of gray level images with minimum filtering, gray level images average block projection profile, segmentation path's candidate area selection, and nonlinear segmentation path is constructed. This system results 78.57% accuracy. Binarization of the palm-leaf documents by Niblack, Sauvola, NICK and Bradley binarization algorithms are discussed in article [12]. By evaluating the output images of all the binarization methods they get into the conclusion that the output of Bradley algorithm is much suitable for segmentation of characters**.** To recognize the palm leaf characters 3D feature of every pixel in an image is used. Here images are divided into zones and the sum of the pixel intensities in each zone is used as a feature vector to recognize the palm leaf characters. Nearest Neighbour classifier is used for classification and this system has 96% accuracy [13].

3. PROPOSED SYSTEM METHODOLOGY

The proposed method transforms characters from ancient Malayalam palm leaf manuscripts into a computer recognizable text file format. The input of these systems should be images of palm leaf documents and the obtained output will be recognized characters in current version of text. Figure 1 is the architecture diagram of proposed system.

The whole implementation process of the system as follows:

A. Image Acquisition
B. Pre-processing
  (a) Gray Scale image
  (b) Binarization
  (c) Noise removal
  (d) Skew detection and Correction
  (e) Skeletonization or Thinning
C. Segmentation
  (a) Line Segmentation
  (b) Word Segmentation
  (c) Character Segmentation
D. CNN based Classification
E. Mapping

A. Image Acquisition

Dataset collection is the task carried out in image acquisition. The system is tested and trained using the collected dataset [1]. Database contains character samples from Malayalam palm leaf manuscript images. We collected images of Malayalam palm leaf manuscript documents from Calicut University's manuscript library. At a resolution of 300 dots per inch (dpi) the images where scanned.

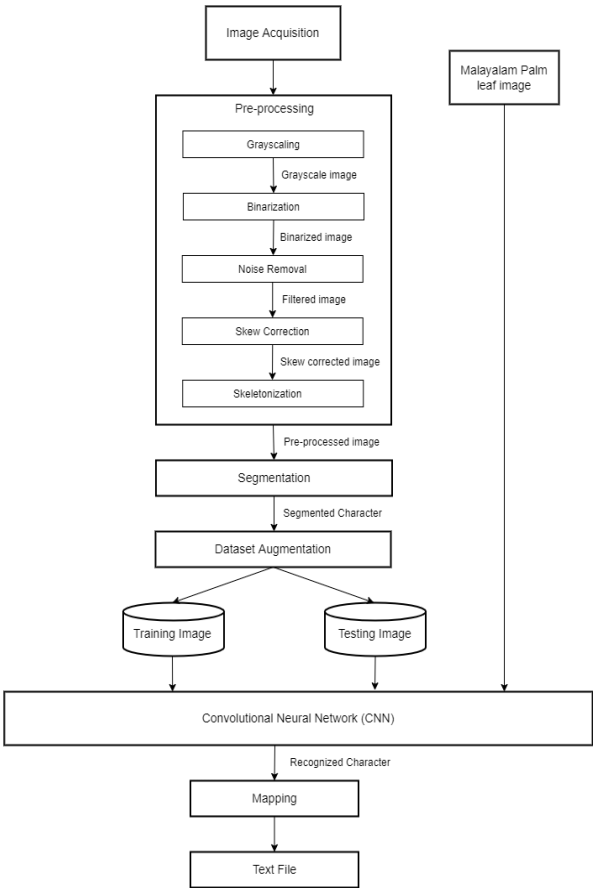

Fig.1 Proposed System Architecture

B. Pre-processing

The goal of pre-processing is to improve the image quality so that we can analyze it better. By doing preprocessing, we can reduce undesired distortions and improve some of the features that are required for the character recognition task [14]. Through the pre-processing steps the noise level in an image should be optimized and areas outside the text should be removed. This step is important for identifying handwritten documents that are more sensitive to noise. Pre-processing allows us to get a clear character image for better image recognition results. The steps that are used in preprocessing are, Binarization, Noise removal, Skew detection and Correction, Thinning or Skeletonization.

(a.) Grey Scale image

The only colors in a grayscale image are shades of grey. The difference between these images and other types of color images is that each pixel requires less information. Grayscale images are sometimes referred to as black and white images.

(b.) Binarization

Binarization is one of the important phases in pre-processing. Binarization is used to separates the foreground

and background information in the input image [12]. One of the popular binarization approaches is thresholding. The binarization is the process to convert the grayscale image to black and white image using certain threshold. Image thresholding is used to binarize the image based on their pixel intensities [10]. If the intensity of a pixel in the input image is greater than a certain limit (threshold), the corresponding output pixel is marked white (foreground), and if the input pixel intensity is less than or equal to the threshold, the position of the output pixel is marked black (background) [5]. There are two types of thresholding, they are global and local. In global thresholding one of the most efficient methods is Otsu's method [2]. Here we used Ostu's method based binarization. The binarization process of palm leaf manuscript document image as follows;

Step1: Read scanned color palm leaf manuscript images
Step2: Palm leaf manuscript images are converted to grayscale image
Step3: Background is removed from grayscale image
Step4: Enhance the grayscale image
Step5: Grayscale image converted into binary image using Ostu's algorithm
Step6: Display the binarized image

(c.) Noise removal

Scanned image may contain noise therefore, it is very important to remove these noises before processing the image [7]. In the palm leaf manuscript image the noise may arise due to deterioration of palm leaf over ages, poor contrast, holes or spots on the palm leaf, any fungal infection on the medium etc. When pixels in an image show various intensity values instead of actual pixel values derived from the image, this is referred to as noise. By smoothing the entire image and leaving sections near contrast boundaries, noise reduction algorithms diminish or eliminate the visibility of noise. Before processing the palm leaf images, noise must be removed with filters. To remove noise, there are numerous filters available. Here we used low-pass filters for noise reduction.

(d.) Skeletonization

Skeletonization, also known as Thinning, is a technique for obtaining a one-pixel-wide representation or skeleton of an object while maintaining the object's connectivity and end points [13]. Thinning reduces visual components to their most important information, allowing for easier analysis and recognition. As a result, detecting important traits becomes much easier in the future. The motive of Skeletonization is to reduce the image into their essential information so that the further analysis and recognition are much easier [6].

(e.) Skew Detection and Correction

A few degrees of skew (tilt) are unavoidable whether a document is fed to the scanner mechanically or by a human operator. The skew angle is the angle formed by the lines of text in a digital image with respect to the horizontal axis [14]. Skew estimation can be done in a number of ways. We used the document's projection profile based method for skew correction.

C. Segmentation

The division of a digital image into many components is known as image segmentation. The purpose of segmentation is to make an image easier to understand by simplifying or changing its representation [15]. The segmentation of Malayalam characters from palm leaf manuscript is an important stage in the character recognition system. In the context of character recognition, segmentation is the process of breaking down a picture into text lines, words, and finally characters that can be used for classification. Segmenting characters from palm leaf manuscripts is extremely difficult, because the character structure and content differ substantially. The accuracy of the recognition system is dependent on the segmentation [9]. The recognition system produces the greatest results when the characters are segmented correctly. In the segmentation step, a picture is separated into regions or objects. Segmentation aims to extract the script's most important elements, which are almost always characters. This is desirable because the classifier perceives these characters only. The segmentation phase is particularly important for reducing inaccuracy caused by touching characters that the classifier fails to recognize correctly [8]. Segmentation is the process of separating individual characters from a piece of handwritten text. The lines, words, and characters are segmented using projection profile-based algorithms.

(a.) Line Segmentation

The process of segmenting lines from an entire text document is called line segmentation. Based on Horizontal Projection (HP), line segmentation is performed. The Horizontal Projection approach, in which peak-valley points are found and employed for line separation is a prominent technique for line segmentation.

(b.) Word Segmentation

Word segmentation is carried out by using a Vertical Projection (VP) and detecting a threshold that exceeds horizontal gaps, after which words are removed from a text line and placed on a separate line [11]. By examining the minima in the vertical projection profile, the peaks and valleys can be found, and words can be separated.

(c.) Character Segmentation

Character segmentation is a crucial phase in the recognition process since incorrectly segmented characters would lead to recognition problems. Vertical projection is the method used for character segmentation. The total of pixel values in each column of the image is the vertical projection value. The vertical projection value of zero indicates a character boundary because the image for character segmentation is a binary image. The following is how the character segmentation is done;

Step 1: Find the size of the image document [called vertical projection M N]
Step 2: Calculate the column sum of pixel value, value for column from 1 to N.
for i=1 to N
$VP(i) = \Sigma^{M}_{j=1}$ Image (j)
end
Step 4: If there are several columns with VP values of zero, the first column with a non-zero VP value is chosen as the segmentation starting point. The last point for segmentation is the first column with a zero VP value. As a result, the Character portion is situated halfway between the first and last points.

D. CNN Based Classification

In a character recognition system, classification is the decision-making phase. The segmented characters are passed to the classifier as input. Convolutional neural networks are employed for categorization in our proposed system. CNN are a subclass of deep models that were inspired by how the human brain processes information [4]. CNNs have been presented as a way to simplify image processing in networks. CNN is a neural network that can transfer values to the next layer without losing spatial information, making it ideal for extracting and categorizing features. These CNN properties make it possible to use spatial information, which helps in uniquely identifying characters in the image [5]. The input, hidden, and output layers make up a CNN. Feature maps and a fully connected layer with convolutional layers and pooling layers make up the layers. The convolutional and pooling layers collect the input values' properties and map the extracted values to the feature map. In this process, the characters in the palm leaf image can be extracted through considering the unique features in the individual characters, and then the fully connected layer has a classification value from features extracted for classification.

The CNN procedure is as follows: Parsed data is passed to the feature maps from the input layer. The data is stored in the convolutional layer of the feature maps at a given position. A convolutional operation is done on the data and mapped to a pooling layer by the convolutional layer. Before being mapped to the pooling layer, the data goes through a max pooling process. The greatest value of the previous layer's results is extracted using max pooling. Following that, CNN builds a fully connected layer that incorporates all of the convolutional and pooling layers. Finally, the result is transmitted to an output layer via the fully connected layer. The developed CNN model, which is used to categorize the input characters, used the functions like Conv2D, MaxPooling_2D, Batch Normalization, flatten, Dropout, Activation, Dense and softmax are used to learn the parameters where an approximation of 855,152 trainable parameters was obtained.

The goal of the classification step is to assign a class label or membership scores to each observation in the defined classes. The backpropagation algorithm is used to train the CNN. During the training, a dropout layer will be used. The network's complexity will be reduced by the dropout layer. The network is initially trained using a training set. The label of the character class to which the input character belongs will be the classifier's output [1]. Because the Malayalam characters are so similar in shape and structure, we deployed multiple networks to improve the system's accuracy.

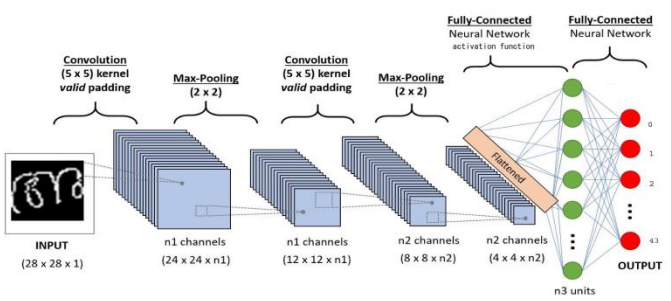

Fig.2 Convolutional Neural Network Architecture

E. Mapping

The process of converting the old Malayalam scripts to its modern form is known as mapping. The output of the classifier is converted into character Unicode. The output that we obtained from the classifier is the label which represents the character class. The relevant Unicode character is mapped according to this label. The Unicode that we used for each character is different and it is drafted in text file.

F. Overall Work Flow of Proposed System

This system helps in automatic recognition of characters from Malayalam palm leaf manuscripts and maps them into their current version of text. Data collection is the first phase. After this phase the data undergoes pre-processing steps. The noise level in an image should be optimised through the pre-processing phases, and portions outside the text should be removed. For pre-processing we used several methods such as; Gray Scaling, Binarization, Noise removal, Skew detection and Skeletonization. Here we use Ostu's algorithm for binarization so the preprocessing process is very efficient to remove unwanted information in the image and it helps in better character recognition results. After the pre-processing steps the segmentation process is carried out.

The segmentation of Malayalam characters from palm leaf manuscripts is a crucial step in the character recognition process. Segmentation is the process of breaking down a picture of palm leaf manuscripts into text lines, words, and finally characters that can be utilized for categorization in the context of character recognition. When the characters are segmented appropriately, the recognition system delivers the best results. So here we used Horizontal and vertical projection profile approach for segmentation. These techniques help in effective segmentation of characters from palm leaf manuscripts. After the segmentation the segmented characters are stored in separate folders, we have collected 43 different Malayalam characters from the palm leaf dataset. Here we have almost 500 numbers of characters in each of the 43 character classes after augmentation. The image augmentation is performed using position augmentation techniques like Resize, cropping, adjusting tilt, Rotation, Transformation functions to artificially increase the size and generate copies of our segmented images.

After the augmentation we have a dataset of 43 classes of individual characters where each individual character consists of 500 different type images of each character. The system is trained and tested using the segmented data from the palm leaf images. After the data augmentation task we are splitting the dataset into Training and Testing dataset. We are using 70% of the collected data for training and rest 30% data for testing. Here we are giving 70% of data in the dataset that is considered as training dataset in to this learning model. The network has been trained to recognize characters. Each character's training set has 350 images. In order to accurately track our model's performance, we additionally set aside distinct image sequences for testing it. Each character in the testing set has 150 images. The error rate was calculated using categorical cross entropy as the loss function. The training phase produces palm.h5 file as its result. The parameters of the network are saved in a palm.h5 file after training and are used during testing. Then CNN automatically analyses each and every characters in the dataset and classify accordingly.

Finally the output of the learning model should be the recognized characters and we are mapping these characters to the current version of text. The mapping is done using classifier label and Malayalam character denoting code. And finally the characters are integrated and output is obtained in text file format.

4. EXPERIMENTAL RESULTS AND ANALYSIS

A. Dataset

Databases containing character samples from Malayalam palm leaf document images were collected for our research. We collected 500 photos of Malayalam palm leaf manuscript documents from Calicut University's manuscript library. At a resolution of 300 dots per inch (dpi) the images where scanned. We use it to recognize characters in palm leaf manuscripts. Noise reduction, background removal, segmentation, classification, and recognition are all part of the system's procedures.

Fig.3 Sample of Malayalam Palm Leaf Manuscript Dataset

B. Experiments

In this Malayalam character recognition system we integrated Ostu's method based binarization and Horizontal, Vertical projection profile for segmentation. And for training the model we used CNN algorithm.

(a.) Pre-processing

In this pre-processing step the unstructured data is transferred into structured form and the unwanted information in the image is removed. An image's noise level should be reduced, and regions outside of the text are deleted. Pre-processing allows us to get a clear character image for better image recognition results. The steps that are used in pre-processing are, Grey scaling, Binarization, Noise removal, Skew detection and Correction, Thinning or Skeletonization.

- Grayscaling: The Grayscaled image contains less pixel information as compared to other colour images. The only colors in a grayscale or graylevel image are shades of grey. The value of each channel in a grayscale image is same.

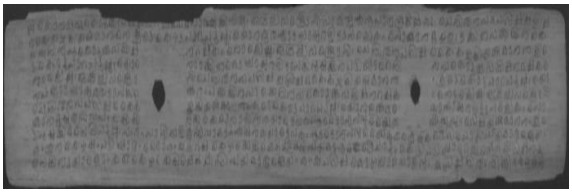

Fig.4 Grayscale image

- Binarization using Ostu's method: The foreground and background information in the input image are separated via binarization. It is the technique of transforming a grayscale image to a black and white image by applying a set of thresholds. Here we done Ostu's method based binarization.

Fig.5 Binarized image

- Skeletonization / Thinning: It aims to reduce the image into their essential information so that the further analysis and recognition are much easier.

Fig.6 Skeletonized image

Pre-processing tasks such as Noise removal, contrast adjusting, skew correction, back ground elimination, unwanted information removal are done in palm leaf image. And it results in better character recognition results.

(b.) Segmentation

The accuracy of a recognition system is determined by the segmentation. When the characters are segmented appropriately, the recognition system delivers the best results. The Malayalam palm leaf manuscript image is separated into lines initially, then individual words, and finally individual characters during segmentation.

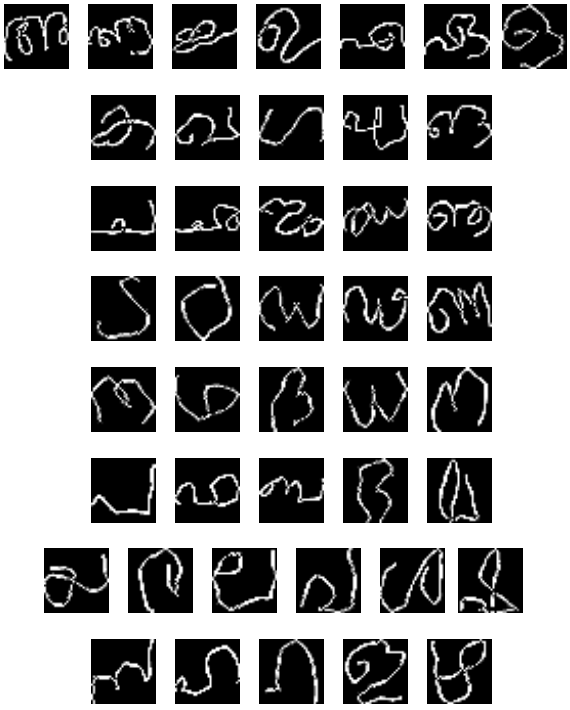

Fig.7 Segmented Characters

(c.) Character Recognition

The recognized Malayalam characters from palm leaf manuscripts are mapped into the current version of Malayalam text.

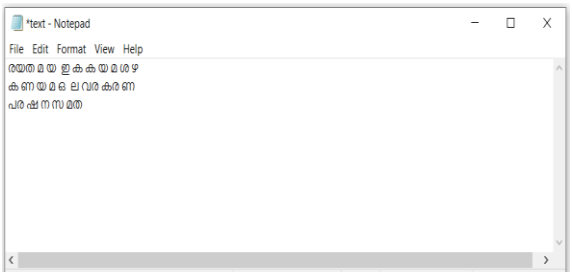

Fig.8 Characters from palm leaf manuscripts are mapped in to current version of text.

C. Results Analysis

The model is tested using 30% of the data from the dataset. Here for testing 150 character images are used for each and every character in the dataset. So 150 character images for 43 different characters, here we have used total of 6450 character images for testing. The total Accuracy that we have obtained for this classification model is 95.30%.

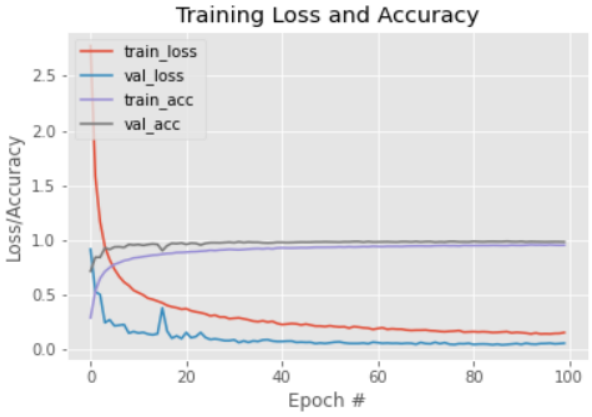

Fig.9 Validation Accuracy and loss plot of the model for 50 epochs

From figure 9 it is very clear that as the number of epochs increases the accuracy of the model also increases where the loss function decreases. While training the model we have maximum accuracy when running 100 epochs. When we try to run more epochs there is a chance of Overfitting. Here the validation accuracy shows the actual accuracy of the developed CNN model. The obtained maximum accuracy for the CNN classification model is 95.30%.

5. CONCLUSIONS

Convolutional neural networks were used in the development of the system to recognize Malayalam palm leaf manuscripts. The input images are pre-processed to improve their quality and binarization is performed using Ostu's approach. Apart from the binarization, Skeletonization also done for pre-processing, it facilitates the process of segmentation more easy. Horizontal and vertical projection techniques are used to segment lines, words and characters. For classification, CNN were deployed, which were trained using the back propagation algorithm. Our system shows the accuracy of 95.30%. The network's output will be a character class label, which will be mapped to the matching character in the current Malayalam character version.

6. FUTURE WORKS

The quality of the palm leaf manuscripts is one of the major problems faced by the developed system. Working with damaged palm leaf manuscripts would create misleading and erroneous results, thus for better accurate results, we must simply avoid the poor quality palm leaf manuscripts. To overcome these issues, an effective prediction system based on Word Sense Disambiguation (WSD) can be devised and implemented in future. This WSD system is more efficient because it analyses contextual meaning of the text. So in future the recognition of characters in ancient documents like palm leaf manuscripts, The Natural Language Processing (NLP) methods also wants to be incorporated. As future work, the proposed method could be expanded to train for all the characters in Malayalam like compound letters, Chillus, Dependent vowels etc.

ACKNOWLEDGEMENT

The author thankfully acknowledge Dr. Geena K P, Assistant Professor Department of Computer Science, University of Calicut for giving support for collecting the Palm leaf manuscripts from universities' Manuscript library.

REFERENCES

1. D. Sudarsan, P. Vijayakumar, S. Biju, S. Sanu and S. K. Shivadas, "Digitalization of Malayalam Palmleaf Manuscripts Based on Contrast-Based Adaptive Binarization and Convolutional Neural Networks," *2018 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pp. 1-4, 2018.

2. Geena K P and G.Raju "Image Segmentation of Malayalam Palm Leaf Manuscripts", International Journal of Innovative Research in Computer and Communication Engineering Vol.1, Issue 2, 2015.

3. D. Valy, M. Verleysen, S. Chhun and J. Burie, "Character and Text Recognition of Khmer Historical Palm Leaf Manuscripts," *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pp. 13-18, 2018.

4. R. S. Sabeenian, M. E. Paramasivam, R. Anand and P. M. Dinesh, "Palm-Leaf Manuscript Character Recognition and Classification Using Convolutional Neural Networks," Lecture Notes in Networks and Systems (LNNS) book series of Springer, volume 75, pp- 397-404, 2019.

5. D. Sudarsan and S. Joseph, "A Novel Approach for Handwriting Recognition in Malayalam Manuscripts using Contour Detection and Convolutional Neural Nets," *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 1818-1824, 2018.

6. C K Savitha, P J Antony "PNN and Deep Learning Based Character Recognition System for Tulu Manuscripts", International Journal of Engineering and Advanced Technology (IJEAT), Volume-8 Issue-5, pp. 2249-8958, June 2019.

7. P. N. Sastry and R. Krishnan, "Isolated Telugu Palm leaf character recognition using Radon Transform — A novel approach," *2012*

World Congress on Information and Communication Technologies, pp. 795-802, 2012.

8. R. Chamchong and C. C. Fung, "A combined method of segmentation for connected handwritten on palm leaf manuscripts," *2014 IEEE International Conference on Systems Man and Cybernetics (SMC)*, pp. 4158-4161, 2014.
9. G. Peng, P. Yu, H. Li and L. He, "Text line segmentation using Viterbi algorithm for the palm leaf manuscripts of Dai," *2016 International Conference on Audio, Language and Image Processing (ICALIP)*, pp. 336-340, 2016.
10. E. Paulus, M. Suryani, S. Hadi and I. N. Yulita, "A novel scheme for handwritten binarization method on sundanese palm leaf document images," *2017 Second International Conference on Informatics and Computing (ICIC)*, pp. 1-5, 2017.
11. Made Windu Antara Kesiman, Jean-Christophe Burie, and Jean-Marc Ogier "A New Scheme for Text Line and Character Segmentation from Gray Scale Images of Palm Leaf Manuscript," 2016 15th International Conference on Frontiers in Handwriting Recognition, pp- 325-330, 2016.
12. Sabeenian R S, Paramasivam M E and Dinesh P M, "Appraisal of localized binarization methods on Tamil palm-leaf manuscripts," *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pp. 793-797, 2016.
13. T. R. V. Lakshmi, P. N. Sastry, R. Krishnan, N. V. K. Rao and T. V. Rajinikanth, "Analysis of Telugu Palm Leaf Character Recognition Using 3D Feature," *2015 International Conference on Computational Intelligence and Networks*, pp. 36-41, 2015.
14. B.Anuradha Srinivas, Arun Agarwal, and C.Raghavendra Rao "An Overview of OCR Research in Indian Scripts," IJCSES International Journal of Computer Sciences and Engineering Systems, Vol.2, No.2, April 2008.
15. D. Valy, M. Verleysen and K. Sok, "Line segmentation for grayscale text images of khmer palm leaf manuscripts," *2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pp. 1-6, 2017.