International Journal of Engineering Trends and Technology
ISSN: 2231-5381/ https://doi.org/10.14445/22315381/IJETT-V74I4P112

Volume 74 Issue 4, 155-169, April 2026

Original Article

Kannada Stone Inscription Image Enhancement using Modified Approach for Binarization and Detection of Text Regions using Deep Learning YOLO Models

Bhagyashri Agasar[1], Gururaj Mukarambi[2]

[1,2]*Department of Computer Science, School of Computer Science, Central University of Karnataka, Kalaburgi, Karnataka, India.*

[2]*Corresponding Author: gmukarambi@gamil.com*

Received: 28 October 2025 Revised: 14 January 2026 Accepted: 06 February 2026 Published: 29 April 2026

Abstract *- In this paper, the YOLOv families, i.e., YOLOv8 to YOLOv11, are used for the detection of TextRegion and Non-textRegion in the stone inscription images. No standard datasets are available in the literature to evaluate the performance of YOLO models on inscription images. Hence, collected 370 inscription images from the ASI Departments, Museums, and ancient temples in the state of Karnataka. The tree structure approach divides each image into four parts based on the contrast of the grayscale image. Then, these four parts undergo contrast enhancement using linear scaling and CLAHE. Further, Otsu's thresholding is applied to greyscale images to convert them to binary images. Then, morphological procedures were applied, including an opening operation to remove minor noise and a closing operation to fill in gaps, resulting in a more enhanced binary image with a PSNR of 35.57 dB. Further, used a set of object detection models based on the YOLO architecture, specifically YOLOv8 to YOLOv11, to distinguish the text and non-text regions. For the experimental setup, the YOLOv (n, s, m, l, x) version has been used. The data augmentation included random horizontal flipping, brightness/contrast jittering, small rotations (±10°), mosaic augmentation, and multi-scale augmentation. For training, used the AdamW optimizer with an initial learning rate in the range 1e-4–1e-3, weight decay in the range 1e-4-1e-3, and batch sizes chosen according to GPU memory constraints; final hyperparameters were selected by empirical validation sweeps (learning rate = 1e-4, weight decay = 5e-4, batch = 32). To evaluate the performance of the YOLO models, precision, recall, F1-score, and accuracy are used. In the experimental results, YOLOv11 performs better than other versions, such as YOLOv8, v9, and v10. The YOLOv11s achieved the highest accuracy of 85.20%. This will help advance the larger goal of preserving and interpreting digitized images of stone inscriptions.*

Keywords *- Inscriptions, YOLO, CLAHE, Binary images, Deep learning.*

1. Introduction

Binarization is a crucial preprocessing step in image analysis, particularly for ancient artefacts such as stone inscriptions. These inscriptions, often characterized by complex formations and diverse contrasts, pose challenges for digital preservation and analysis [3, 2, 9, 15]. These inscriptions are fading because of weathering, erosion, and human interference. While conventional methods are crucial for cultural conservation, they often prove insufficient [19, 1, 8, 7, 20, 17, 25]. Diverse communities have various strategies to preserve this history. Hand-waxed estampages of Kannada stone inscription epigraphs are likely to be maintained at the State Department of Archaeology and the Archaeological Survey of India. Automating epigraph retrieval and digitization reduces the time and effort required for manual procedures [18]. Improvements in image processing and machine learning enable the automation of digitization and the study of inscriptions. Automatic text recognition remains challenging because inscriptions are in many languages, lighting is uneven, inscriptions are physically damaged, and contrast is low [6, 5]. To help this, epigraphers sometimes use white chalk or paint to make inscriptions easier to see before taking pictures of them [13, 16, 10]. Even yet, noise, uneven backdrops, and uneven texturing still make it hard to recognize characters. The Multi-Level Improvised Binarization Technique (MLIBT) [16] and adaptive thresholding methods that use improved binarization [24] have both helped make inscriptions easier to read. But they are unsuccessful with real-world variability and with generalizing across diverse languages and writing situations.

1.1. Motivation and Novelty

Ancient Kannada stone inscriptions pose significant challenges and create a research gap for automatic text

extraction due to uneven lighting, poor contrast, rough surfaces, and partial damage, further complicating the process. We are observing that traditional methods, such as global CLAHE, Sauvola, and Niblack techniques, often fail in these situations - they either make clear areas too bright or make faint text even more complicated to read, resulting in incomplete or noisy results. Furthermore, text region detection is required for high-dimensional inscription images that include both text and non-text regions (i.e., graphical regions), as shown in Figures 2(a), (c), and (e). Hence, the text region needs to be localized for further character recognition.

To solve these problems, this work proposes a simple processing method to identify text and non-text areas in Kannada stone inscriptions for digitization. The first step utilizes preprocessing and binarization methods to further enhance and contrast the inscription images. This process itself enhances image quality for improved analysis. Binarization is undoubtedly a key preprocessing step that separates text from background, making further processing much easier. Moreover, this separation of foreground and background elements simplifies all subsequent image analysis tasks.

Basically, after binarization, YOLO models like YOLOv8 to YOLOv11 are used to detect text and non-text regions, which is the same approach as in object detection, here applied to text regions. According to the procedure, detection is crucial for identifying the text area within the complete inscription image. Regarding text detection, this method helps locate written parts within the full image. These bounding text regions surely represent the final step of the process. Moreover, they will help in character-level segmentation and recognition for future work.

The main new contributions of this work surely include several significant findings. Moreover, these contributions significantly enhance the existing research.

Key novel contributions of this work include:

- Region-Adaptive Contrast Enhancement uses a tree-structured subdivision to enhance faint regions further while preserving details in clearer areas.
- Basically, this pipeline combines adaptive enhancement with Otsu's method and morphological operations to robustly create clean binary images.
- We are seeing that using YOLOv8–YOLOv11 deep learning methods helps to find text areas correctly in old and damaged inscriptions only.
- The dataset contains 370 real-time sample images of Kannada stone inscriptions collected through fieldwork conducted at central government institutions, the Archaeological Survey of India (ASI), Government Museums, Sites, and Ancient temples in the state of Karnataka.

We are seeing that this method helps reduce human work only by making digital copies of old inscriptions more accurate. The structure of the paper is organized as follows: The introduction is in Section 1, Related Work is given in Section 2, data collection is explained in Section 3, while the proposed framework explanation is in Section 4, while Section 5 presents the experiment results and it is discussion, Section 6 performance analysis, Section 7 presents the comparative study, Section 8 about discussion and future work and conclusion is given in Section 9.

2. Background Study

Jayanthi, J., & Maheswari, P. U. (2024) [13]. A total of 200 images of ancient Tamil stone inscriptions from the 8th-century Brihadeeswarar Temple in Tanjore were examined. To enhance the clarity of the faded text, 34 distinct image preprocessing techniques were applied, including monochrome conversion, contrast adjustment, filtering, and noise reduction. This paper explored both global and local thresholding methods for distinguishing text from background, and it was concluded that local methods, such as Niblack and Sauvola, were most effective for intricate images.

Munivel, M., & Enigo, V. S. F. (2024) [16] proposed a Multi-Level Improvised Binarization Technique (MLIBT) for Tamizhi inscriptions. This approach helps address challenges such as uneven lighting, similar backgrounds, and limited datasets. A total of 396 annotated images, derived from 150 inscriptions encompassing cave paintings and memorial stones, were drawn from Iravatham Mahadevan's book, Early Tamil Epigraphy. The MLIBT achieved a binarization accuracy of about 92.19%, which was better than that of traditional thresholding methods across several image quality metrics.

Bhuvaneswari, S., & Kathiravan, K. (2024). [3] The study describes how to interpret Tamil temple inscriptions and understand their context. Advanced preprocessing approaches, including adaptive binarisation, orientation correction, contrast enhancement, and noise reduction, improve photographs. We utilize the Stroke Width Transform to separate characters and CNNs with Vision Transformers (ViT) and multi-head attention to recognize them. After training a ViT model for a specific task, transfer learning allows you to adapt it. The 100 high-resolution images of old writings from different times and nations make this collection even more interesting. It performed well with an F1-score of 95.17%, a recall rate of 95.05%, and a character recognition accuracy of 97.25%. The program corrects 95% of dates, 94% of context-specific phrases, and 98.92% of relevant historical keywords. This illustrates its versatility for research in history and epigraphy. This study outlines the methodology for investigating Tamil temple inscriptions and interpreting their characters. A blend of advanced preprocessing techniques improves images. These include adaptive binarization, noise reduction, contrast enhancement, and orientation correction.

Here, this work uses CNNs, Vision Transformers (ViT), multi-head attention, and the Stroke Width Transform to distinguish things and differentiate characters. Transfer learning adapts pre-trained ViT models to the task. The dataset contains 100 high-resolution images of historical inscriptions from diverse eras and locations. The experiment measured recall of 95.05%, F1-score of 95.17%, and character recognition accuracy of 97.25%. The model's 95% date recognition, 94% context-specific word recognition, and 98.92% recognition of relevant historical phrases aid epigraphic and historical researchers.

Researchers Rajnish et al. (2023) [17] examined two approaches to adaptive thresholding: fuzzy entropy-based and modified bi-level entropy. We also looked at pictures of stone carvings. to make fifty images of handwritten Brahmi scripts from Google Images and the ASI easier to read and of better quality. Advanced methods such as GAN-based denoising and image inpainting were used to restore damaged regions. Features were extracted using Histogram of Oriented Gradients (HOG) and Zernike Moments. Sukanthi et al. (2021) [27] developed a new thresholding method, called Modified Bi-Level Thresholding (MBET). They used PSNR and standard deviation as measures to compare their effectiveness to other thresholding methods. The collection contains various images of stone inscriptions discovered on land and in water. The results show that the MBET method increases PSNR by 49% and SD by 39%.

Chandrakala et al. (2021) [7] discuss several approaches to improve and study older handwritten texts. Normalization, hyperspectral imaging, directional wavelet transform, curvelet, shearlet, median filtering, rapid independent component analysis based on natural gradients, and the Retinex enhancement method are some of the techniques used. This study uses the EHHKSI dataset, which contains 14 images of ancient Kannada stone carvings made by local artisans by hand. Each image is 2330 x 3658 pixels. Thus, the dataset has 14 images. The document describes the TVR enhancement results for the EHHKSI dataset. The 3 TVR technique performed well on this dataset, with a Global Contrast Factor (GCF) of 1.1528 and character segmentation accuracy of 79.2%.

Bhat and Seshikala (2018) [2] utilize phase congruency features and Gaussian mixture models to enhance and segment old inscription images. Log-Gabor wavelets are used to remove noise from images, detect edges, and assess phase congruency. Binarization eliminates background noise by combining the components using a Gaussian model. The stone and palm leaf epigraph databases were used. The collection contained more than 50images of Kannada writing. Finally, the testing revealed that the introduced method significantly improved the binarization of the inscription image. The novel method was evaluated against five well-known local binarization algorithms: Niblack, Bernsen, Bradley, Sauvola, and Wolf. According to Carrero-Pazos (2018) and Liu and Ma (2016), [6, 14, 25] digital imaging techniques such as photogrammetry, Structure-from-Motion (SfM), and Digital Image Modelling (DIM) can successfully capture the degradation of Roman inscriptions. These methods include acquiring images from multiple views and then using software to create a 3D model. The paper talks about digitizing archaeological and epigraphic materials for Zaragoza's "Virtual Museum of Los Ba¸nales." When working in the field, in bad weather, or with limited time, DIM methods are better than manual methods because they are faster and more accurate. DIM methods in epigraphic research improve archaeology by producing more accurate and reliable data that can be shared digitally.

An author move this result into higher-precision metrology: they use calibrated close-range scanning (structured light / high-resolution laser or optical triangulation) and mesh denoising to recover faint glyph profiles on Greek inscriptions, demonstrating that micron-level surface detail available from accurate 3-D scans can reveal stroke geometry invisible in conventional images.

More recently, Hu et al. (2025) [26] propose an RMSE-constrained Laplacian smoothing and collision-detection pipeline that terminates smoothing adaptively (using RMSE as a stop criterion) and then computes depth-to-fitted-surface values to generate measurable digital rubbings from fine 3-D models; the paper reports feasibility and quantitative control over smoothing iterations (Hu et al., 2025). (Pei 2023) [27] Hyperspectral Imaging (HSI) and multispectral methods are now established tools for revealing low-contrast strokes, pigment traces, and material heterogeneities on stone surfaces that RGB images miss.

Yang et al. (2025) [28] publish the HD-SC-1 hyperspectral dataset of stone cultural relics (Dazu Rock Carvings), which contains 28,193 hyperspectral layers covering 400–1000 nm at 2048×2046 pixels per frame. The dataset is explicitly intended to support supervised learning and restoration experiments where spectral signatures correlate with deterioration modes. Advances in deep learning and script-specific modelling have substantially raised recognition rates for regional and ancient scripts on stone. However, performance still depends heavily on dataset size and domain adaptation [29].

Bhuvaneswari & Kathiravan (2024) [30] propose a deep-learning pipeline combining a ViT/transfer-learning backbone with Stroke Width Transform(SWT) and report strong metrics: precision 97.25%, recall 95.05%, F1 95.17%, and recognition rate figures around 96 to 99% on key terms and frequent lexemes in their temple-inscription corpus, showing that transformer transfer learning and stroke-aware preprocessing are effective for degraded epigraphic images.

3. Data Collection

The dataset comprises 370 real-time sample images of Kannada stone inscriptions collected through fieldwork conducted at central government institutions, including the Archaeological Survey of India (ASI), Government Museums, Sites, and Ancient temples in the state of Karnataka. The DSLR camera was used to capture these images, as illustrated in Figures 2(a), 2(c), and 2(e).

3.1. Dataset Description

We tested our binarization and YOLO text detection system on 370 high-quality images of ancient Indian stone inscriptions. The same framework was evaluated on this custom dataset to assess its performance. Each inscription image was captured using DSLR and mobile phone cameras in outdoor light conditions only, resulting in significant changes in lighting, surface roughness, and texture damage. The original image resolutions range from 1091×901px to 6000×4032px, which further enables clear visibility of character strokes and surface textures themselves. The dataset has been resized to 960×960 pixels to help the model learn properly, and we kept the original shape by adding padding so the images do not get distorted. This resizing method was chosen because early tests showed that smaller sizes, such as 640×640, were likely to cause loss of small inscription details, especially in weakly carved or damaged areas.

Furthermore, larger images, typically around 960-1280 pixels, maintain the carved text structure while making YOLO training computationally feasible. Each image was surely marked by hand using a YOLO-based format to separate text areas from non-text areas. Moreover, this manual annotation clearly distinguished between different regions. The dataset surely includes inscriptions carved on granite, sandstone, laterite, and marble stones. To better understand light and surface changes, the dataset was grouped using the K-means method based on average brightness, texture differences, and image sharpness, yielding three visual groups.

- The low-light-rough category contains 206 images that surely show poor lighting conditions with mild shadows. Moreover, these images exhibit uneven surface reflectance, making them challenging to analyze.
- As per the collection, 116 images show well-lit inscriptions with clear character boundaries regarding the highlight-soft category.
- Moreover, this category surely includes 49 images that show moderate contrast levels. Moreover, these images show clear signs of erosion or granular texture.

This categorization illustrates the range of real-world situations encountered at ancient historical sites and helps in the systematic evaluation of method performance across these situations. This further shows the variety found in old Indian stone inscriptions. The collection work is still ongoing because site access rules and conservation guidelines further restrict detailed digital documentation.

4. Proposed Framework

The proposed framework introduces domain-specific methodological innovations for robust text region detection in Kannada stone inscriptions by combining tree-based image processing with YOLO-based deep learning. Regarding the method, it uses both preprocessing and detection techniques together. Each grayscale image is further divided into smaller regions based on local contrast, calculated as the standard deviation of pixel intensities. Regions with low contrast or small size remain the same without further division, while other regions are split into four quadrants. To make low-contrast areas more visible, we surely apply adaptive contrast enhancement methods. Moreover, moderate-contrast regions use the CLAHE technique, while high-contrast areas use linear scaling to expand their intensity range. After enhancement, Otsu's method essentially converts the grayscale image to a binary form, and morphological operations perform a similar task of removing noise and filling gaps to produce a clean binary image.

The processed image is surely fed into YOLOv8–YOLOv11 models that are trained to find and separate Text and Non-Text areas. Moreover, these models can clearly identify different regions in the image. YOLO models are enhanced using an Optuna-based Bayesian method for parameter tuning, and their performance is evaluated using mAP, precision, recall, and FPS measures in accordance with standard evaluation criteria.

The optimization focuses on hyperparameter tuning to improve model performance. According to this combined method, the system yields better results for various regions and accurately identifies text, even in complex stone images that are damaged or contain excessive writing. Based on preprocessing and deep learning methods, the framework provides robust, fast text detection, suitable for the study of old documents and for inscription digitization.

Figure 1 shows the same framework we propose for detecting Kannada inscription text. The pipeline consists of two main stages, as clearly explained in Sections 4.1 and 4.2, along with the proposed framework shown in Figure 1.

4.1. Tree-Structured Image Binarization

This method divides images into smaller parts using a tree structure, where each division is based on the same local contrast patterns. Each part is enhanced and converted to binary using the same combination of CLAHE, linear scaling, and Otsu's method, and morphological processing cleans up the resulting binary. The proposed tree-structured image binarization method uses a hierarchical approach to convert images to black and white.

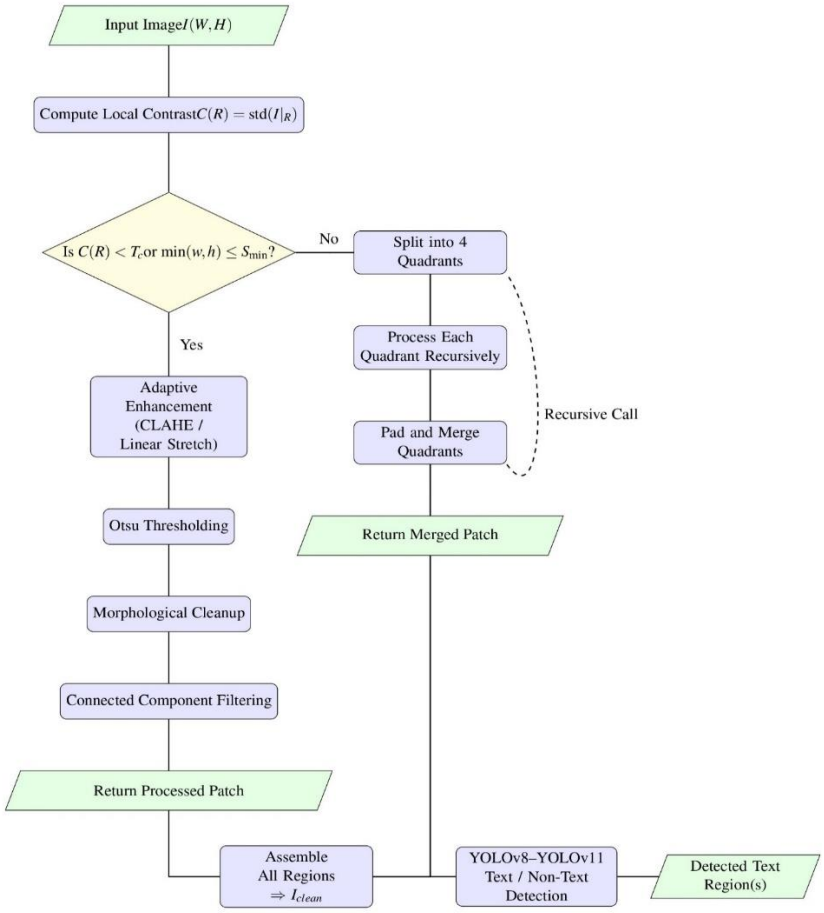


Fig. 1 Architecture of the proposed framework pipeline tree-structured image Binarization and YOLO-based TextRegion detection

4.2. Tree-Structured Recursive Partitioning

Tree-structured recursive partitioning is a method that splits data into smaller parts using a tree-like structure. Then, given a grayscale image I with width W and height H, each region is shown only as a smaller matrix inside it.

Now, *I(l, b, w, h)* shows a region where *(l, b)* gives the top-left corner position, and *w* and *h* are the area's width and height. Moreover, this notation helps define the rectangular region simply. The local contrast *C* of a region is surely calculated as the standard deviation of pixel intensities. Moreover, this method provides a simple way to measure the extent of brightness variation within that specific area.

4.3. Local Contrast

$$C = \sigma(I) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(I_k - \mu)^2}$$

Here, I_k denotes the intensity value of the k^{th}pixel, $\mu = \frac{1}{N}\sum_{k=1}^{N} I_k$ is the mean pixel intensity, and $N = w \times h$represents the total number of pixels in the region.

The image is recursively divided into four subregions based on the local contrast value C. A region is not subdivided further if any of the following conditions are satisfied:

Stopping Condition

$(w \le \text{min\_size}) \ \vee \ (h \le \text{min\_size}) \ \vee \ (C <$ contrast_threshold)

If none of these conditions holds, the region is split into four equal quadrants [4]:

Quadrant Definition

$$\text{Quad}_{i,j} = [l + i\tfrac{w}{2},\ b + j\tfrac{h}{2},\ \tfrac{w}{2},\ \tfrac{h}{2}], i,j \in \{0,1\}$$

Here, $i = 0$and $i = 1$correspond to the left and right halves, respectively, while $j = 0$and $j = 1$correspond to the bottom and top halves.

4.4. Adaptive Contrast Enhancement

Adaptive contrast enhancement actually improves image quality by automatically adjusting brightness levels. This method definitely makes dark and bright areas in the image

more visible. To improve visibility in low-contrast areas, pixel brightness values are likely scaled non-linearly. Moreover, this method helps enhance image details that are otherwise difficult to see. The enhancement method is chosen based on the contrast value \$C\$ for image processing.

1. Contrast-Limited Adaptive Histogram Equalization (CLAHE)

For regions with moderate contrast:

$30 < C < 60$

CLAHE is applied to enhance local details while preventing excessive noise amplification:

CLAHE Enhancement

$I_{\text{enhanced}} = \text{CLAHE}(I)$

CLAHE [23] redistributes pixel intensities within local neighbourhoods, improving contrast while maintaining noise control.

2. Linear Intensity Scaling

For regions with high contrast:

$C \geq 60$

A simple linear scaling operation [12] is applied to expand the dynamic range:

Linear Scaling

$I_{\text{scaled}} = \alpha \cdot I, \alpha = 1.5$

This method is computationally efficient and enhances already well-defined regions without altering their structure.

Together, CLAHE and linear scaling provide an adaptive framework that adjusts contrast differently across image regions, ensuring balanced enhancement throughout the image.

4.5. Thresholding Using Otsu's Method

After contrast enhancement, Otsu's method is used to get a binary image. This method determines the optimal threshold that best separates foreground and background regions.

Otsu's technique maximizes the between-class variance:

Between-Class Variance

$\sigma^2_{\text{between}}(\tau) = P_1(\tau)P_2(\tau)(\mu_1(\tau) - \mu_2(\tau))^2$

Here:

- $P_1(\tau)$and $P_2(\tau)$: The probabilities of the two classes separated by the threshold τ,
- $\mu_1(\tau)$ and $\mu_2(\tau)$: The corresponding class means.

The optimal threshold is defined as:

Optimal Threshold

$T^* = \arg\max_T \sigma^2_{\text{between}}(T)$

This ensures maximum separability between foreground and background regions.

4.6. Morphological Post-Processing

Morphological operations are applied to the binary image to remove noise and improve structural continuity. Specifically:

- Small object removal eliminates isolated noisy pixels.
- Hole-filling repairs small gaps within detected regions.

Noise Removal

$I_{\text{clean}} = \text{RemoveSmallObjects}(I_{\text{binary}}, \text{min\_size})$

Hole Filling

$I_{\text{clean}} = \text{RemoveSmallHoles}(I_{\text{binary}}, \text{min\_size})$

These operations significantly improve the visual quality and reliability of the binary image.

4.7. Connected Component Filtering

Let $\{C_i\}$ represent the connected components in the cleaned binary image, with each element having an area A_i. Components are retained only if their area satisfies:

Area Constraint

$A_i \geq \alpha \cdot (H \times W)$

where $\alpha = 0.0005$. It is a small area threshold.

The final binary image is defined as:

Final Binary Image

$$B_{\text{final}}(x,y) = \begin{cases} 1, & (x,y) \in \bigcup_{A_i \geq \alpha HW} C_i \\ 0, & \text{otherwise} \end{cases}$$

Overall Processing Framework

The complete processing pipeline is summarized as:

Overall Model

$$B_{\text{final}} = \text{Clean}(\text{RecProcess}(I, T_c, S_{\min}))$$

where:

$$\text{RecProcess}(I, T_c, S_{\min}) = P(\Omega)$$

These post-processing steps enhance object boundary continuity and produce a clean, noise-free binary image. The tree-structured recursive approach adapts processing to different image regions, yielding superior black-and-white segmentation results, as illustrated in Figures 2(b) and 2(d).

4.8. Deep Learning Model(YOLO family) based Detection of TextRegion and Non-textRegion in the Kannada Inscribed Images

The author set up the deep learning models based on the You Only Look Once (YOLO) architecture [21, 22], from YOLOv8 to YOLOv11, to detect and separate TextRegion and Non-textRegion in binarized stone inscription images. We chose these models because they can recognize objects in real time and are strong at handling dense, complex layouts like those found in historical sites. Using annotated bounding boxes, each model in the YOLOv8–YOLOv11 series has learned how to find and mark both text and non-text areas. All models are trained on a small dataset.

4.8.1. Experimental Setup

Data annotation: The detection framework was evaluated on a custom dataset of 370 high-resolution stone-inscription images. Images were resized with aspect-ratio preserving padding to 960×960 pixels for training to preserve fine inscription details. Manual annotations in YOLO-based format using the Computer Vision Annotation Tool (CVAT) [32]. Experiments compare all YOLO variants v8–v11. Models were initialized with corresponding pretrained weights for transfer learning. The dataset was split into training, validation, and test sets at 8:1:1. We used the validation set to monitor training performance and guide model selection. Training used the AdamW optimizer with an initial learning rate in the range 1e-4–1e-3, weight decay in the range 1e-4–1e-3, and batch sizes chosen according to GPU memory constraints; final hyperparameters were selected by empirical validation sweeps (learning rate = 1e-4, weight decay = 5e-4, batch = 32). Data augmentation included random horizontal flipping, brightness/contrast jittering, small rotations (±10°), mosaic augmentation, and multi-scale augmentation. All models were trained until convergence with early stopping (patience=5) and evaluated on a held-out test set. Evaluation metrics comprise mAP@0.5 (mAP50), precision, recall, inference throughput (FPS measured on a fixed GPU at test resolution), model size (parameters), and approximate FLOPs. GPU specifications, the batch size during inference, and the input size for evaluation are reported alongside the results.

Table 1. Evaluated YOLO versions and architectural highlights

Version	Variants	Core Features
YOLOv8	n, s, m, l, x	CSPDarkNet backbone; anchor-free head
YOLOv9	t, s, m, c, e	GELAN backbone; Programmable Gradient Flow (PGF)
YOLOv10	n, s, m, l, x	Efficient unified head; end-to-end alignment
YOLOv11	n, s, m, l, x	Transformer backbone; multi-scale fusion

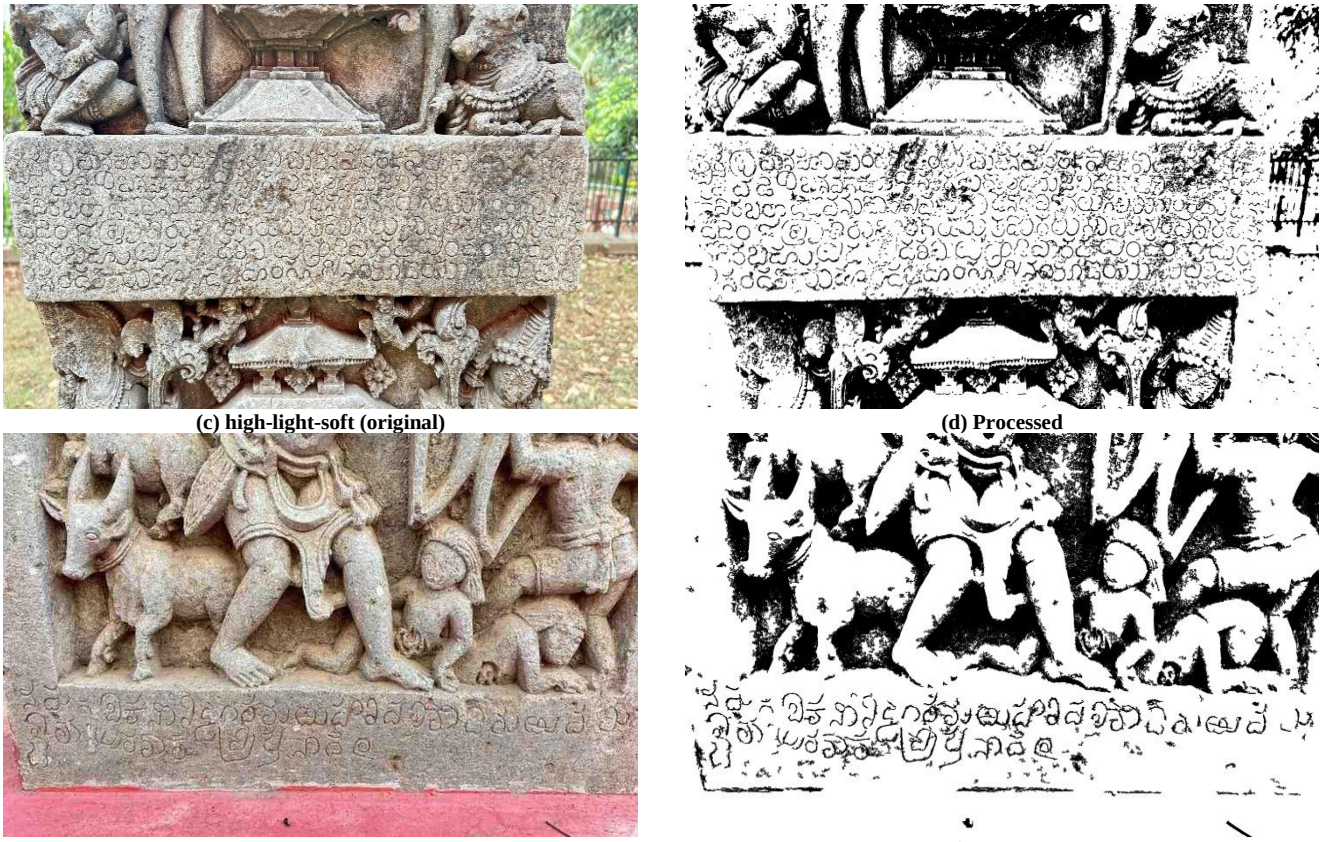
4.8.2. Models and Architectures

We evaluated the full range of YOLO versions and their respective model variants as released by Ultralytics. Each model was initialized with pretrained COCO weights and fine-tuned on the custom dataset. Table 1 summarises the models and their key architectural features.

5. Experimental Results

(a) midtone-grainy (original)

(b) Processed



(c) high-light-soft (original) (d) Processed

(e)low-light-rough(original) (f)processed

Fig. 2 The proposed technique has been utilized to process real-time images (a, c, e), resulting in the acquisition of effective binarized images(b,d,f)

We evaluate the stone inscription image analysis pipeline, which incorporates image binarization, TextRegion and Non-textRegion detection, and performance evaluation utilizing standard metrics.

The procedure commences with the binarization of the inscription images shown in Figure 2(b), (d), and (f), the first step in the process. Preprocessing enhances inscription images so that the text can be accurately read against background noise.

After that, the binarized images are fed to the YOLOv models as input to detect two target classes: TextRegion and Non-textRegion. The predicted bounding boxes over validation images are shown in Figure 3, which presents the qualitative results: predictions for TextRegions and Non-textRegions. The model can accurately identify the TextRegion (blue) and the Non-textRegion (cyan) using confidence scores.

The Precision-Recall (PR) curve (Figure 4(a) demonstrates performance for the TextRegion class (AP = 0.986) and marginally regional performance for Non-textRegion (AP = 0.912), indicating some ambiguity with background elements. The F1-confidence curve (Figure 4(b) indicates an ideal with a confidence level of 0.477 and an F1 score of 0.93, properly balancing precision and recall. The training dynamics Figure 4(c) demonstrates a constant reduction in training and validation losses, which show that learning has been successful (box, classification, and distribution focal losses).

Achieving 0.5 in the Accuracy, Recall, and mAP evaluation metrics shows consistent growth across epochs. Figure 4(d) shows that the well-performed YOLOv11s model separates TextRegion, Non-TextRegion, and background. Diagonal entries show samples that have been correctly classified.

The model achieved predictions of text regions and non-text regions, with misclassifications occurring between Background and the other two classes. Overall accuracy of 85.20%, indicating strong performance in identifying TextRegions and Non-TextRegions.

The pipeline demonstrates strong effectiveness in segmenting components of historical inscriptions, making it well-suited for subsequent tasks such as OCR or layout analysis.

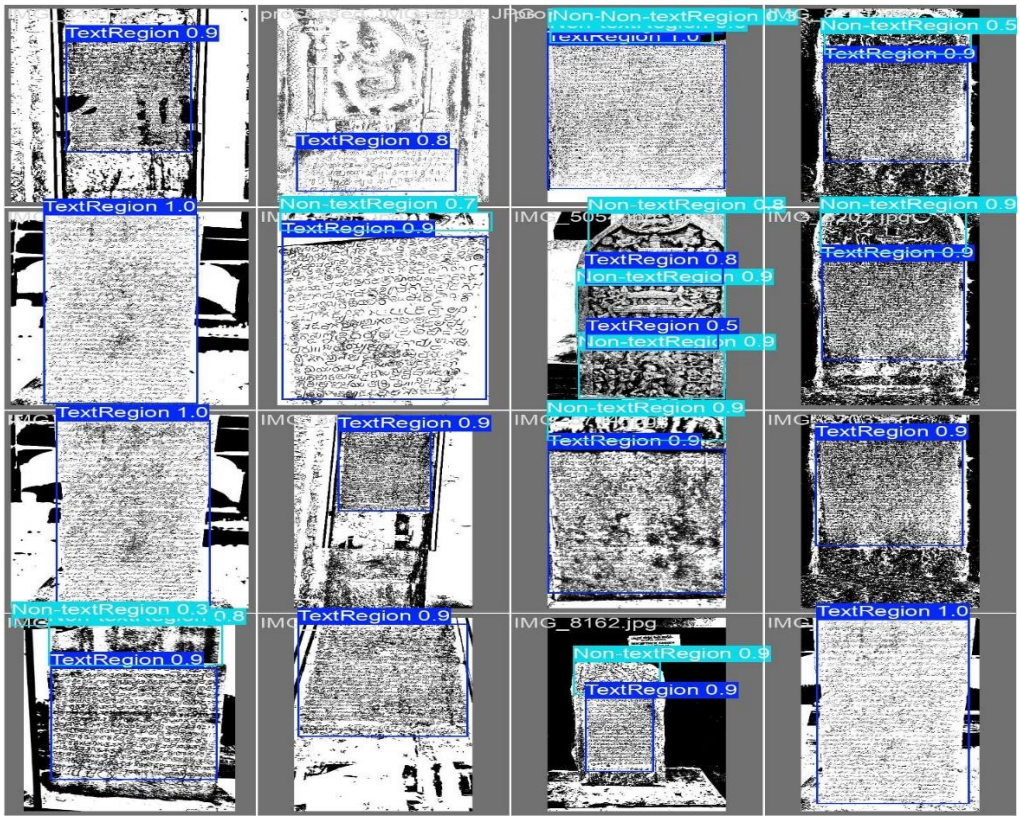


Fig. 3 YOLO predictions of TextRegion

6. Performance Analysis

6.1. Evaluation Metrics for Image Binarization

A quantitative evaluation of the proposed image processing method employed two widely used metrics: *MSE* and *PSNR*. These metrics provide objective measures of the similarity between the original and processed images.

Mean Squared Error (MSE) measures the average squared difference between the original image and the processed image. A lower MSE value indicates that the improved image is more similar to the original image. It is defined as:

$$MSE = \frac{1}{MN}\sum_{k=1}^{M}\sum_{l=1}^{N}[\, I(k,l) - K(k,l)]^2$$

where $I(k,l)$and $K(k,l)$indicate the pixel intensities of the original image and the processed image at the position (k, l), respectively, and $M$and Ndenote the dimensions of the image.

Peak Signal-to-Noise Ratio (PSNR) is a logarithmic measure that expresses the ratio between the maximum possible pixel intensity and the power of the noise affecting image fidelity. The high PSNR value indicates better image quality and greater similarity to the original image. PSNR is said to be:

$$PSNR = 20\,Log\left[10, \frac{MAX_I}{Sqrt[MSE]}\right]$$

where MAX_I says the maximum possible pixel value of the image (basically 255 for 8-bit).

The relationship between PSNR and MSE indicates that as the reconstruction error decreases, PSNR increases, reflecting improved image quality.

6.2. Evaluation Metrics for TextRegion Detection

The parameters Tp, Tn, Fp, and Fn, four conventional classification metrics extracted from the confusion matrix, evaluate how effectively the YOLO-based classifier distinguishes between TextRegions and Non-textRegions in stone Kannada inscribed images.

Accuracy is an indicator of how many regions, both TextRegion and Non-textRegion, were correctly computed as the proportion of all areas.

$$Accuracy = \frac{Tp + Tn}{Tp + Tn + Fp + Fn}$$

Although accuracy provides an overall perspective of performance, it can be inaccurate when there is a mismatch across classes.

Precision is described as the

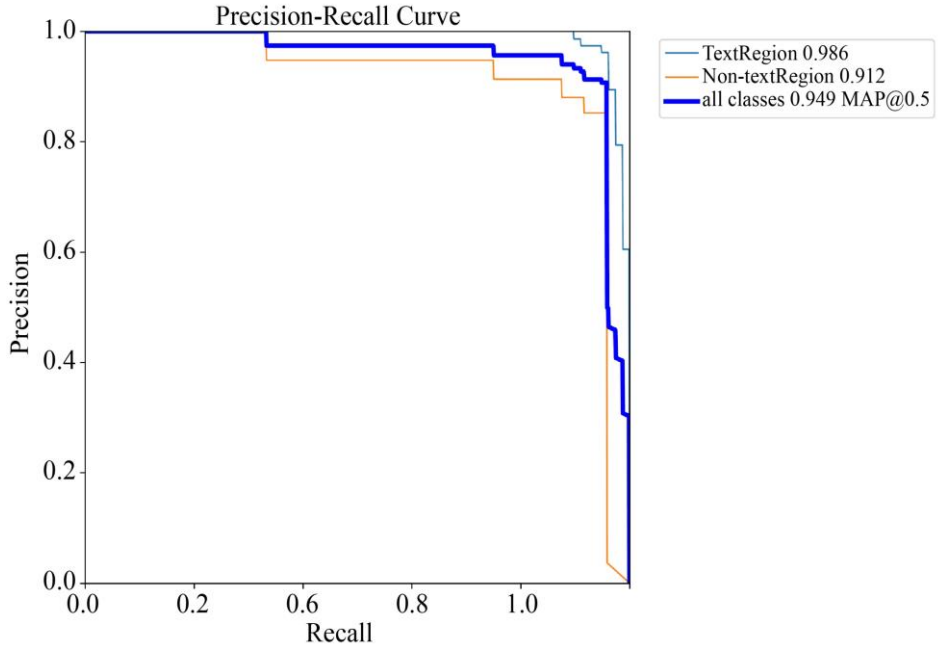
$$Precision = \frac{Tp}{Tp + Fp}$$

In a strategy to reduce misidentifications in document analysis, high precision means that the majority of regions identified as text actually contain text.

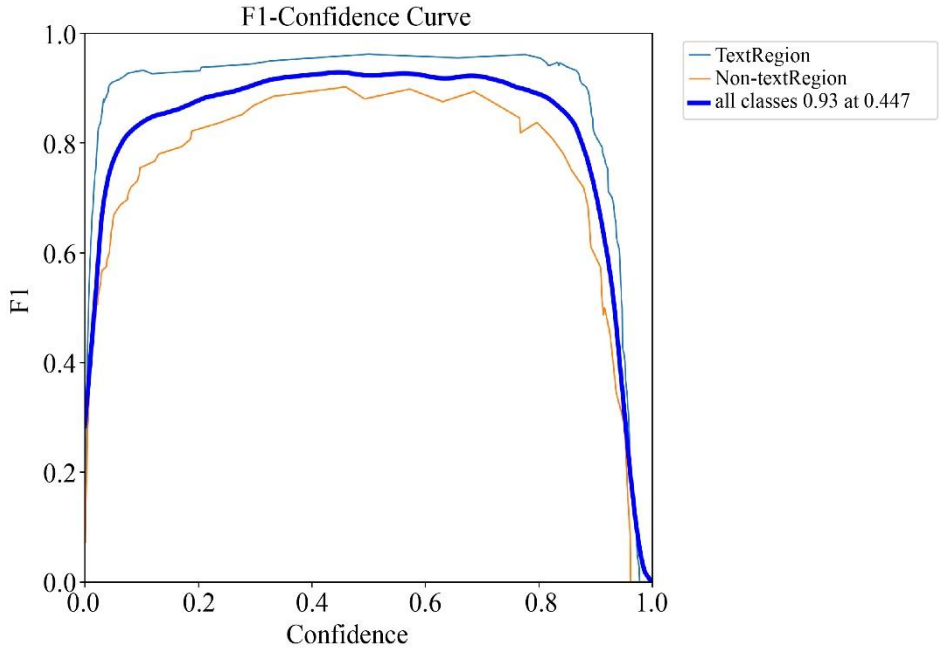
Recall, also known as Sensitivity, is a metric that assesses how well the model identifies all true text regions.

$$Recall = \frac{Tp}{Tp + Fn}$$

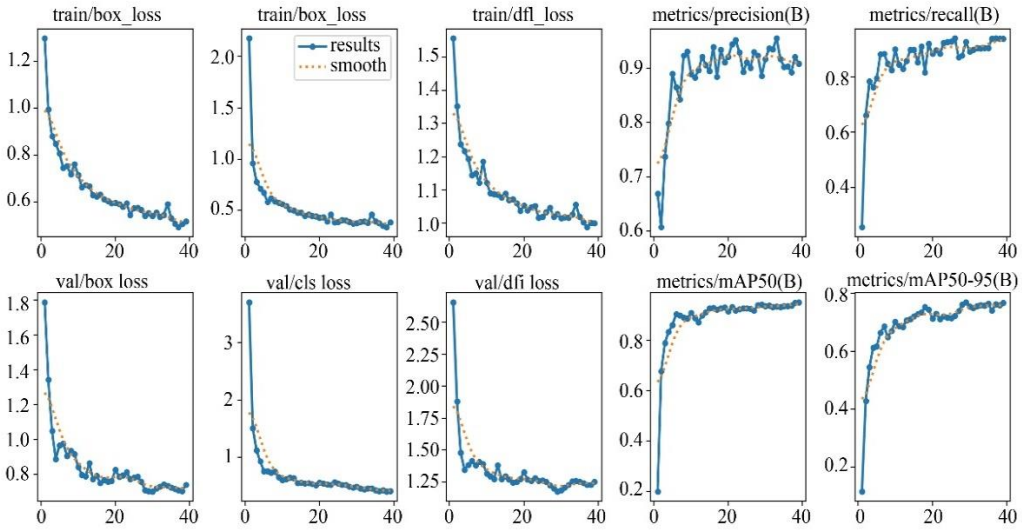
The model can find the majority of the text, regardless of background or font size, if it has a high recall score.



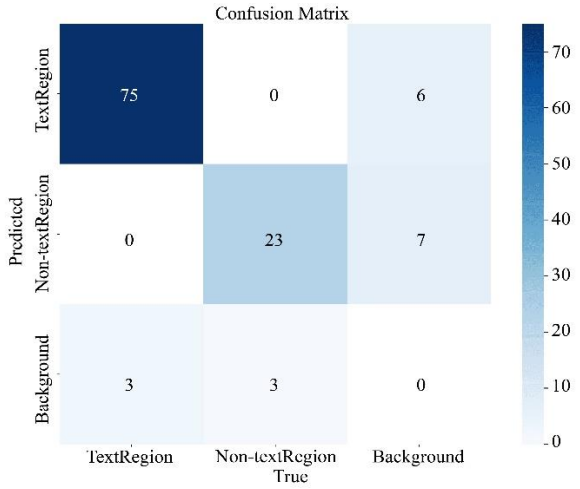
(a) Precision-Recall (PR)



(b) F1 Score



(c) Visual detection results of text and non-text regions using YOLO.



(d) Confusion matrix

Fig. 4 Comprehensive evaluation of the YOLOv11s model, including (a) PR curve showing the model's balance between recall and precision, (b) F1 curve illustrating the harmonic mean of recall and precision over different thresholds. (c) visual results, and (d) confusion matrix for the YOLOv11s model predictions on test data.

Table 2. Comparison of performance evaluation metrics on YOLOv8 to YOLOv11 models.

Method	Precision	Recall	mAP@50	Accuracy
YOLOv11n	0.936	0.919	0.953	0.820
YOLOv11s	0.899	0.959	0.959	0.852
YOLOv11m	0.956	0.905	0.931	0.759
YOLOv11l	0.916	0.943	0.937	0.840
YOLOv11x	0.945	0.931	0.951	0.742
YOLOv10n	0.861	0.831	0.950	0.737
YOLOv10s	0.926	0.925	0.950	0.873
YOLOv10m	0.943	0.871	0.955	0.790
YOLOv10l	0.916	0.924	0.949	0.843
YOLOv10b	0.952	0.825	0.909	0.787
YOLOv10x	0.920	0.919	0.952	0.888
YOLOv9t	0.906	0.947	0.944	0.859
YOLOv9s	0.893	0.933	0.939	0.852

YOLOv9m	0.903	0.940	0.927	0.795
YOLOv9c	0.929	0.932	0.961	0.862
YOLOv9e	0.928	0.945	0.966	0.801
YOLOv8n	0.903	0.955	0.952	0.765
YOLOv8s	0.929	0.942	0.922	0.765
YOLOv8m	0.886	0.940	0.935	0.738
YOLOv8l	0.892	0.951	0.935	0.848
YOLOv8x	0.899	0.907	0.944	0.827

These measures work together to show how well the YOLO model can find TextRegions and Non-textRegions. Table 2 shows a detailed evaluation of the efficiency of various versions of YOLOv8, YOLOv9, YOLOv10, and the new YOLOv11 models on four necessary measures for assessment: mAP@50, recall, precision, and Accuracy. Observations: The YOLOv11m model achieved the highest accuracy of 0.956, indicating it could make correct predictions. YOLOv11s outperformed the others in recall (0.959) and mAP@50 (0.959), indicating its efficacy in accurately identifying relevant items.

YOLOv11s exhibits the highest overall accuracy among the YOLOv11 variants, with a score of 0.852. This suggests that it excels across all metrics. Among all the models, the YOLOv11s model achieved the highest accuracy of 85.20%. YOLOv10s and YOLOv10l, respectively, achieved impressive accuracy values of 0.873 and 0.843. With the highest mAP@50 (0.966) and the most robust recall (0.945), the YOLOv9e model stands out as an effective.

Both the YOLOv9c and YOLOv9t models achieved accuracies greater than 0.85. Despite being old, YOLOv8 models still achieve high accuracy; among them, YOLOv8l achieved 84.80%. Figure 4 summarizes the evaluation of the YOLOv11s model using the test dataset and includes both numerical indicators and example detection outputs. As shown in Figure 4(a), the Precision–Recall curve suggests that the model maintains a reasonable trade-off between precision and recall across different confidence thresholds, implying controlled false-positive and false-negative rates. Figure. 4(b) presents the corresponding F1-score curve, where a clear maximum can be observed, indicating a threshold at which precision and recall are jointly balanced. Figure 3(c) shows that the model can generally identify and localize targets across different scales, orientations, and backgrounds. The confusion matrix in Figure 3(d) shows class-wise predictions, with most correct classifications along the diagonal and a small amount of misclassification. Finally, the results show stable performance on the test data, though some limitations remain. In view of the Figure. 5, four YOLO model variants are considered: YOLOv8l, YOLOv9c, YOLOv10x, and YOLOv11s for ROC calculation. In this scenario, the True Positive Rate (Recall) versus the False Positive Rate, the dotted line shows random guessing (AUC = 0.5). The YOLOv11s (AUC = 0.798), next YOLOv8l (0.774), YOLOv9c (0.759), and finally, YOLOv10x's performance (0.508). It is easier to distinguish between positives and negatives when the AUC is higher. Hence, YOLOv11s is the best.

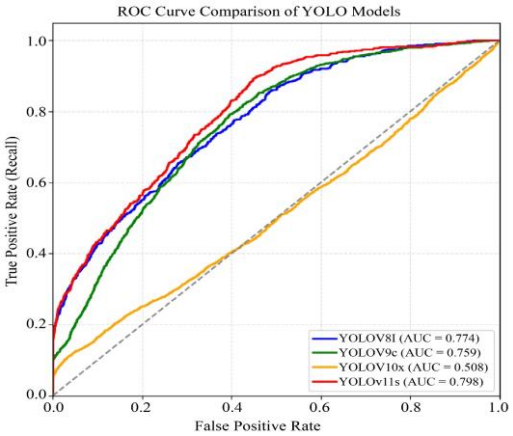


Fig. 5 ROC curves for four variants of YOLO versions

7. Comparative Analysis

7.1. Detecting TextRegion

An exhaustive review of the literature is carried out on Kannada Stone Inscription images. It is found that there is no direct work on the detection of TextRegion and Non-textRegions. Hence, the YOLO detection model is not compared in this work. Based on the preliminary survey, we claim that this is the first of its kind, i.e., the detection of

TextRegions and Non-textRegions in the field of Kannada language inscriptions.

Table 3. Comparison of methods in related domains based on the detection of TextRegion and Non-textRegion.

Domain / Method	Target Material	Text / Non-text Region
EAST (Scene Text)	Printed / Scene	Yes
CRAFT	Natural Scenes	Yes
Historical Manuscripts	Paper	Yes
YOLO Method (Proposed in this paper)	Stone Inscription	Yes

7.2. Binarization

Table 4 presents the Peak Signal-to-Noise Ratio (PSNR) values, in decibels (dB), for multiple image binarization and enhancement algorithms tested under three lighting conditions: midtone-grainy, low-light-rough, and high-light-soft. The PSNR value tells you how well an image is binarized compared to a reference image. Higher numbers mean better preservation. The Proposed approach has the best PSNR in all categories, with scores of 37.34 dB, 36.48 dB, and 32.89 dB. The average for all of them was 35.57 dB. The PSNR values for standard techniques such as Otsu, Niblack, Gatos, Li, Kapur, and others typically range from 27.90 dB to 27.97 dB.

Table 4. Average PSNR (dB) per method and category of the real-time stone inscription image

Methods	midtone-grainy	low-light-rough	high-light-soft	overall average
Proposed	37.34	36.48	32.89	35.57
Otsu	27.92	27.91	28.04	27.96
RC	27.92	27.91	28.04	27.96
Niblack	27.90	27.92	27.99	27.94
Gatos	27.92	27.91	28.04	27.96
Li	27.90	27.92	28.02	27.95
Mean	27.91	27.91	28.04	27.95
Percentile	27.92	27.93	28.05	27.97
Intermodes	27.93	27.93	28.04	27.97
Kapur	27.92	27.92	28.01	27.95
Moments	27.90	27.92	27.99	27.94
Huang	27.91	27.92	28.02	27.95
Bersen	27.91	27.92	28.01	27.95
Contrast	27.91	27.92	28.02	27.95
Median	27.90	27.91	28.04	27.95
Midgray	27.85	27.86	27.98	27.90
Triangle	27.91	27.93	28.03	27.96
Isodata	27.92	27.91	28.04	27.96

8. Discussion and Future Work

8.1. Limitation

Despite the encouraging results, there are significant caveats that must be highlighted. The dataset is relatively small and collected through fieldwork, limiting the feasibility of formal statistical significance testing and extensive cross-validation. As a result, performance evaluation relies on a held-out test set using standard detection metrics. Next, the experimental validation is limited to Kannada stone inscriptions, and the approach's generalizability to other scripts or multilingual inscriptions has not yet been empirically verified due to the study's limited scope. In addition, the current framework relies exclusively on RGB images, which are insufficient for handling inscriptions with extreme erosion or minimal visual contrast. Finally, the proposed pipeline focuses on the binarization and text-region detection.

8.2. Future Work

Future research will focus on extending the proposed framework to larger, multilingual inscription datasets to evaluate its generalizability and support statistically robust performance analysis. Incorporating multimodal information, such as 3D surface geometry and multispectral or hyperspectral imaging, is expected to improve detection in severely degraded scenarios.

In addition, efforts will be made to curate and release a publicly accessible subset of the dataset, subject to cultural heritage and site-access constraints, to support reproducibility and facilitate comparative research. The final focus is on integrating Optical Character Recognition (OCR) and semantic analysis modules with the proposed detection framework to enable complete transcription, linguistic interpretation, and content-level analysis of stone inscriptions.

9. Conclusion

This work summarizes the modified binarization and YOLOv families (YOLOV8–YOLOV11) for detecting TextRegions and Non-textRegions in Kannada stone inscribed images. The performance of the Inscription images is evaluated using Precision, Recall, F1-score, and Accuracy. The proposed approach, YOLOv11, outperformed the other models, namely, YOLOv8, YOLOv9, and YOLOv10, in the experiments. The YOLOv11s achieved the highest accuracy of 85.20%. The proposed experimental results show that modern YOLO architectures are suitable for analyzing ancient Kannada inscription images.

The proposed work in this paper will be helpful for further studies, such as segmentation and recognition of isolated characters in inscription images. In this direction, the work is in progress. This will have an enormous impact in the field of digital epigraphy and the broader effort to apply modern AI techniques to protect and understand ancient scripts by addressing the problems that arise with complex, deteriorated historical inscription image data.

Acknowledgements

The authors are grateful to the University Grants Commission (UGC), New Delhi, India, for research support, as well as the Archaeological Survey of India (ASI), India, and the Department of Archaeology, Museum, and Heritage, Government of Karnataka, India, for supporting data collection.

Data Availability

The author collected the dataset generated and analyzed during the current study through fieldwork, and it is not publicly available at present. The authors plan to release the dataset in a public repository after completion of data curation and quality verification.

References

[1] Shashaank M. Aswatha et al., "A Method for Extracting Text from Stone Inscriptions using Character Spotting," *Computer Vision - ACCV 2014 Workshops: Singapore*, Singapore, vol. 9009, pp. 598-611, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[2] Sachin Bhat, and G Seshikala, "Preprocessing and Binarisation of Inscription Images using Phase-based Features," *2018 Second International Conference on Advances in Electronics, Computers and Communications (ICAECC)*, Bangalore, India, pp. 1-6, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[3] S. Bhuvaneswari, and K. Kathiravan, "Enhancing Epigraphy: A Deep Learning Approach to Recognize and Analyze Tamil Ancient Inscriptions," *Neural Computing and Applications*, vol. 36, no. 31, pp. 19839-19861, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[4] N. Shobha Rani, and Arun Gopi, "A Quad Tree based Binarization Approach to Improve Quality of Degraded Document Images," *International Journal of Computer Science Engineering (IJCSE)*, vol. 3, no. 1, pp. 1-8, 2024. [Google Scholar] [Publisher Link]

[5] Miguel Carrero-Pazos, and David Espinosa-Espinosa, "Tailoring 3D Modelling Techniques for Epigraphic Texts Restitution: Case Studies in Deteriorated Roman Inscriptions," *Digital Applications in Archaeology and Cultural Heritage*, vol. 10, pp. 1-28, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[6] H.T. Chandrakala, G. Thippeswamy, and Roshan Joy Martis, "Impact of Total Variation Regularization on Character Segmentation from Historical Stone Inscriptions," *Pattern Recognition and Image Analysis*, vol. 31, no. 1, pp. 35-48, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[7] Ushasi Chaudhuri, Partha Bhowmick, and Jayanta Mukherjee, "A Novel Rough Set based Technique for Character Spotting on Inscription Images," *2017 Ninth International Conference on Advances in Pattern Recognition (ICAPR)*, Bangalore, India, pp. 1-6, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[8] Sugata Das, Sekhar Mandal, and Amit Kumar Das, "Binarization of Stone Inscripted Documents," *2015 IEEE International Conference on Computer Graphics, Vision and Information Security (CGVIS)*, Bhubaneswar, India, pp. 11-16, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[9] K. Durga Devi et al., "Pattern Matching Model for Recognition of Stone Inscription Characters," *The Computer Journal*, vol. 66, no. 3, pp. 554-564, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] K. Durga Devi, and P. Uma Maheswari, "RETRACTED ARTICLE: Digital Acquisition and Character Extraction from Stone Inscription Images using Modified Fuzzy Entropy-based Adaptive Thresholding," *Soft Computing*, vol. 23, no. 8, pp. 2611-2626, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[11] Rafael C. Gonzalez, and Richared E. Woods, *Digital Image Processing*, 3rd ed., Pearson, 2009. [Google Scholar]

[12] J. Jayanthi, and P. Uma Maheswari, "Comparative Study: Enhancing Legibility of Ancient Indian Script Images from Diverse Stone Background Structures using 34 Different Pre-Processing Methods," *Heritage Science*, vol. 12, no. 1, pp. 1-17, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[13] Ke Liu, and Jun Ma, "3D-Scanning and Computer Reverse Engineering Technology to Preserve Inscriptions at Beihai Park," *International Journal of Simulation: Systems, Science and Technology*, vol. 17, no. 26, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[14] H.S. Mohana et al., "Interactive Segmentation for Character Extraction in Stone Inscriptions," *Second International Conference on Current Trends in Engineering and Technology - ICCTET 2014*, Coimbatore, pp. 321-327, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[15] Monisha Munivel, and V.S. Felix Enigo, "MLIBT: A Multi-Level Improvised Binarization Technique for Tamizhi Inscriptions," *Expert Systems with Applications*, vol. 236, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[16] Pranav Rajnish et al., "Improving the Quality and Readability of Ancient Brahmi Stone Inscriptions," *2023 2nd International Conference for Innovation in Technology (INOCON)*, Bangalore, India, pp. 1-8, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[17] N. Sasipriyaa et al., "An Approach for Tamil Handwritten Recognition using ABC-Enabled GAN," *2024 2nd International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), Paralakhemundi Campus, Centurion University of Technology and Management*, Odisha, India, pp. 1-5, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[18] Indu Sreedevi et al., "NGFICA based Digitization of Historic Inscription Images," *International Scholarly Research Notices*, vol. 2013, no. 1, pp. 1-7, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[19] S. Sukanthi, S. Sakthivel Murugan, and S. Hanis, "Binarization of Stone Inscription Images by Modified Bi-Level Entropy Thresholding," *Fluctuation and Noise Letters*, vol. 20, no. 6, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[20] Ultralytics, Explore Ultralytics YOLOv8, 2023. [Online]. Available: https://docs.ultralytics.com/models/yolov8/

[21] Ultralytics, YOLOv9 vs. YOLO11: A Technical Deep Dive into Modern Object Detection, 2024. [Online]. Available: https://docs.ultralytics.com/compare/yolov9-vs-yolo11/

[22] Karel Zuiderveld, "Contrast Limited Adaptive Histogram Equalization," *Graphics Gems IV*, pp. 474-485, 1994. [Google Scholar] [Publisher Link]

[23] Bapu D. Chendage, and Rajivkumar S. Mente, "Enhancement of Ancient Marathi Script using Improved Binarization Method," *Sādhanā*, vol. 48, no. 4, pp. 1-5, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[24] Bapu D. Chendagel, Rajivkumar S. Mente, and Bapu D. Chendage, "A Comparative Study of Contrast Enhancement and Brightness Preservation of Ancient Inscription Image using Different Histogram Equalization Algorithms," *International Journal of Natural Sciences*, vol. 14, no. 80, pp. 61336-61343, 2023. [Google Scholar]

[25] A.I. Papadaki et al., "Accurate 3D Scanning of Damaged Ancient Greek Inscriptions for Revealing Weathered Letters," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 40, pp. 237-244, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[26] Zhongming Pei, Yong Mao Huang, and Ting Zhou, "Review on Analysis Methods Enabled by Hyperspectral Imaging for Cultural Relic Conservation," *Photonics*, vol. 10, no. 10, pp. 1-19, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[27] Haiqing Yang et al., "Hyperspectral Data Set of Stone Cultural Relics in High-Precision Machine Vision Scene," *Scientific Data*, vol. 12, no. 1, pp. 1-10, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[28] Balasubramanian Murugan, and P. Visalakshi, "Ancient Tamil Inscription Recognition using Detect, Recognize and Labelling, Interpreter Framework of Text Method," *Heritage Science*, vol. 12, no. 1, pp. 1-21, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[29] Boris Sekachev, Andrey Zhavoronkov, and Nikita Manovich, "Computer Vision Annotation Tool: A Universal Approach to Data Annotation," *Intel [Internet]*, 2019. [Google Scholar]