Electronic Letters on Computer Vision and Image Analysis 25(1):1-21, 2025

Preserving the Heritage: Evaluating the Character Segmentation Quality in Palm Leaf Manuscripts by Comparing the Classical and Noise2Void Denoising Techniques

Deepa Unnikrishnan* and Vignesh Radhakrishnan+

* *School of Information Science,*
+ *School of Computer Science and Engineering,*

Presidency University, Itgalpur, Rajanakunte, Yelahanka,Bengaluru, Karnataka, India, 560 119

Received 14th of august, 2025; accepted 27th of november, 2025

Abstract

The Palm Leaf Manuscripts are a rich source of information about ancient India. It shares a vast amount of knowledge about the past, including art, culture, literature, and medicine. As the Manuscripts were developed organically, they are prone to getting damaged very easily. Various methodologies are used to preserve actual copies of manuscripts; however, due to climatic conditions, the deterioration of these manuscripts is inevitable. This work presents a comparative analysis of classical and deep learning-based approaches for denoising degraded palm leaf manuscripts, with a focus on the quality of the segmented text inscribed on the PLMs. The classical pipeline consists of denoising, followed by binarisation and finally segmentation of the entire image. This sequence of processes was implemented using both Fast Non-Local Means and a self-trained Noise2Void (N2V) model for denoising. However, the segmented characters, particularly from the Fast NL-based approach, appeared visually distorted. In contrast, the N2V-based difference image showed better structural preservation and closer alignment with the ground truth. To tackle these limitations, a novel and innovative processing pipeline was developed, which commences with denoising Palm Leaf Manuscript images using the N2V model, proceeds with the direct extraction of text, and culminates in the targeted application of binarisation exclusively on the segmented patches. This restructured approach minimises the noise, enhances text clarity, and preserves character details more effectively. It eliminated the crucial step of removing distortions, thereby saving the process time. Quantitative evaluation reveals improved performance with lower MSE values (0.97, 1.15, 1.02), higher PSNR scores (27.17 dB, 26.61 dB, 29.09 dB), and a structural similarity index (SSIM) of 91superiority of the proposed method over the traditional workflow.

Key Words:Binarising, Degraded, Denoising, Distorted, Palm Leaf Manuscripts, Segmentation.

Correspondence to: deepa.20233csc0006@presidencyuniversity.in

Recommended for acceptance by Angel D. Sappa
https://doi.org/10.5565/rev/elcvia.2320
ELCVIA ISSN:15108-5097
Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

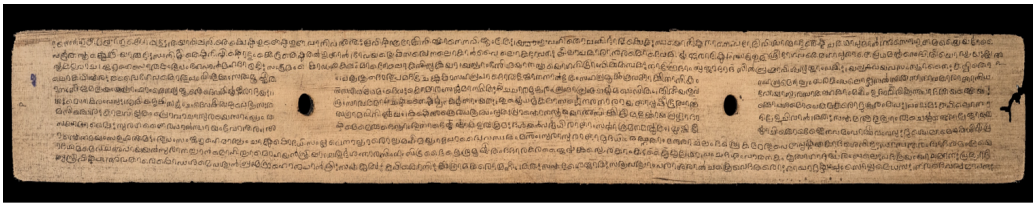
1 Introduction

South and Southeast Asia have the most collections of different types of manuscripts. The word manuscript originally derived from two Latin words: 'manu', meaning hand and 'scriptus', meaning written [1]. India boasts one of the most extensive and ancient collections of manuscripts in various languages, encompassing diverse themes ranging from painting and art to fiction and scientific innovations [2, 3]. The manuscripts from India reveal a great deal about the history, for example, how Islamic culture began to influence India over a thousand years ago. Linguistic experts may be particularly interested in translating Buddhist manuscripts, especially those that contain Ayurveda or pertain to Buddhist texts, as the roots of Buddhist monasteries still persist in India [4]. The manuscripts were developed from either dried and treated palm leaves or processed and flattened papery wooden bark. The life of a manuscript can be more than a thousand years if it is treated well, but climatic conditions can also affect its deterioration due to its organic nature. In the hot, humid areas of India and Sri Lanka, the lifespan of Palm leaf Manuscripts (PLMs) can be reduced to 400–500 years. PLMs typically lack a standardised structure or design; their shapes and dimensions may vary, with some being circular and others in rectangular forms. These manuscripts usually contain one or two punch holes used to tie the bundle together with thread [5, 6]. Due to their age, PLMs are often physically damaged. As part of preservation tasks, they may have undergone treatments involving oil or turmeric. But such preservation needs periodic maintenance. Another way to preserve the PLMs is by making scanned or digital copies; however, this necessitates substantial memory space, and maintenance can be costly.

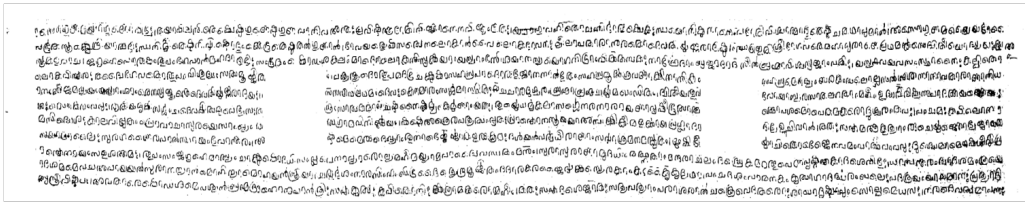
1.1 Contributions of the Present Work

Recent CNN and GAN-based approaches for processing historical, degraded documents have shown excellent results in denoising and reconstructing missing text; however, these supervised methods require large paired datasets of degraded and clean images. Such datasets are rarely available for PLMs, whose degradations vary greatly. Moreover, existing deep models often generalise poorly across different writing styles or physical variations in the leaves, producing over-smoothed text or artefacts when applied to unseen manuscript types. To overcome these limitations, we propose a self-supervised, character-level segmentation framework exclusively for PLMs. The Noise2Void (N2V) model is employed for denoising without the need for paired data, enabling the network to learn directly from corrupted images and adapt to diverse degradation patterns. In contrast to the Baseline Framework, which processes the entire image, the proposed framework primarily denoises the PLMs implementing the N2V model and then extracts the text into individual character regions prior to binarisation. This processing order remarkably improves structural continuity, retains fine textual details, and yields segmented outputs of text-patch images.

The study provides a comparative examination of these two segmentation strategies - the Baseline Framework and the proposed Character Segmentation Pipeline - and analyses that the latter is more robust and applicable for real PLM collections, where leaf dimensions, shape and content density vary significantly. Because each text region is considered separately, the proposed approach achieves consistent quality in binarisation regardless of the design or scale of the PLMs. A comprehensive evaluation using full-reference (MSE, PSNR, SSIM) and no-reference (NIQE, PIQE, BRISQUE) metrics, supported by Visual Quality Estimation (VQE), proves that the character-level framework achieves higher PSNR and SSIM values, generating visually sharper, distortion-free text. These outcomes surpass existing CNN/GAN-based strategies for denoising and binarising and establish the Character Segmentation Pipeline as a practical and straightforward solution for the digital preservation of PLMs.



Original image of PLM



Groundtruth image of PLM

Figure 1: The image of the original Palm Leaf Manuscript and the Corresponding Ground truth generated .

2 Related Works

This section provides an overview of some of the earlier works carried out in the area of denoising, binarisation, and segmentation of text from degraded documents. The section is organised into two parts: (1) Conventional ethodology encompassing the traditional methods for working with the degraded PLMs, (2) Deep learning and Generative methods, which leverage deep learning and generative models for document enhancement and segmentation.

2.1 Conventional Methodology Followed in Denoising and Binarising

Sulaiman et al. described various types of noises that can be present on degraded documents, including tearing marks, ink bleed, smudges, and contrast variation [7]. They introduced a hybrid method by combining Otsu's binarisation and Sauvola's method. The paper also discussed the various evaluation strategies that can be used to evaluate the quality of the resultant images. Employing the Image Gradient Approach (IGA), Dhanya Sudarsan and Deepa Sankar developed binarised images of PLMs, without removing the Punch Hole Mark (PHM), by achieving a PSNR value of 14.64 dB [8]. They carried out a similar work on Malayalam PLMs to construct a novel method for completely denoising and binarising the PLMs. The method involved a series of steps to achieve a PSNR value of 19.40 dB [9].

Ge Peng et al. proposed a model specifically designed for PLMs using a stroke edge detection. The method utilised an adaptive algorithm to detect areas with intensity differences by computing the contrast map and then binarising it. The canny edge detector isolates the prominent edges in the map and combines them with a binarised contrast map using a bitwise AND operator, helping to remove the small strokes on the images. Since the method is highly dependent on Canny, the performance may drop for the blurred strokes. The proposed method got a PSNR value of 19.68 dB [10]. Ekta Vats et al. proposed a method for document image binarisation, which utilised adaptive median filtering for denoising, employing appropriate filters to preserve text details and mask background noise. The final binarisation is done using Kitter's minimum error thresholding. The hyperparameters were automatically tuned using Bayesian Optimisation, resulting in a PSNR value of 18.13 dB [11].

By combining the Hybrid thresholding technique and Circular contour analysis for detecting the PHM, Unnikrishnan D et al. presented a work that achieves excellent quality binarised images by completely removing noise and PHM from degraded PLMs using a minimal procedure [12]. Singh M et al. carried out a study on

degraded PLMs by combining a Hybrid denoising technique and conservative Gaussian filtering with a smoothing operation in the HSY colour space to enhance the distorted manuscript images for character extraction. The method is validated using entropy and PSNR metrics. [13]. B Gatos introduced a novel method implementing a local adaptive binarisation method that requires no tuning of parameters to effectively handle the degradations, preserving the textual integrity as confirmed by Levenstein distance analysis [14] Wadgy et al. proposed a method for image binarisation on degraded documents implementing Tetinex and global thresholding, was designed to reduce the adverse impact of degradations, showing a strong visual OCR performance even under inadequate lighting illumination [15].

2.2 Deep learning and Generative methods

Jyothi RL proposed a work on analysing and comparing the efficiency of various wavelet transform families for recognising segmented text from PLMs containing Grantha scripts, which were historically used in ancient India. The PLMs were binarised using Otsu's binarisation algorithm, and the resultant images underwent thinning of the text using the Hilditch algorithm. The segmentation of the text from the PLMs was done using a combination of projection analysis and connected components labelling. The Discrete Meyer Wavelet provided the highest accuracy of 71% [16]. Su et al. utilised Generative Adversarial Networks (GANs), a model that learns to generate realistic images, to reconstruct missing strokes or parts from damaged characters in ancient documents. To enforce the restoration quality, the Font Style Loss is also proposed to preserve the original calligraphic style, achieving excellent SSIM and PSNR values [17].

There were multiple works done by Uzan et al. [18], Chen, K. et. al. [19], Pastor-Pellicer, J. [20], Renton, G et. al. [21], Alaasam et al. [22], A. Prusty et al. [23], Watanabe, K. et al. [24], Ziran et al. [25], Cai et al. [26] on degraded documents analysis, employing CNN-based models focusing on text restoration, segmentation and recognition. Most of the strategies include optimised convolutional networks, GANs, and FCNs with enhancements like adversarial losses. Even though the datasets used were different (Arabic, Chinese, or Latin scripts), all the methods aim to preserve the structure and legibility under degradation. The SSIM and PSNR values of the resultant images showed an improvement in the efficiency of the models.

Abbas Ali et al. conducted a detailed study on the works aimed at enhancing degraded documents, concluding that future research should focus on developing complex neural networks that are more adaptable to varied languages and writing styles, based on the degradation levels [27].

Following the critical analysis of existing work on PLMs, the specific contribution and novelty of the present study are outlined in section 1.1.

3 Materials and Methods

From the literature, it was clear that although many earlier works had been conducted on denoising and binarising, including advanced deep learning models, none of the methods had particularly emphasised the quality of the segmented objects after binarisation. While many CNN models achieved strong results in preprocessing and restoration of degraded documents, they typically require a pair of training data or extensive labels. In contrast, Noise2Void performs self-supervised noise removal without requiring clean images. N2V also addresses fine-grained noise patterns and background degradation, thereby improving the recognition accuracy after segmentation. This work focused on evaluating the segmentation quality obtained when two different approaches were employed in conjunction with various binarisation methods. This section describes the various methodologies employed in the study to facilitate the evaluation.

3.1 Dataset Preparation

The images of the PLMs are collected from the Kanipayur Shankaran Namboothirippad Memorial Library, Thrissur, Kerala. The 'thaliyola kettu' contained a total of 330 olas of Shrimad Bhagavath Gita, a beautiful

epic from ancient India. The original image is shown in Figure 1. The PLMs are more than 100 years old, and most of the pages are damaged due to fungal or bacterial attacks, with the characters engraved fading. The images of the palm leaves were taken by scanning the PLMs in RGB colour mode. The scanned images were then subjected to post-processing to give the border in black colour using Adobe Photoshop. The images were stored in TIFF format, and approximately one image takes 12.1 MB of memory space. The leaves in the bundle vary in size and shape. Some of the measurements of the leaves, in centimetres (cm), at 300 DPI, along with the resolution, are provided in Table 1 for reference. The original image is cropped to reduce the size

PLM-Length in CM	PLM-Height in CM	Resolution (Width X Height) in pixels
64.77	7.44	7650 x 879
64.20	7.43	7583 x 878
64.04	6.88	7564 x 813
63.87	7.27	7544 x 859
63.70	6.93	7524 x 819

Table 1: Dimensions of PLMs-not all the dimensions are included

without losing the necessary information. The images were then converted to greyscale and denoised for further processing. The coding part of the work is completed in Python 3.10 using Google Colab, and MATLAB is used for image evaluation. To avoid exceeding the Colab's RAM and slowing down I/O, the dataset of 330 images was divided into n batches of images to stay within a 12-16 GB GPU memory envelope.

3.2 Ground truth image Generation

The ground truth images are generated by denoising and binarising the images using an adaptive thresholding method [28]. The resultant images are then manually corrected to get clear output images. The ground-truth images are used to compare the resultant images generated by the proposed method. The ground truth image, which is generated from the original image, is depicted in Figure 1 for reference.

3.3 Assessment Methodologies Used in the Study

For the thorough and effective evaluation of the resultant images that are generated from the proposed method, a combination of different assessment metrics was chosen. The blind image quality assessment includes Natural Image Quality Evaluator (NIQE) [29], Blind/No Reference Image Spatial QUality Evaluator (BRISQUE) [30], and Perception-based Image Quality Evaluator (PIQE) [31], which don't require any ground truth image for the evaluation. Then, a full reference metric is used to compute the MSE (Mean Squared Error), PSNR (Peak Signal-to-Noise Ratio) [32], and SSIM (Structural Similarity Index Metric) [33] values of the image by comparing the processed image with the corresponding ground-truth image. The final one employs a subjective, human-based evaluation through VQE [34]. A summary of the evaluation metrics used in the study is depicted in Table 2.

Metric	Type and Description	Equation	Eq. No	Interpretation
MSE	**Full Reference:** Computes the average squared difference between the processed image (pi) and the ground truth image (gt).	$\text{MSE} = \frac{1}{KL}\sum_{y=1}^{K}\sum_{x=1}^{L}[pi(x,y) - gt(x,y)]^2$, Where K is the number of Rows and L is the number of columns	(1)	Lower MSE values indicate fewer distortions.
PSNR	**Full Reference:** Ratio between the maximum possible pixel value and distortion power, derived from MSE.	$\text{PSNR} = 10\log_{10}\left(\frac{MAX_i^2}{MSE}\right)$,Where MAX is the maximum possible intensity value of the image	(2)	Higher PSNR indicates better image quality.
SSIM	**Full Reference:** Compares luminance, contrast, and structure between the processed image and ground truth.	$\text{SSIM}(pi, gt) = l(pi, gt)\, c(pi, gt)\, s(pi, gt)$	(3)	Values close to 1 indicate better structural similarity.
NIQE	**No Reference:** Compares statistical features of the test image with a natural scene model.	$\text{NIQE} = \sqrt{(\mu_x - \mu_y)^T\left(\frac{\Sigma_x+\Sigma_y}{2}\right)^{-1}(\mu_x - \mu_y)}$, Where, μ_x and $\sum x$, μ_y and $\sum y$is a mean vector and Covariance matrix from the pretrained model on NSS and test image respectively.	(4)	Lower NIQE values indicate better natural image statistics.
PIQE	**No Reference:** Evaluates perceptual distortions block-wise based on human visual sensitivity.	$\text{PIQE} = \frac{1}{N}\sum_{i=1}^{N} w_i D_i$, where wi is the perceptual weight, Di is the Distortions/ block.	(5)	Lower PIQE indicates fewer visible distortions.
BRISQUE	**No Reference:** Extracts NSS features and compares them with an SVR-based opinion model.	$f(i,j) = \frac{I(i,j)-\mu(i,j)}{\sigma(i,j)+1}$, Where, $I\,(i,j)$is the pixel intensity of the image, $\mu\,(i,j)$ is the intensity average, $\sigma\,(i,j)$ is the local standard deviation, the constant prevents division by zero $\text{BRISQUE}_{score} = f_{SVR}(f)$, Where f is the feature vector from Eq. 6, f_{SVR} is the Support Vector Machine (SVM)-pretrained model	(6) (7)	Lower BRISQUE indicates better perceived quality.
VQE	**Subjective:** Evaluates human ratings for readability, clarity, and noise suppression.	Not applicable	–	Higher VQE score indicates better readability.

Table 2: Image quality metrics used in this work.

4 Creating the Denoised Dataset Using Fast NL-Means and Noise2Void

As the PLMs were very old and degraded, the scanned copies of the PLMs contain many types of noise. The noise is caused by the fungal or bacterial attack, by tearing the edges, the punch hole mark (PHM), due to the natural texture of palm leaf or wood from which the PLMs were developed, and a significant amount of illumination variation because of the digital cameras ' captures, hence the denoising plays a vital role in the preprocessing of ancient documents. To evaluate the segmentation quality of the extracted text, two different sets of datasets were developed using two internal statistical methods. The first denoising method employed is the Fast NL Means algorithm [35], and the second method used is a deep learning technique, Noise2Void (N2V) [36], which is a self-trained neural network trained on noisy images.

4.1 Fast NL-Means Denoising Method

In the Fast NL Means (FNLM) Algorithm, each pixel in the image is processed one at a time. Around every pixel, a window of fixed size known as a template window is taken. The search window is created around the same pixel, but larger than the template window, where the algorithm searches for a similar patch to the template window. Then it calculates the similarity weight between the template window and the other patches. The pixel, also known as a target pixel, will be assigned a weighted mean of similar patches identified through the searches. The size of the template window and search window matters in the execution time. If the sizes are large, it can slow down the process even though it can produce clear images. In this study, we have used a template and search window size of 15×15 and 20×20 pixels, respectively.

4.2 Noise2Void (N2V)

Unlike classical CNN-based denoising, which requires a pair of noisy and clean images, N2V doesn't require clean images to denoise. It works directly on the noisy images. Both the input and the target images are from a single noisy image. A patch or receptive field is extracted from the image, which is the same as the input patch, but with the central pixel hidden. The N2V architecture is designed in such a way that the prediction of the central pixel is affected by all the other pixels in the receptive field. The advantage of the N2V architecture is that it does not learn the identity. During the study, each PLM grayscale image was divided into 8064 non-overlapping patches of size 64×64 pixels. As N2V employs blind-spot architecture, there are 8 blind spots chosen randomly per image. For each blind spot, the central pixel is masked, and the model is trained to predict the pixel intensities based on the surrounding context. The patches were divided into training and validation sets in an 80:20 ratio, resulting in 6448 patches for training and 1616 for validation per image. These parameters were selected through preliminary experiments that balanced denoising and text clarity, especially when thin strokes are present. During the experiments, it was noted that a larger patch size increased computation time while smaller ones produced artefacts. The model reached convergence near epoch 15; however, after epoch 12, the output images became overly smooth, and the text began to fade. Hence, epoch 12 was selected as the optimal training point to maintain the clarity of the text. One of the patches from the validation and training sets is shown in Figure 2. The resultant images obtained after denoising are shown in Figure 3. The N2V image has blurred the background and text, but without affecting the continuity of the text, resulting in a clear, noise-free, legible resultant image. However, the Fast NL Means algorithm results in smudged and clustered text, even though it appeared to be distinct from the background.

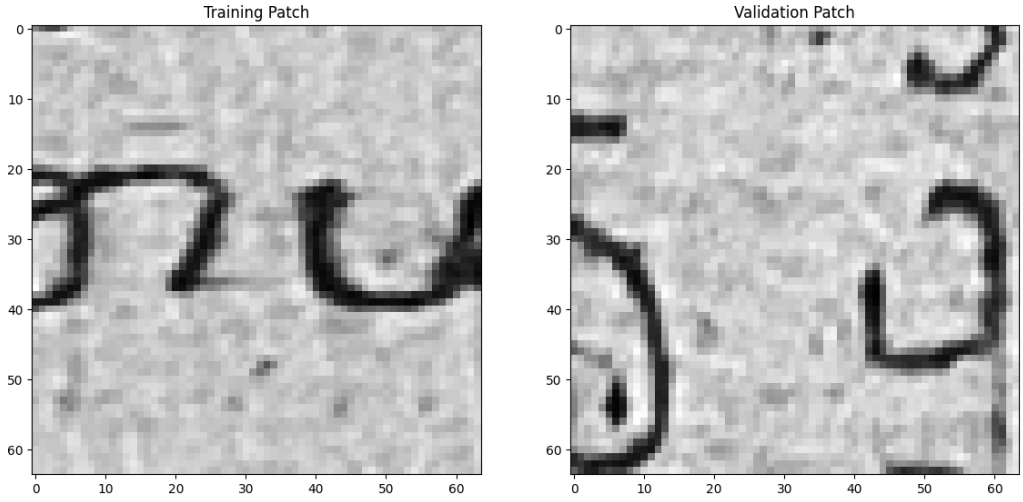


Figure 2: Patches generated while training the N2V model on noisy PLMs.

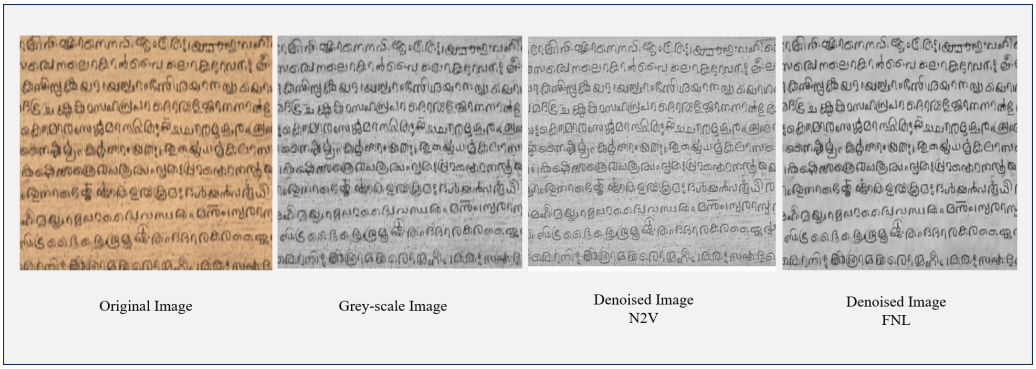


Figure 3: Resultant denoised images obtained from the Original PLM image.

5 Analysis of Denoised Images

5.1 Baseline Framework– using the full image of PLMs

In this study, two different denoising strategies—Fast Non-Local Means (FNLM) and N2V — were employed to establish a baseline comparison framework. Each denoised image underwent three classical binarisation algorithms, Otsu [37], Sauvola [38], and a Hybrid binarisation technique, from this, six unique datasets were produced - FNLM-Otsu, FNLM-Sauvola, FNLM-Hybrid, N2V-Otsu, N2V-Sauvola, and N2V-Hybrid. The workflow of Baseline Framework, shown in Figure 4. Algorithm 1 describes the steps to complete the Baseline framework.

Algorithm-1: Baseline Framework

Inputs: $I_FNLM[]$: Array of images denoised using Fast Non-Local Means
$I_N2V[]$: Array of images denoised using Noise2Void (N2V)
Tr_area : Threshold value for contour area
$Tr_circularity$: Threshold value for contour circularity
Outputs: Cleaned binarised images stored in separate folders for each method
Distortion count per image
Evaluation metrics for binarised outputs

1. Let Methods = $FastNL_Means, Noise2Void$
Let Binarisation$_M ethods$ =Otsu, Sauvola, Hybrid$Foreachdenoisingmethodm Methods$,
Retrieve the denoised image array Images []
2. For each image $IImages[]$ and each binarisation method $bBinarisation_M ethods$:
2.1 Compute the binarised image:
B = Binarise (I, method=b)
// Apply Otsu, Sauvola or Hybrid algorithm
3. Initialise large_distortions ← 0
Detect external contours: Contours = DetectContours(B)
For each contour C Contours:
3.1 Compute geometric properties:
area ← Area(C)
perimeter ← Perimeter(C)
If perimeter = 0, continue to the next contour
3.2 Compute circularity:
C ← $(4 \times \pi A)/(parameter)$
3.3 Evaluate distortion criteria:
If (A Tr_area) (C Tr_circularity):
large_distortions ← large_distortions + 1
Replace C in image B with white mask:
B = MaskContour(B, C)
4. Construct a path and save image B to the path with an appropriate naming convention.
5. For each processed image B:
Compute evaluation metrics:
NIQE ← Compute NIQE(B)
PIQE ← Compute PIQE(B)
BRISQUE ← Compute BRISQUE(B)
Store metrics along with image name and distortion count.

In this study, we employed a comprehensive set of image quality evaluation techniques to assess the effectiveness of the resultant denoising and binarising processes on ancient PLM images. We employed a combination of no-reference metrics, such as BRISQUE, NIQE, PIQE, and VQE, which considered the difficulty in developing ground-truth images for each PLM, along with full-reference metrics, including MSE, PSNR, and SSIM, to capture both global and local characteristics of the images. For implementing no-reference metrics, MATLAB is used.

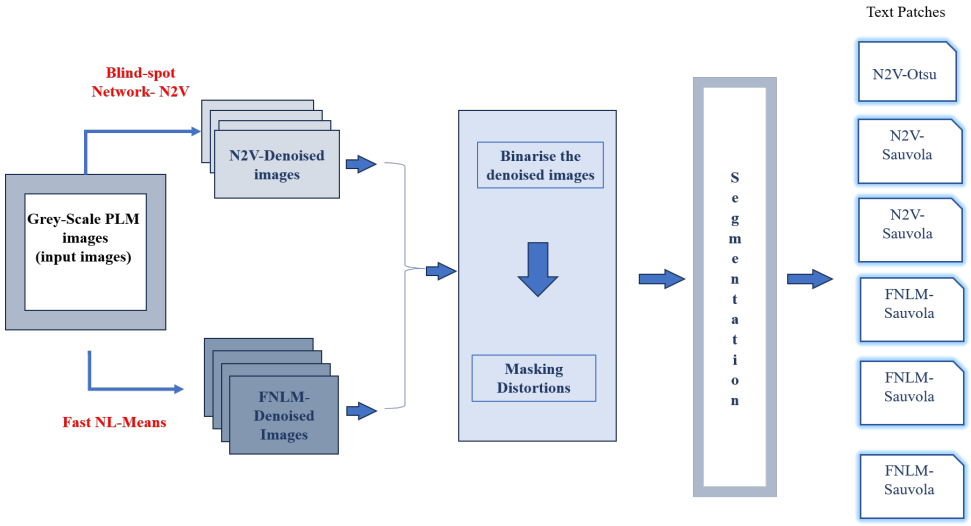


Figure 4: The Flow diagram of the Baseline Approach.

Method	MSE	PSNR	NIQE	PIQE	BRISQUE	SSIM	VQE
FNLM-Otsu	0.18	7.87	24.13	79.16	52.2	49.03	Bad
FNLM-Sauvola	0.16	7.34	22.73	78.59	52.44	46.56	Bad
FNLM-Hybrid	0.16	7.86	20.81	78.43	52.54	52.03	Bad
N2V-Otsu	0.016	17.95	16.6	81.75	47.42	86.34	Excellent
N2V-Sauvola	0.022	16.42	16.27	84.52	48.98	86.45	Fair
N2V-Hybrid	0.017	17.87	19.1	79.85	50.68	89.13	Good

Table 3: Comparison of the Results of Baseline Approach Obtained from various evaluation metrics on PLMs

5.1.1 The Results Obtained and Limitations from the Baseline Framework

The grey-scale images obtained from the two denoising methods, the Fast NL Means algorithm and the Noise to Void deep learning method, are then subjected to binarisation using the existing techniques, Otsu, Sauvola, and Hybrid binarisation, to evaluate the performance of the denoising techniques mentioned earlier. The results of the evaluation are given in Table 3. Figure 5 shows the resultant images after binarising the image employing the Baseline Framework.

From the results of NIQE, it was noted that the results obtained from the images are only moderate, as the values range from 15 to 25, even though the N2V images show slightly better values than the Fast NL Means images, indicating that there are a lot of distortions and illumination inconsistencies that still exist in the resultant images. PIQE provides two types of results: the PIQE score and the distortion map. The PIQE score of all the images ranges between 75% and 85%, showing a significant amount of degradation or noise present on the images. The corresponding distortion map obtained showed high localised distortions in the areas where text was present, which can seriously affect the image segmentation quality. BRISQUE values range from 45 to 55; these moderate values showed that although the N2V performed better than the FNLM, neither methodology was able to restore the expected quality of the manuscript images.

The MSE and PSNR results were computed on the normalised resultant images of PLMs. From the results, it was understood that the Fast NL Means denoising algorithm does not yield a good result on the developed datasets that implement binarisation methods when compared to the N2V datasets. The MSE and PSNR values for all the datasets developed using the Fast NL-Means denoising algorithm ranged from 0.18 to 0.16 and 7.34 dB to 7.87 dB, respectively, on the normalised images. In parallel, the N2V images, MSE, and PSNR values were in the range of 0.016 to 0.022 and 16.42 to 17.95, respectively, confirming N2V's effectiveness in suppressing noise while retaining textual details, such as strokes.

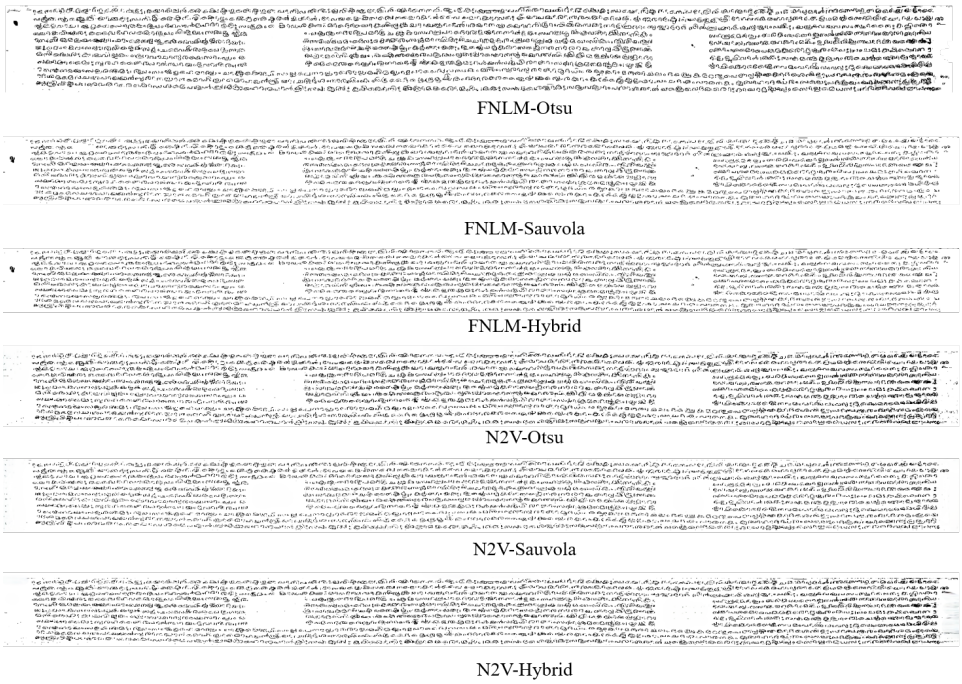


Figure 5: Binarised results obtained from Fast NL-Means and N2V.

To study the structural preservation of the PLMs, the resultant images were also evaluated using SSIM, which requires a ground truth image developed from the original image. The high values of SSIM indicate a good similarity index between the processed and the ground truth images. N2V images demonstrated good structural fidelity when compared to the Fast NL-Means images and the ground truth images developed. The VQE values

for each method were computed with the help of a human observer. The results of binarised images from all the datasets were presented to the experts for grading on a scale of 5 based on the legibility and clarity of the images. As per their analysis, N2V-Sauvola images were the clearest with less noise. They also noted that these image outputs displayed clear text and fewer clustering sets of FNLM images.

From the results obtained after applying the Baseline framework, and by looking at the holistic performance, we noted that the N2V images consistently excelled over those developed from the FNLM algorithm. However, the results also showed that significant improvements were still needed in denoising and binarisation of PLMs. From the segmented text from the images, it became evident that the text often lacked quality, appearing broken and distorted, as shown in Figure 6, especially the text from the right side of the image, showing effectiveness of the denoising and binarisation process was found to be influenced by both the physical dimensions and the resolution of the input images. Figure 7 illustrates the difference image between N2V-Otsu and FNLM-Otsu with the corresponding ground truth image. N2V-Otsu exhibits a significant illumination difference on the right side of the PLM, most likely because of a camera flash during image capture. FNLM-Otsu showed a considerable amount of variation throughout the entire resultant image. This excessive lighting led to blurred and clustered regions, adversely impacting the quality of the segmented text and resulting in the loss or partial segmentation of certain characters in most degraded areas [39].

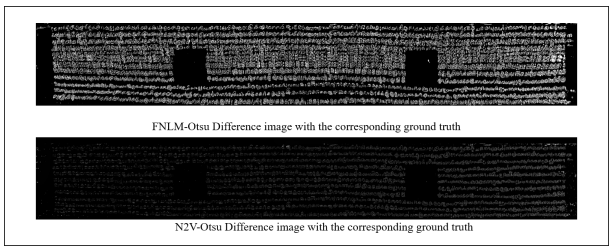


Figure 6: Difference images obtained for FNLM-Otsu and N2V-Otsu.

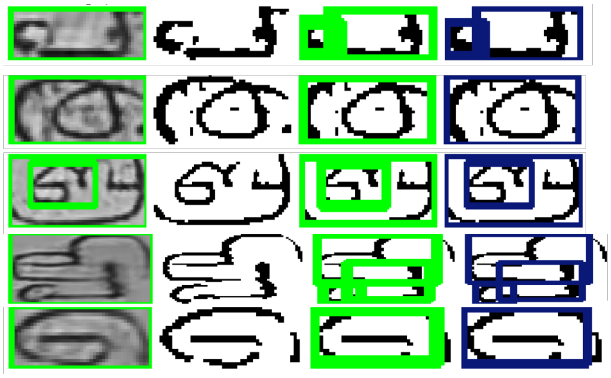


Figure 7: Samples of segmented text from Baseline Approach.

To address the issues faced in the baseline method, it is necessary to divide the image into multiple parts; however, this approach will significantly impact the continuity of the text, and the computation cost will also increase. Hence, to improve the segmentation quality of the text present on the PLMs, another approach was tested with the existing methodologies to denoise and binarise the images, that is, the second evaluation strategy, Character Segmentation.

5.2 Character Segmentation Pipeline-Proposed Method

From the N2V-denoised images, it was visually clear that the model is producing a good result when compared to the Fast NL–Means–denoised images. The text appeared smudged, even though there was a noticeable pixel intensity difference between the foreground and background in the images developed using the Fast NL-Means algorithm. In contrast, in the N2V–denoised image, the text remained intact and clear, even though the overall image seemed to be blurred.

In the proposed method, the N2V denoised images were used. The text regions of the images are cropped using contour analysis and then extracted to apply the binarisation. Since the N2V could suppress a good amount of noise present on the PLM images, the contour analysis on the PLMs was easy and effective. Figure 8 shows the flow diagram of the proposed method.

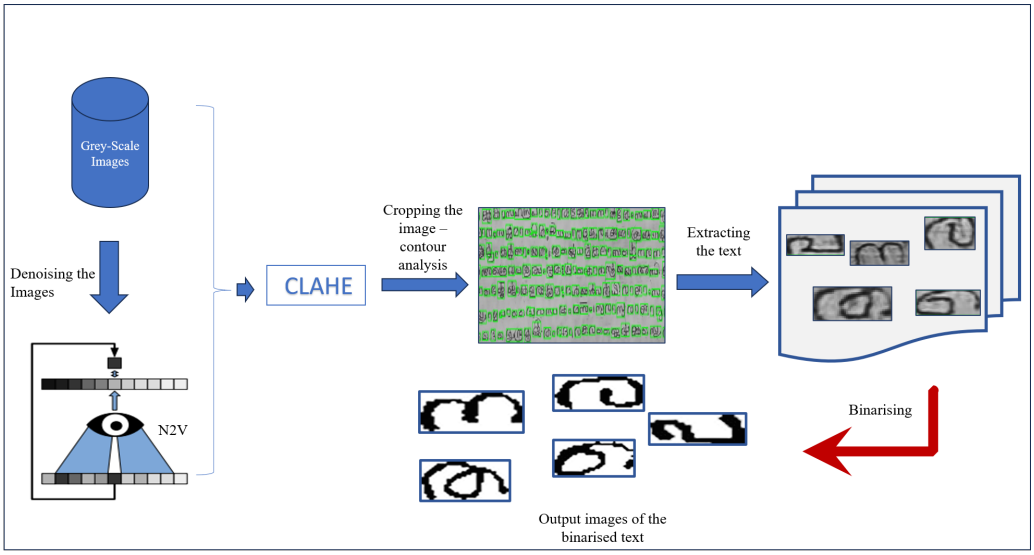


Figure 8: the Flow Diagram of the Proposed method.

The steps followed to complete the proposed method are described in Algorithm 2.

In this method, the earlier developed N2V–denoised images were used. The parameters of the N2V-deep learning method were chosen on the basis of trial and error. Figure 9 shows the output of the denoised images at different epochs while training the model. From the figure, it is evident how the N2V model was working on the degraded images of PLMs. At the 15th epoch, the background noise was completely suppressed without much affecting the quality of the texts.

Figure 10 illustrates the loss curve of the model, showing the decrease in training and validation losses over the 15 epochs, which denotes the model's convergence. These images were then subjected to Contrast Limited Adaptive Histogram Equalisation (CLAHE) [40] to enhance the text by increasing the contrast between the foreground and background, while preventing the global amplification of background noise. The step, where the removal of punch holes and other distortions from the image was taken out, was performed because while segmenting the text from the PLMs, the distortions like PHM and tearing marks were also detected and segmented and can be discarded in the next step of processing the images, that is, identification of the text segments.

By omitting step 4 in the baseline comparison, the time taken to loop through every contour in the image can be avoided, as most Indian scripts often include a large number of curved and circular patterns. The text segmentation was performed using contour analysis, and a bounding box was drawn over every independent text element based on its pixel intensity.

Finally, these texts were extracted and stored as a list of independent image files in the order it was segmented. To assess the performance of the proposed method, six sets of segmented characters were developed: three sets of Otsu, Sauvola, and Hybrid binarised segmented text from the Fast NL-Means denoising algorithm, and three additional sets of binarised segmented characters implementing the above-mentioned binarisation techniques from the N2V binarisation method. All six types of datasets for the cropped text were stored in the order they appeared in the image as separate files in PNG format and were compared to determine the optimal one.

Algorithm-2: Character Segmentation

Inputs: $D_{set} = Im, Im, ..., ImN$: Set of N grey-scale images denoised implementing Noise2Void (N2V)
Outputs: Cropped textual patches CT_P
Binarised patches Binarised stored in structured output directories

Procedure: 1. Define dataset:
$D_{set} = Imfori = 1toN$,
where each Im is a grey-scale denoised image.
2. For each image Im D_{set} :
Apply Contrast Limited Adaptive Histogram Equalisation (CLAHE):
Im ← CLAHE(Im)
// Enhances local textural contrast for better contour detection
3. Detect external contours in Im:
Contours ← Detect$_Contours(Im)$
For each contour C Contours:
3.1 Compute bounding box:
(x, y, w, h) ← Bounding$_Box(C)$
3.2 Crop the text region:
CT_P_ij ← Im[y : y + h, x : x + w]
// Patch $CT_P_ijrepresentsthejcroppedtextualregionfromimagei$
4. Define the directory path and persist each cropped patch
Save $(CT_P_ij, path + "patch_ij.png")$
5. For each patch CT_P_ij:
5.1 Compute optimal global threshold T$_otsubymaximisingbetween - classvariance$:
T_otsu_Otsu(CT_P_ij)
5.2 Apply binarisation:
$Binarised_(CT_ij_PT_otsu)$? 1 : 0
// Generates binary patch B from the greyscale input
Save each binary patch to the corresponding output directory

6 Results and Discussions of the Proposed Method

The dataset of PLMs exhibits noticeable variation in leaf dimensions, design and the number of characters that can be present on the leaf depending on the style of writing. As N2V divides the image into different patches, it does not require the input images to be of the same size, making it very suitable to for a heterogeneous set of data. The results obtained from both strategies were taken into account to evaluate the performance and to determine which method produced the best results. To address the slight blurring of the N2V denoised images, CLAHE was employed to enhance the clarity of the foreground; thereafter, a contour-based segmentation algorithm was applied to identify the characters on the leaf. The extracted characters underwent a binarisation for futher processing.

As the dimension of the textual patch images was tiny, the global binarisation produced vivid images. Each patch was stored as a separate PNG file in the order in which it appeared on the PLMs. To ensure robustness and quality, the method was evaluated using a combination of no-reference, full-reference, and subjective metrics. Table 4 depicts the results obtained from both the FNLM denoising and N2V implementations of the Character Segmentation Pipeline.

The classical FNLM-based methods showed nominal performance across all the metrics. Among these, FNLM-Hybrid attained the least MSE value (1.14) and the highest PSNR value (27.91 dB), indicating relatively superior pixel-wise restoration within the FNLM bucket. However, the method still suffered from poor visual quality, as indicated by a PIQE of 41.13, a NIQE of 19.18, and a BRISQUE of 40.56, showing the limited

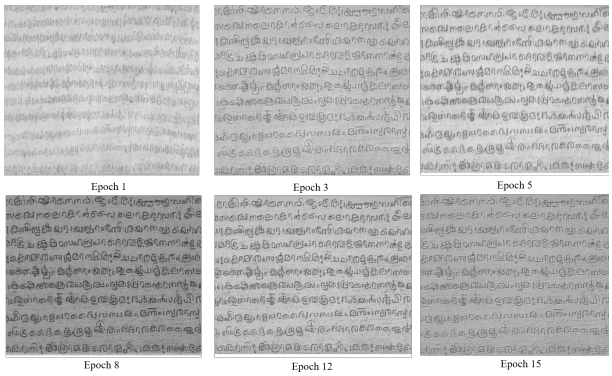


Figure 9: The resultant images generated on different epochs when N2V was implemented.

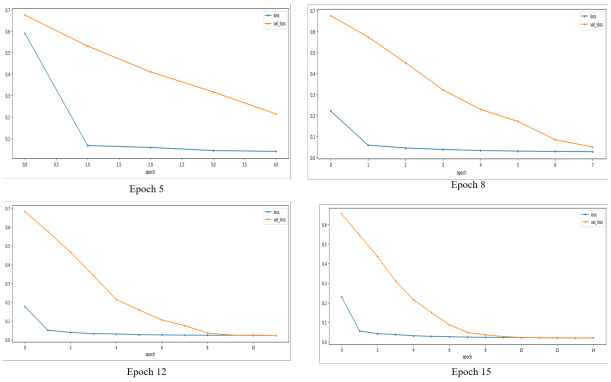


Figure 10: The loss curve of the Model at different epochs.

efficacy of classical approaches in safeguarding fine-grained textual details.

In significant difference, the proposed method – Character segmentation pipeline achieved a promising result, among the no-reference image quality metrics, as it could achieve an average NIQE, PIQE, and BRISQUE score of 15.29, 26.20, and 32.94, respectively, for N2V-Otsu images and 16.01, 25.91 and 33.27, respectively, for N2V-Hybrid images, showing a substantial improvements in visual image quality, indicating that the resultant images are visually more clearer that the original degraded images. In the proposed method, the NIQE values dropped, depicting that the denoised and binarised character patches matched the statistical behaviour of clean images. PIQE results showed improved values, reflecting a noticeable decrease in distortion. The reduction in BRISQUE values indicated that the proposed method not only eliminated the residual noise but also retained the clarity of the text, which was slightly lost in the baseline Fast NL-Means output images.

In terms of the Character segmentation pipeline, the MSE resultant values of 0.97, 1.15, and 1.02 for the N2V images were consistently low, reflecting a low error rate. The high PSNR values of 29.15, 28.6, and 29.09, respectively, explained excellent similarity with the ground truth image. The method yielded high SSIM values of 0.91, 0.90, and 0.92 for N2V-Otsu, N2V-Sauvola, and N2V-Hybrid images, respectively. The output images underwent a subjective evaluation strategy, VQE; the resultant images received high subjective ratings, with consistent scores highlighting the legibility and preservation of text. The improvement over the Baseline framework in terms of PSNR reflected patch-wise processing with CLAHE and local thresholding, which minimised the noise accumulation. SSIM results showed a superior retention of character morphology, especially when strokes were present.

From the findings and the resultant values obtained, it was clear that the N2V images showed a strong dominance over FLNM denoising images. Collectively, these results affirm that the proposed method is not only quantitatively strong but also subjectively reliable, making it a robust solution for denoising and segmentation of culturally significant, degraded palm leaf manuscripts. Figure 11 presents the visual comparison of the extracted image patches using the proposed method, specifically character segmentation and Fast NL-Means, when implementing the Baseline Approach. The table also contains the distortions extracted from the PLM scanned images. The distortions include the punch hole mark, the tearing mark, and the decay mark created because of the whitish, small ant-like insect that feeds on wood or wooden-like structures. These segmented distortions can be discarded in the next step of processing. Figure 12 illustrates the image quality metrics for Baseline and Character segmentation, as well as the proposed method. In both methods, the N2V-based methods excelled over the classical methods. With the proposed method, we further improve PSNR, SSIM, and perceptual quality values while reducing the MSE results.

Method	MSE	PSNR	NIQE	PIQE	BRISQUE	SSIM	VQE
FNLM-Otsu	1.23	27.17	44.24	20.13	44.22	89	Fair
FNLM-Sauvola	1.21	26.61	40.52	22.63	48.14	87	Bad
FNLM-Hybrid	1.14	27.91	41.13	19.18	40.56	88	Fair
N2V-Otsu	0.97	29.15	26.20	15.29	32.94	91	Excellent
N2V-Sauvola	1.15	28.6	30.19	19.27	35.89	90	Good
N2V-Hybrid	1.02	29.09	25.91	16.01	33.27	92	Excellent

Table 4: Comparison of the Results of Baseline Approach Obtained from various evaluation metrics on PLMs

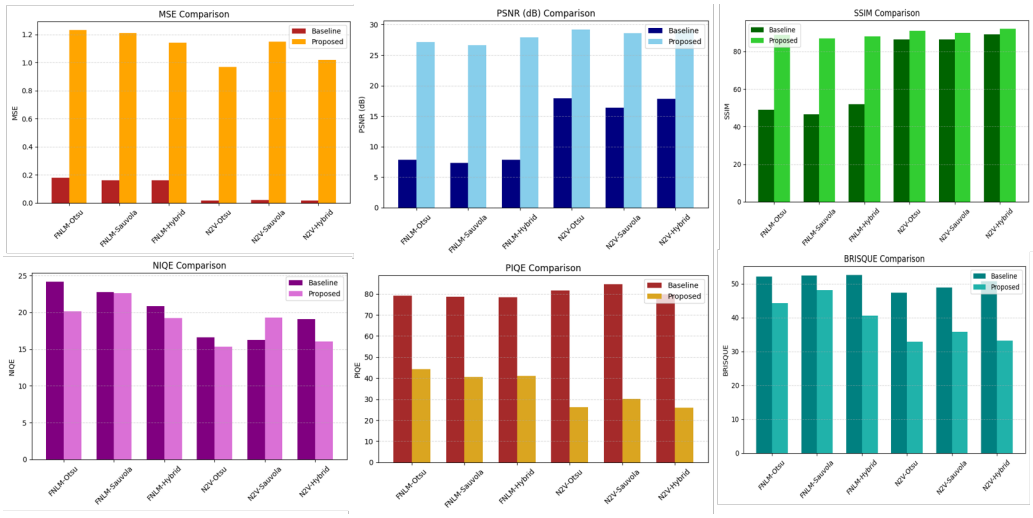


Figure 11: Samples of segmented images from Fast NL-Means-Baseline Approach and N2V-Character Segmentation .

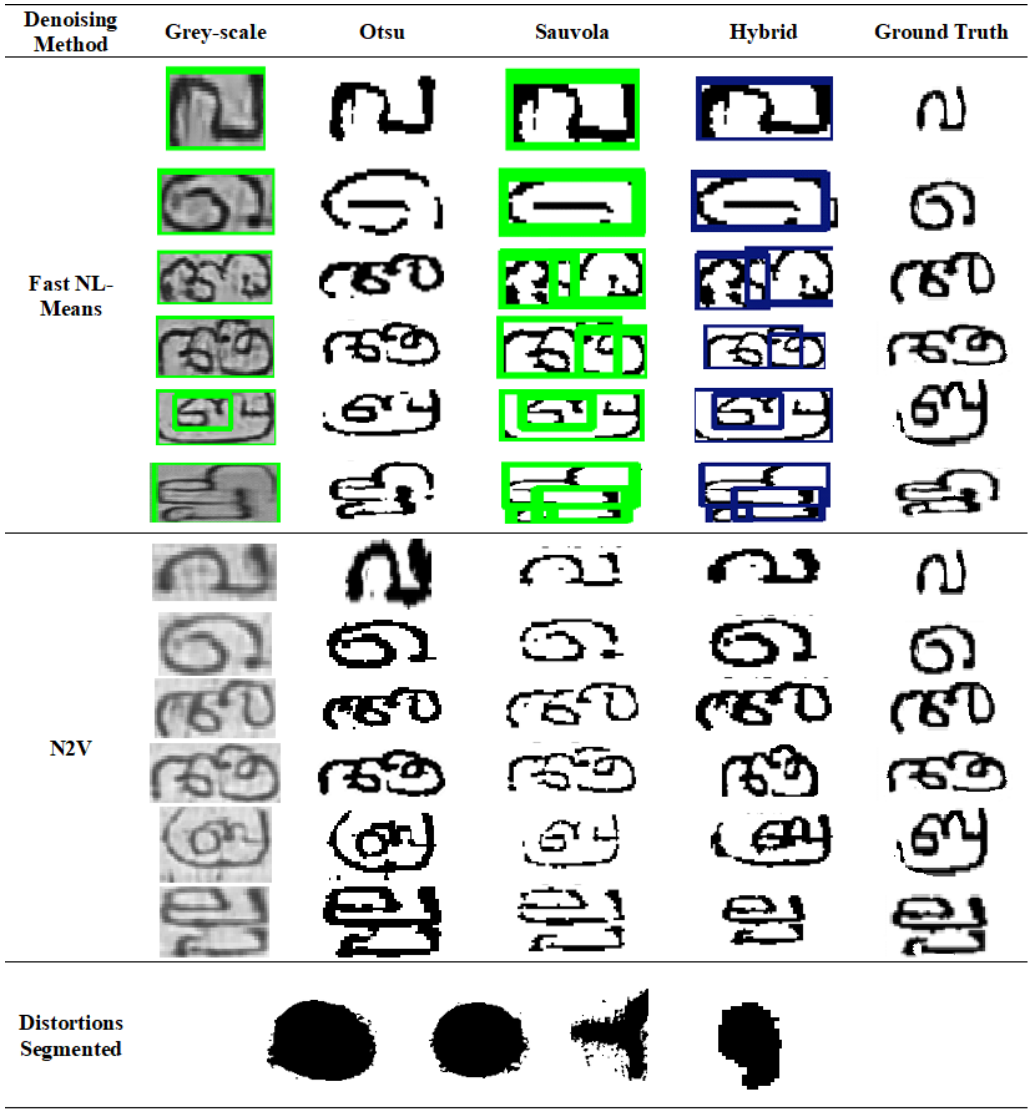


Figure 12: the Flow Diagram of the Proposed method.

7 Limitations and Future Work

Despite the clear advantages of the proposed work, it also introduces some computational constraints. Processing the individual image patches of the text will increase the computational cost in terms of both time and memory requirements, depending on the number of characters in the leaf. For highly deteriorated regions where the pixel intensity of the text and noise is identical, CLAHE occasionally amplifies the noise, resulting in an adverse effect. The Future work will focus on developing an independent deep learning model that can act appropriately on different images, suppressing noise and selectively enhancing characters by identifying regions erased due to removal distortions.

8 Conclusion

Palm Leaf Manuscripts are not just documents, but they can be called heritage artefacts that hold sacred knowledge and literary history. In the changing climatic conditions, preserving manuscripts is very difficult, as they are highly vulnerable. While many libraries and institutions have embarked on the digitisation of palm leaf manuscripts by photographing or scanning them, such efforts often result in high memory usage and computational costs without fully addressing the core goal of digital preservation. True digitisation should not merely involve image capture but must also encompass the extraction of textual content from the manuscripts and the structured storage of that information in digital formats. In the present study, we compared the segmentation quality achieved through two different methodologies (Baseline Framework and the Character Segmentation Pipeline) using two internal statistical methods for denoising degraded PLMs. From the results obtained by implementing both the methods using Fast NL Means and N2V, we could summarise that there is a clear dominance of N2V over Fast NL Means. Based on the experiments we conducted, a novel pipeline for character segmentation tailored to degraded palm leaf manuscripts was also introduced. The newly developed methodology offered several advantages; it used a self-trained deep learning approach, N2V, for denoising, which eliminated the need for clean images during the training. The choice of N2V is motivated by the inherent challenges in working with PLMs-namely, the wide variability in noise patterns, inconsistent illumination due to manual scanning or photographic capture, and the absence of uniform structure or size across the manuscript leaves. The PLMs were crafted from dried palm leaves or thin wood, resulting in natural irregularities in dimensions and surface quality. Traditional denoising techniques often fall short under such variability, whereas N2V excels as it does not require uniform or noise-free input data. Binarisation is performed on individual character patches rather than the entire image, resulting in clearer outputs and making the process independent of variation in design, size, or content density of the PLMs. Moreover, the method inherently eliminates the need for a separate punch-hole or distortion removal step, as it can be effectively excluded during the subsequent steps of processing the text. To our knowledge, this is the first application of Noise2Void on degraded PLMs. The proposed method significantly outperforms classical binarisation and segmentation techniques, achieving an average PSNR of 28.94 dB and 91% similarity with the ground truth, thus demonstrating its robustness and effectiveness in heritage document processing.

References

[1] J. Sahoo and B. Mohanty, "Digitization of Indian manuscripts heritage", *IFLA Journal* 41(3):238, 2015.

[2] D. G. Suryawanshi, M. V. Nair, and P. M. Sinha, "Improving the flexibility of palm leaf", *Restaurator* 13(1):37–46, 1992.

[3] B. Davids, "From palm leaf to book: A South Asia quest", *Printing History* 10(10):25–37, 2011.

[4] Y. Van Dyke, "Sacred leaves: The conservation and exhibition of early Buddhist manuscripts on palm leaves", *The Book and Paper Group Annual* 28:83–97, 2009.

[5] D. Wujastyk, *Indian Manuscripts*, p. 159.

[6] B. Wilson and J. M. Rice, "Palm leaf manuscripts in South Asia", *School of Information Studies – Post-doc and Student Scholarship* 8, 2019.
https://surface.syr.edu/ischoolstudents/8

[7] A. Sulaiman, K. Omar, and M. F. Nasrudin, "Degraded historical document binarization: A review on issues, challenges, techniques, and future directions", *Journal of Imaging* 5(4):48, 2019.
https://doi.org/10.3390/jimaging5040048

[8] D. Sudarsan and D. Sankar, "A novel approach for denoising palm leaf manuscripts using image gradient approximation", in *Proc. IEEE ICECA*, 2019.

[9] D. Sudarsan and D. Sankar, "A novel complete denoising solution for old Malayalam palm leaf manuscripts", *Pattern Recognition and Image Analysis* 32(1):187–204, 2022.
https://doi.org/10.1134/S1054661822010096

[10] P. Ge, P. Yu, H. Li, and H. Li, "Stroke edge based binarization algorithm for the palm leaf manuscripts", in *Proc. ICIVC*, pp. 778–782, 2017.
https://doi.org/10.1109/ICIVC.2017.7984660

[11] E. Vats, A. Hast, and P. Singh, "Automatic document image binarization using Bayesian optimization", in *Proc. International Workshop on Historical Document Imaging and Processing*, pp. 89–94, 2017.
https://doi.org/10.1145/3151509.3151520

[12] D. Unnikrishnan, D. Sudarsan, and R. Vignesh, "A novel method of absolute noise removal from the degraded palm leaf manuscripts", *Indian Journal of Traditional Knowledge* 24(6):595–605, 2025.
https://doi.org/10.56042/ijtk.v24i6.16782

[13] M. Singh and S. Indu, "Denoising of palm leaf manuscripts using Gaussian filter and conservative smoothing", *AIP Conference Proceedings* 2521(1), 2023.
https://doi.org/10.1063/5.0142237

[14] B. Gatos, I. Pratikakis, and S. J. Perantonis, "Adaptive degraded document image binarization", *Pattern Recognition* 39(3):317–327, 2006.
https://doi.org/10.1016/j.patcog.2005.09.010

[15] M. Wagdy, I. Faye, and D. Rohaya, "Document image binarization using Retinex and global thresholding", *ELCVIA Electronic Letters on Computer Vision and Image Analysis* 14(1):61–73, 2015.
https://doi.org/10.5565/rev/elcvia.648

[16] J. Jyothi and A. Malangai, "Comparative analysis of wavelet transforms in the recognition of ancient Grantha script", *International Journal of Computer Theory and Engineering* 9:235–241, 2017.
https://doi.org/10.7763/IJCTE.2017.V9.1144

[17] B. Su et al., "A restoration method using dual generative adversarial networks for Chinese ancient characters", *Visual Informatics* 6(1):26–34, 2022.

[18] L. Uzan, N. Dershowitz, and L. Wolf, "Qumran letter restoration by rotation and reflection modified PixelCNN", in *Proc. ICDAR*, vol. 1, pp. 23–29, 2017.

[19] K. Chen, M. Seuret, J. Hennebert, and R. Ingold, "Convolutional neural networks for page segmentation of historical document images", in *Proc. ICDAR*, vol. 1, pp. 965–970, 2017.

[20] J. Pastor-Pellicer, M. Z. Afzal, M. Liwicki, and M. J. Castro-Bleda, "Complete system for text line extraction using convolutional neural networks and watershed transform", in *Proc. DAS*, pp. 30–35, 2016. https://doi.org/10.1109/DAS.2016.58

[21] G. Renton et al., "Fully convolutional network with dilated convolutions for handwritten text line segmentation", *International Journal of Document Analysis and Recognition* 21:177–186, 2018.

[22] R. Alaasam, B. Kurar, and J. El-Sana, "Layout analysis on challenging historical Arabic manuscripts using Siamese network", in *Proc. ICDAR*, pp. 738–742, 2019. https://doi.org/10.1109/ICDAR.2019.00123

[23] A. Prusty, S. Aitha, A. Trivedi, and R. K. Sarvadevabhatla, "Indiscapes: Instance segmentation networks for layout parsing of historical Indic manuscripts", in *Proc. ICDAR*, pp. 999–1006, 2019.

[24] K. Watanabe et al., "Japanese character segmentation for historical handwritten official documents using fully convolutional networks", in *Proc. ICDAR*, pp. 934–940, 2019.

[25] Z. Ziran et al., "Text alignment in early printed books combining deep learning and dynamic programming", *Pattern Recognition Letters* 133:109–115, 2020.

[26] J. Cai, L. Peng, Y. Tang, C. Liu, and P. Li, "TH-GAN: Generative adversarial network based transfer learning for historical Chinese character recognition", in *Proc. ICDAR*, pp. 178–183, 2019.

[27] A. Abbas, K. Baheeja, and A. M. N. Alzubaidi, "Ancient textual restoration using deep neural networks: A literature review", in *Proc. AICCIT*, pp. 64–69, 2023. https://doi.org/10.1109/AICCIT57614.2023.10218159

[28] D. Bradley and G. Roth, "Adaptive thresholding using the integral image", *Journal of Graphics Tools* 12(2):13–21, 2007. https://doi.org/10.1080/2151237x.2007.10129236

[29] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer", *IEEE Signal Processing Letters* 20(3):209–212, 2013. https://doi.org/10.1109/LSP.2012.2227726

[30] N. Venkatanath, D. Praneeth, K. C. Bh, S. S. Channappayya, and S. S. Medasani, "Blind image quality evaluation using perception based features", in *Proc. NCC*, pp. 1–6, 2015. https://doi.org/10.1109/NCC.2015.7084921

[31] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain", *IEEE Transactions on Image Processing* 21(12):4695–4708, 2012. https://doi.org/10.1109/TIP.2012.2214050

[32] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks", *Information Processing & Management* 45(4):427–437, 2009.

[33] E. Chebbi, F. Benzarti, and H. Amiri, "An improvement of structural similarity index for image quality assessment", *Journal of Computer Science* 10(2):353–360, 2014. https://doi.org/10.3844/jcssp.2014.353.360

[34] M. Gogoi and M. Ahmed, "Image quality parameter detection: A study", *International Journal of Computer Science and Engineering* 4:110–116, 2016.

[35] C. Pang, O. C. Au, J. Dai, W. Yang, and F. Zou, "A fast NL-means method in image denoising based on the similarity of spatially sampled pixels", in *Proc. IEEE International Workshop on Multimedia Signal Processing*, pp. 1–4, 2009.

[36] A. Krull, T.-O. Buchholz, and F. Jug, "Noise2Void – Learning denoising from single noisy images", in *Proc. CVPR*, pp. 2124–2132, 2019.
https://doi.org/10.1109/CVPR.2019.00223

[37] N. Otsu, "A threshold selection method from gray-level histograms", *IEEE Transactions on Systems, Man, and Cybernetics* 9(1):62–66, 1979.

[38] J. Sauvola and M. Pietikäinen, "Adaptive document image binarization", *Pattern Recognition* 33(2):225–236, 2000.
https://doi.org/10.1016/S0031-3203(99)00055-2

[39] D. Chen and R. M. Haralick, "Recognition of degraded characters using weighted shape similarity", *Pattern Recognition* 33(7):1037–1048, 2000.
https://doi.org/10.1016/S0031-3203(99)00165-6

[40] A. M. Reza, "Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement", *Journal of VLSI Signal Processing* 38(1):35–44, 2004.
https://doi.org/10.1023/b:vlsi.0000028532.53893.82