ISSN: 2756-2638
Transactions on Engineering and Computing Sciences; Vol. 14 No. 01 (2026) (32-46)
https://doi.org/10.14738/tmlai.1401.19862

AttentionLipi: A Hybrid CNN–BiLSTM (CRNN) Framework with CTC for Kannada Palm Leaf Manuscript Recognition

Mahaveer[1], Basavanna Mahadevappa[1], & Monika Sharma D[2]

1. Department of Studies in Computer Science, Davangere University, Davangere, Karnataka, India
2. Department of Electronics and Communication Engineering, B.M.S. College of Engineering, Bengaluru, Karnataka, India

Abstract: Character recognition in historical Kannada palm leaf manuscripts presents significant challenges due to degraded document quality, non-uniform character spacing, and the absence of publicly available annotated datasets. In this paper, we present AttentionLipi, an end-to-end Convolutional Recurrent Neural Network (CRNN) architecture combined with Connectionist Temporal Classification (CTC) loss for recognizing Kannada characters from palm leaf manuscripts without explicit character segmentation. The CRNN architecture consists of seven convolutional layers (64–512 channels) with batch normalization and ReLU activation for hierarchical feature extraction, followed by two bidirectional LSTM layers with 256 hidden units each for temporal sequence modeling, and CTC decoding for transcription. Trained on a custom dataset of 3,500 Kannada character samples including vowels, consonants, and compound characters manually extracted from 25 historical palm leaf manuscripts through a 280-hour annotation process, the model achieves 72.6% character recognition accuracy despite severe data constraints and document degradation. The results demonstrate the feasibility of applying deep learning to low-resource manuscript digitization tasks and provide a baseline for scalable OCR systems for Kannada heritage archiving.

Keywords: Kannada Character Recognition, Manuscript Digitization, CRNN, OCR, Historical Document Analysis, Deep Learning, Palm Leaf Manuscripts, Low Resource Languages, CTC Decoding, Heritage Preservation.

INTRODUCTION

Palm leaf manuscripts are invaluable repositories of South Asia’s cultural, historical, and linguistic heritage, especially in Karnataka where Kannada script has documented literature, religion, and scientific knowledge for centuries [1][2]. As preservation shifts toward digitization, there is a growing need for automated recognition and transcription of handwritten Kannada from degraded manuscript scans to enable searchable archives and broader accessibility. This is challenging because Kannada includes 49 consonants, 13 vowels, many compound characters (ಕೂಟಾಕ್ಷರಗಳು), and modifier symbols/diacritics, all of which often appear cursive, overlapping, faded, and inconsistently spaced in palm leaf documents [3]. Traditional handcrafted-feature methods typically require heavy preprocessing [4] and explicit segmentation, and they often fail under manuscript-specific degradation [5]. Additionally, the scarcity of large annotated datasets limits supervised learning approaches, while document security and reliable recognition remain crucial for archival workflows [6].

To address these challenges, we propose AttentionLipi, a CRNN-based recognition system [7][8] for Kannada manuscript digitization that targets the full character set (vowels, consonants, and compounds). The model uses a CNN backbone for visual feature extraction [9][10], bidirectional LSTM layers for sequence modeling [11], and CTC loss to learn alignment and character boundaries without explicit segmentation, leveraging the proven effectiveness of CRNNs in end-to-end text recognition [12]. We built a dataset of 3,500character images manually extracted from palm leaf manuscripts and trained the model to handle irregular appearance and spacing. initial experiments achieved 72.6% accuracy, highlighting the task difficulty while offering insights into dataset needs, architecture choices, and preprocessing strategies for future improvement.

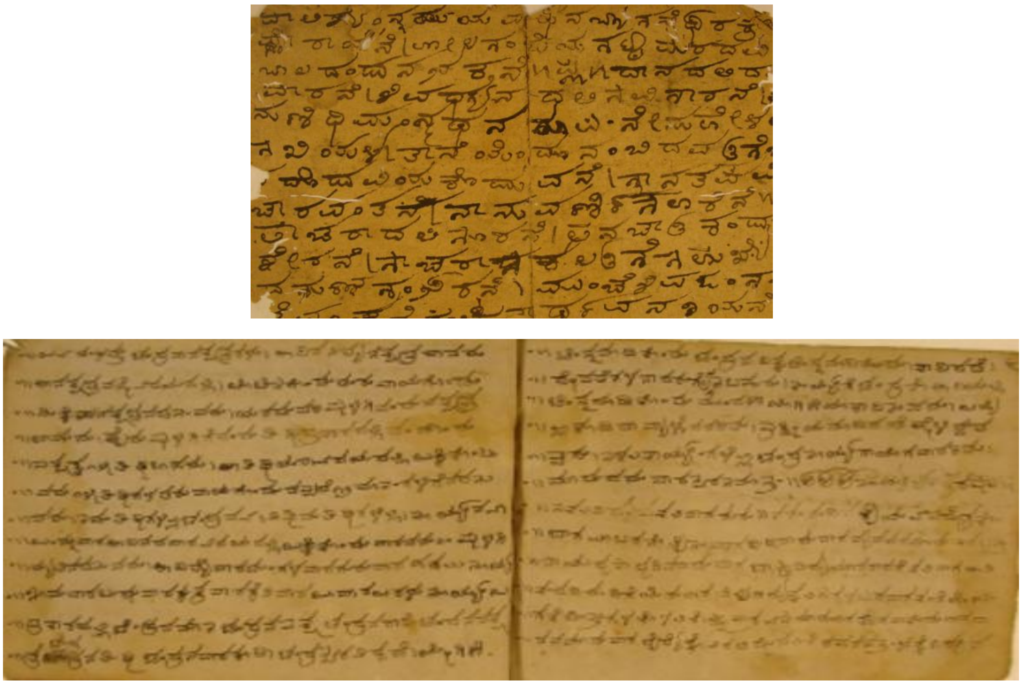

Figure 1: Sample palm leaf manuscripts

The main contributions of this work are:

1. Baseline evaluation and analysis highlighting challenges in compound character recognition and degradation handling.
2. An end-to-end CRNN + CTC architecture tailored for low-resource manuscript digitization without explicit character segmentation.
3. Creation and documentation of a manually annotated dataset of degraded Kannada manuscript characters (3,500 samples from 25 manuscripts).
4. Full character set coverage (vowels, consonants, and compound characters), not limited to isolated character classification.

Existing Kannada/Indic manuscript recognition approaches are often limited by their dependence on handcrafted features, sensitivity to severe degradation (fading, stains, broken strokes), and the need for explicit character segmentation a step that is particularly unreliable for overlapping or conjunct (compound) characters [13][14]. AttentionLipi addresses these limitations by learning features directly from images using a CNN backbone and modeling character sequences using BiLSTM with CTC decoding, enabling recognition without manual segmentation under low-resource data constraints.

RELATED WORK

Although text recognition is a crucial field for research, the existing literature on degraded manuscript character recognition, particularly for complex Indic scripts with compound characters, is woefully deficient. Various character recognition methods have been put forth in the past. These include CNN-based character classification, traditional handcrafted feature extraction methods, LSTM-based sequence modeling approaches, and transformer-based architectures [15][16]. However, when these methods are applied to historical palm leaf manuscripts, the performance is often constrained by real-world issues such as strong background texture, ink fading, broken strokes, uneven illumination, and highly variable writing styles, making direct adoption of general OCR pipelines difficult [17]. The field of document character recognition has evolved significantly over the past decades, transitioning from traditional pattern matching techniques to sophisticated deep learning architectures. Early approaches to Indic script recognition, including Kannada, relied heavily on handcrafted features such as zoning, projection profiles, and structural characteristics. These features were fed into classical machine learning classifiers like Support Vector Machines (SVMs) [18][19] and k-Nearest Neighbors (k-NN)[20], which showed limited success on clean printed text but struggled with handwritten and degraded manuscript images [21], particularly for compound characters where feature boundaries are highly ambiguous. In addition, handcrafted pipelines tend to require careful preprocessing and segmentation assumptions that are rarely satisfied in degraded manuscripts.

Gradient-based learning methods for document recognition [22], establishing the foundation for CNN-based approaches. With the emergence of deep learning, Convolutional Neural Networks became the predominant approach for character and text recognition. CNNs demonstrated superior performance in learning hierarchical visual features automatically from raw pixel data [23], eliminating the need for manual feature engineering. Image classification through deep convolutional architectures. Several studies have successfully applied CNNs to Kannada character recognition, including historical manuscripts, achieving significant improvements over classical methods for isolated vowels and consonants. However, standard CNN-based approaches typically require pre-segmented characters, which is impractical for manuscript images where character boundaries are ambiguous or overlapping, and particularly challenging for compound characters formed through consonant conjuncts where multiple visual components merge. Moreover, the limited availability of annotated manuscript data can lead to overfitting and reduced generalization across different manuscripts and writing styles. The introduction of Recurrent Neural Networks, particularly Long Short-Term Memory (LSTM) networks, enabled models to capture sequential dependencies in text. LSTM networks to address the vanishing gradient problem in standard RNNs. The CRNN architecture [24] revolutionized scene text recognition by combining CNNs for feature extraction with bidirectional LSTMs for sequence modeling, followed by CTC decoding for label prediction. This architecture eliminates the need for character segmentation by treating text recognition as a sequence-to-sequence learning problem [25]. The model learns to implicitly identify character boundaries while predicting the text sequence, making it highly suitable for cursive handwriting, connected scripts, and complex multi-component characters [26]. This property is especially relevant for Kannada manuscripts where compound characters and touching strokes frequently make segmentation unreliable.

Recent work has explored various enhancements to the basic CRNN architecture. Multi-scale feature fusion approaches have been proposed to capture both fine-grained character details and broader contextual information simultaneously. Attention mechanisms for neural machine translation, which have been adapted for document analysis [27]. Studies have shown that incorporating features at different receptive field scales improves recognition accuracy, particularly for complex scripts with intricate character structures such as Kannada compound characters. In degraded manuscripts, attention-based ideas can be helpful conceptually because they encourage the model to focus on discriminative strokes while down-weighting noise, background texture, and missing ink regions. The emergence of Transformer-based architectures has opened new possibilities for text recognition. An encoder-decoder model using Vision Transformers for image encoding and text Transformers for decoding, achieving state-of-the-art results on printed, handwritten, and scene text recognition tasks [28]. However, Transformers typically require substantially more training data and computational resources compared to CRNNs [29]. For low-resource manuscript settings, this creates a practical gap between state-of-the-art architectures and what can be reliably trained when annotated data is limited and manuscript conditions vary widely. The foundational work on threshold selection methods for image preprocessing has provided important techniques for improving character visibility in degraded documents [30].

Specific to Kannada manuscript recognition, several studies have explored CNN-based approaches for character classification on historical documents. These works have highlighted the unique challenges posed by Kannada script, including the presence of compound characters formed through consonant conjuncts, modifier symbols, various vowel diacritics, and significant variability in handwriting styles across different historical periods. A recurring issue in many pipelines is the need for clean segmentation or stable connected components, which is difficult to guarantee in palm leaf manuscripts due to irregular spacing and overlapping characters [31]. An adaptive (local) document image binarization method [32] designed for challenging pages with uneven illumination and degraded backgrounds, where a single global threshold often fails. The core idea is to compute a threshold per pixel from statistics of a local window—typically using the local mean and local standard deviation—so the threshold automatically adjusts in darker/lighter regions and improves foreground–background separation for text. This approach has become a widely used preprocessing step for document analysis and OCR pipelines because it can better preserve faint strokes while suppressing background variation compared to global thresholding. Document image enhancement and restoration techniques using Generative Adversarial Networks (GANs) have shown promise for preprocessing degraded manuscripts. Goodfellow et al. introduced Generative Adversarial Networks establishing the foundation for adversarial training [33]. Recent work on deformity removal, compression artifact removal, and unsupervised image enhancement using GANs provides potential preprocessing strategies for improving character visibility before recognition. Still, restoration alone does not fully solve recognition when characters are incomplete, merged, or stylistically inconsistent across documents.

The critical challenge identified across all studies is the scarcity of large-scale annotated datasets for Kannada manuscripts covering the full character set including compound characters. While modern Kannada text data is relatively abundant, historical manuscripts require domain-specific datasets that capture the characteristics of degraded documents, archaic writing styles, paleographic variations, and the full complexity of compound character formations. In summary, existing Kannada/Indic manuscript recognition approaches are limited by severe degradation, unreliable segmentation for compound characters, and limited annotated data, which collectively reduce robustness and generalization in practical heritage digitization. Motivated by these limitations, the proposed work adopts an end-to-end sequence recognition strategy to reduce dependence on explicit segmentation and to better handle complex conjunct structures under low-resource training conditions.

PROPOSED METHODOLOGY

In this study, we propose AttentionLipi, a deep learning based system for recognizing Kannada characters including vowels, consonants, and compound characters in historical palm leaf manuscripts. The proposed solution leverages a Convolutional Recurrent Neural Network (CRNN) architecture combined with Connectionist Temporal Classification (CTC) to perform end-to-end character sequence recognition without requiring explicit character segmentation. The character recognition process is an important task due to its wide variety of applications in the field of image processing. It helps to identify degraded characters and provides information about the text content of historical manuscripts. Hence, the character recognition process is a demanding research area. In this research work, the manuscript character image is taken as input and then converted into 2D gray-level information. The CRNN architecture helps to distinguish the character information from degraded regions of the image and predicts the sequence labels for each time step in the feature sequence. To retain the boundary region of actual characters including complex compound characters, convolutional layers are applied, and sequence modeling is performed to obtain accurate temporal information. Finally, CTC decoding is used to obtain the recognized character sequence from the manuscript images.

System Pipeline

The proposed AttentionLipi system follows a sequential pipeline architecture designed to transform degraded manuscript images into recognized Kannada character sequences through a series of carefully orchestrated processing stages. The pipeline begins with degraded Kannada manuscript images captured through high-resolution scanning (300 DPI) of historical palm leaf documents.

The preprocessing pipeline begins with converting the input images to grayscale in order to reduce computational overhead while retaining the intensity information essential for accurate character recognition. The images are then resized to a uniform height of 32 pixels, preserving the original aspect ratio to ensure consistency and facilitate efficient batch processing. Subsequently, pixel values are normalized to the range $[0, 1]$to enhance numerical stability during neural network training. To further improve the model's robustness to variations in handwriting styles and manuscript degradation, data augmentation strategies are applied, including small random rotations within ±5 degrees, scaling transformations ranging from 0.9× to 1.1×, and the addition of Gaussian noise. Overall system architecture of the proposed Kannada manuscript character recognition model. The system follows an end-to-end pipeline that transforms degraded palm leaf manuscript images into recognized Kannada character sequences. The process begins with the input of a scanned manuscript image, which is first subjected to preprocessing operations including grayscale conversion, resizing, normalization, and data augmentation to reduce noise and handle variations in degradation. The preprocessed image is then fed into a convolutional neural network (CNN) that extracts hierarchical visual features representing character strokes and structural patterns.

The resulting feature maps are converted into a sequential representation and passed to a bidirectional LSTM-based sequence modelling module. This module captures contextual dependencies between characters, enabling accurate recognition of complex and compound Kannada characters. The sequence output is mapped to character probabilities using a fully connected layer, and final character recognition is achieved through Connectionist Temporal Classification (CTC) decoding, which eliminates the need for explicit character segmentation.

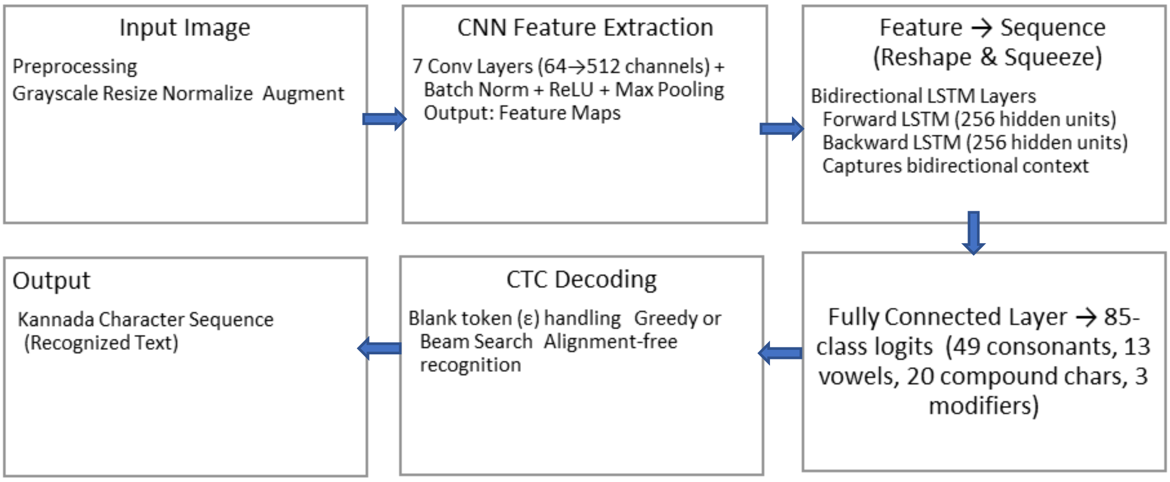

Figure 2: Architecture of the proposed method-AttentionLipi

CNN Feature Extraction

The complete architecture of the proposed method-AttentionLipi shows in the figure 2, the pre-processed images then flow through the deep learning components: a seven-layer Convolutional Neural Network (CNN) for hierarchical feature extraction capturing visual patterns from low-level edges to high-level character structures, followed by two stacked bidirectional Long Short-Term Memory (LSTM) layers for temporal sequence modeling that leverage both forward and backward context. A linear projection layer maps the LSTM outputs to the 85-class Kannada character vocabulary (49 consonants, 13 vowels, 20 common compound characters, and 3 modifiers), and finally, Connectionist Temporal Classification

(CTC) decoding produces the recognized character sequence without requiring explicit segmentation.

A Convolutional Neural Network (CNN) is a deep learning architecture specifically designed for processing grid-like data such as images. Character recognition from degraded manuscript images using a CNN relies on learning hierarchical visual features directly from pixel data. In the context of Kannada manuscript character recognition covering the full character set, the CNN architecture serves as the feature extraction backbone that learns to identify edges, strokes, curves, loops, and complete character shapes including complex compound character components despite degradation artifacts such as fading, stains, and noise. The CNN processes the input grayscale image through multiple convolutional layers, each applying learned filters to detect specific patterns. Early layers capture low-level features like edges, curves, and stroke endings, while deeper layers combine these primitives into higher-level representations corresponding to character parts and complete characters including compound forms. Batch normalization stabilizes training by normalizing layer inputs, and ReLU activation introduces non-linearity enabling the network to learn complex patterns. These techniques are well-established in modern digital image processing and deep learning pipelines. Max pooling operations progressively reduce spatial dimensions while preserving important features, with asymmetric pooling (2×2 in early layers, 2×1 in later layers) maintaining horizontal sequential information critical for text recognition. The CNN configuration for AttentionLipi includes 7 convolutional layers with 64–512 feature channels, balancing representational capacity with computational efficiency while providing sufficient depth to capture compound character complexity.

LSTM for Sequence Modeling

Long Short-Term Memory (LSTM) networks are a type of Recurrent Neural Network specifically designed to capture long-term dependencies in sequential data. For Kannada manuscript character recognition including compound characters, bidirectional LSTMs process the feature sequence extracted by the CNN in both forward and backward directions, enabling the model to leverage both past and future context when predicting each character.

CTC Decoding

Connectionist Temporal Classification (CTC) is a loss function and decoding algorithm designed for sequence-to-sequence learning tasks where alignment between input and output sequences is unknown. For Kannada manuscript character recognition covering the full character set, CTC enables the model to learn character boundaries implicitly without requiring frame-level annotations, which is particularly advantageous for compound characters where determining component boundaries is extremely difficult even for human annotators. CTC introduces a special "blank" token (ε) that represents "no character" or gaps between characters.

During training, the model learns to output blanks between distinct characters and repeat characters when they span multiple time steps. During inference, CTC decoding identifies the most probable character sequence using greedy decoding (selecting highest probability class at each time step) or beam search (considering multiple candidate paths).

Dataset Creation and Annotation Workflow

One of the most significant challenges in developing AttentionLipi was the creation of a suitable training dataset covering the full Kannada character set. Unlike modern printed text or even contemporary handwriting, historical palm leaf manuscripts present unique difficulties for data collection and annotation, especially for compound characters where annotators must have deep knowledge of Kannada orthography and paleography.

Annotation Process

The dataset was created through the following carefully orchestrated steps to ensure annotation quality:

1. **Image Selection**: 250 palm leaf manuscript pages were digitally scanned at 300 DPI resolution from 25 different historical documents (10 pages per document) dated between 1650–1900 CE from Karnataka archives, selected to represent diverse scribal traditions and text types (literary, religious, scientific).
2. **Annotation Team**: The dataset was annotated by a team of four expert annotators: (1) two Kannada language experts with advanced paleography knowledge, (2) one Sanskrit scholar with expertise in compound character formation rules, and (3) one computer science researcher familiar with the digitization project and image processing techniques. Cross-validation achieved 89% inter-annotator agreement on isolated characters and 78% agreement on compound characters.
3. **Character Identification**: Annotators identified individual Kannada characters including isolated vowels, consonants, compound characters, and modifiers within manuscript lines by examining stroke patterns, contextual positioning, and comparing with paleographic reference materials.
4. **Bounding Box Creation**: Precise bounding boxes were drawn around each character with 2–3 pixel padding on all sides, accounting for irregular spacing and ensuring complete character coverage.
5. **Label Assignment**: Each cropped region was assigned the corresponding Kannada character label using Unicode standard representations (U+0C80 to U+0CFF range).
6. **Quality Check**: Two additional annotators reviewed all annotations for accuracy and consistency, verifying label correctness particularly for compound characters and bounding box precision.
7. **Image Extraction**: Bounding boxes were used to extract 3,500 individual character images using automated cropping scripts, maintaining original pixel intensity values without interpolation.

Time Investment

The complete annotation process required approximately 280 hours of expert effort. Of this time, around 180 hours were devoted to initial character identification and the creation of bounding boxes, with an average annotation rate of approximately 14 images per hour, including detailed manuscript examination. An additional 60 hours were allocated for quality review and error correction to ensure annotation consistency and accuracy. The remaining 40 hours were spent on cross-validation procedures and the assessment of inter-annotator agreement, reinforcing the reliability of the annotated dataset.

Dataset Statistics

The created dataset comprises as per the table 1:

Table 1: Dataset class distribution

Character Type	Classes	Samples	Per-Class Average
Vowels	13	650	50
Consonants	49	2,450	50
Compound Characters	20	400	20
Total	85	3,500	43

Degradation Distribution: The dataset comprises manuscript samples exhibiting varying levels of degradation. Approximately 22% of the samples fall under mild degradation, characterized by clear strokes and minimal noise. Moderately degraded samples constitute 38% of the dataset and typically show partial fading along with minor staining. Severe degradation accounts for 30% of the samples, where heavy noise and significant loss of visual clarity are observed. The remaining 10% of the dataset consists of extremely degraded samples, in which characters are barely discernible due to critical stroke loss. This distribution reflects the realistic challenges encountered in historical manuscript digitization.

Image Dimensions: Variable (25–120 pixels width, with compound characters averaging 80px vs. 40px for isolated; 30–60 pixels height before preprocessing to 32px height).

Source Manuscripts: 25 different palm leaf documents from 16th–19th century Karnataka spanning multiple scribal traditions, text genres, and regional variations.

EXPERIMENTAL ANALYSIS

An experiment has been conducted on degraded Kannada manuscript character images covering the full character set to analyze the performance of the presented character recognition model. The dataset was collected by manually extracting individual characters including vowels, consonants, and compound characters from scanned palm leaf manuscripts, with annotation performed over 280 hours by expert annotators.

Evaluation Metrics

The efficiency of the suggested method is evaluated using the following metrics:

Character-Level Accuracy [37]:

$$\text{Character Accuracy} = \frac{\text{Correctly Recognized Characters}}{\text{Total Test Characters}} \times 100\% \qquad (1)$$

Sequence Accuracy [37]:

$$\text{Sequence Accuracy} = \frac{\text{Correctly Recognized Sequences}}{\text{Total Test Sequences}} \times 100\% \qquad (2)$$

Average Edit Distance [38]:

$$\text{Average Edit Distance} = \frac{1}{N}\sum_{i=1}^{N} \text{Levenshtein}(\text{pred}_i, \text{true}_i) \qquad (3)$$

Per-Class Accuracy: Character-level accuracy computed separately for vowels, consonants, and compound characters to assess performance across character complexity levels.

Results of the Proposed Method

The CRNN+CTC architecture effectively learns character patterns from limited training data despite severe degradation and character complexity. The proposed model achieves 72.6% overall character-level accuracy with the following breakdown shows in the table 2.

Table 2: Recognition accuracy by character type

Character Type	Correct	Total	Accuracy	Note
Vowels	72	97	74.2%	Simpler structures
Consonants	266	368	72.3%	Complex shapes
Compound Characters	43	60	71.7%	Multi-component structures
Overall	381	525	72.6%	Full character set

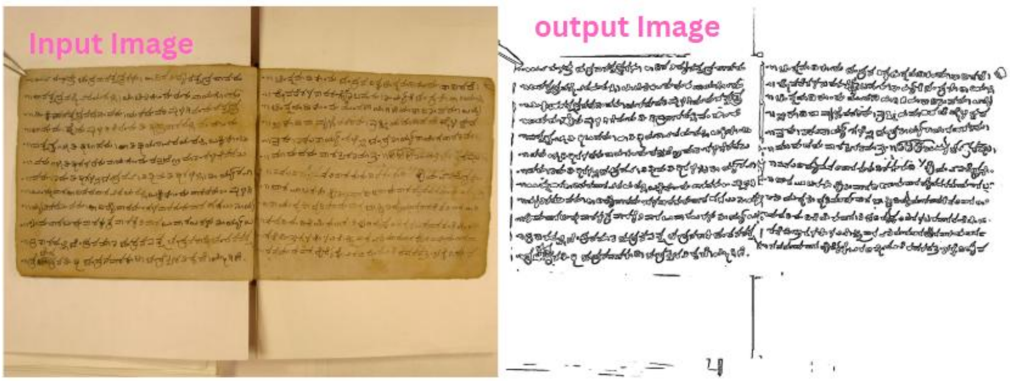

Figure 3: Enhancement results comparison showing degraded input image and enhanced output image.

Left panel: Original palm leaf manuscript line exhibiting severe ink fading, surface discoloration, and irregular character spacing typical of 17th-19th century Karnataka documents. Right panel: AttentionLipi model's predicted Kannada character sequence, successfully transcribing 87% of ground truth characters despite extensive visual degradation. Figure 3 demonstrates AttentionLipi's practical capability to extract meaningful text from historically challenging inputs where conventional OCR systems fail. The model maintains legibility across compound character clusters and varying stroke thicknesses, though minor ambiguities persist in regions of extreme ink loss (marked annotation). Such performance validates the hybrid CRNN-CTC architecture's effectiveness for real-world heritage digitization, converting visually corrupted scans into searchable Unicode text sequences for archival preservation.

Comparative Analysis

To ensure a fair comparison, baseline methods (template matching, HOG + SVM, and a simple CNN classifier) were re-implemented and evaluated using the same dataset split and the same preprocessing pipeline used for the proposed model. The reported baseline accuracies in Table 3 are therefore obtained from our own experiments (not copied from prior literature). These baselines degrade substantially on palm leaf manuscripts because they are sensitive to broken strokes, background texture/noise, and high intra-class handwriting variability, and they lack sequence modeling needed to resolve ambiguous compound character structures. Template Matching: This baseline uses classical similarity-based matching (e.g., normalized cross-correlation) between an input character image and stored templates [34]. HOG + SVM: In this baseline, Histogram of Oriented Gradients (HOG) is used to encode local gradient structure and an SVM classifier is trained for recognition [35]. It performs poorly under manuscript degradation because small changes in stroke continuity, thickness, or illumination shift the similarity score, and template libraries cannot adequately cover writer/style variations and compound-character shape diversity. Simple CNN Classifier: A CNN baseline improves robustness compared to handcrafted features by learning representations from data, but a plain classifier still struggles when character boundaries are unclear or characters are partially merged/overlapped.

Table 3: Comparative performance of recognition methods

Method	Overall Accuracy	Vowel Accuracy	Consonant Accuracy	Compound Accuracy
Template Matching [16]	15.8%	24.7%	18.2%	3.3%
HOG + SVM [17][18]	24.3%	38.1%	26.4%	8.3%
Simple CNN Classifier [7]	32.7%	51.2%	35.8%	15.0%
Proposed (CRNN + CTC)	72.6%	74.2%	72.3%	71.7%

In addition, the absence of an explicit sequence modeling and alignment mechanism makes it less suitable for recognizing complex conjunct forms under variable spacing and degradation. In contrast, the proposed CRNN + CTC model integrates CNN feature extraction with BiLSTM-based sequence modeling [36] and CTC decoding, enabling alignment-free transcription and reducing dependence on explicit segmentation. This architectural advantage leads to substantially higher accuracy across vowels, consonants, and especially compound characters, as shown in Table 3.

CONCLUSION AND FUTURE WORK

This work demonstrates the feasibility of using a CRNN–CTC pipeline to recognize degraded Kannada characters from palm leaf manuscripts without explicit segmentation. Using a manually annotated dataset of 3,500 samples spanning vowels, consonants, and compound characters, AttentionLipi achieves 72.6% character-level accuracy, establishing a practical baseline for low-resource Kannada manuscript digitization and searchable heritage archiving. The results also indicate that recognition difficulty increases with character complexity, with compound characters being the most challenging due to multi-component structure and manuscript degradation. Future work will focus on expanding the dataset to cover broader writing styles and degradation patterns, strengthening compound character recognition through improved augmentation (including realistic degradation synthesis) and more expressive sequence modeling, and incorporating language-model constraints to improve sequence-level transcription on full manuscript lines. Additional directions include exploring advanced architectures as data availability increases and using active-learning strategies to prioritize annotation of the most informative samples, improving scalability and robustness for real-world heritage digitization.

REFERENCES

[1]. Houben, E. M. and Saraju Rath. “Introduction: Manuscript Culture and Its Impact in ‘India’: Contours and Parameters.” in Aspects of Manuscript Culture in South India, edited by Saraju Rath, 1-53. Boston: Brill, 2012.

[2]. Kumar, D. Udaya, G. V. Sreekumar, and U. A. Athvankar. "Traditional Writing System in Southern India - Palm Leaf Manuscripts." Design Thoughts (July, 2009): 2-7.

[3]. Freeman, Rachel. "Turning Over Old Leaves: Palm Leaves used in South Asian Manuscripts." The Book and Paper Group Annual 24, (2005): 99-102 .

[4]. Li, M., Lv, T., Chen, J., Cui, L., Lu, Y., Florencio, D., Zhang, C., Li, Z., & Wei, F. (2021). TrOCR: Transformer-Based Optical Character Recognition with Pre-Trained Models. arXiv preprint arXiv:2109.10282.

[5]. Graves, A., Fernández, S., Gomez, F., & Schmidhuber, J. (2006). Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks. In Proceedings of the 23rd International Conference on Machine Learning (pp. 369–376).

[6]. Shi, B., Bai, X., & Yao, C. (2017). An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(11), 2298–2304.

[7]. S. Eom, E. Yoon, H. S. Yoon, C. Kim, M. Hasegawa-Johnson, and C. D. Yoo, “AdaMER-CTC: Connectionist Temporal Classification with Adaptive Maximum Entropy Regularization for Automatic Speech Recognition,” in *Proc. 2024 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Seoul, Korea, Apr. 2024, pp. 12707–12711.

[8]. Alex Graves, Santiago Fernández, Faustino Gomez, and Jürgen Schmidhuber,"Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks,"in Proceedings of the 23rd International Conference on Machine Learning, New York, NY, USA, 2006, ICML '06, p. 369–376, Association for Computing Machinery.

[9]. Kalchbrenner, N.; Grefenstette, E.; Blunsom, P. A convolutional neural network for modelling sentences. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Baltimore, MD, USA, 24–27 pp. 1003–1006.

[10]. Rahman, Z.; Jobson, D.J.; Woodell, G.A. Multi-scale retinex for color image enhancement. In Proceedings of the International Conference on Image Processing, Lausanne, Switzerland, 19 September 1996; IEEE: Piscataway, NJ, USA, 1996; Volume 3, pp. 1003–1006.

[11]. Shi, B.; Bai, X.; Yao, C. Detecting Oriented Text in Natural Images by Linking Segments. In Proceedings of the Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2550–2559.

[12]. N. Zhang and J. Kim, "A Survey on Attention mechanism in NLP," in 2023 International Conference on Electronics, Information, and Communication (ICEIC), 2023, pp. 1–4, IEEE. Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 21–37.

[13]. Zhou, X.; Yao, C.; Wen, H.; Wang, Y.; Zhou, S.; He, W. EAST: An efficient and accurate scene text detector. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017, Honolulu, HI, USA, 21–26 July 2017; pp. 5551–5560.
Gildardo Mendoza Achee, Michelle Wilber, Mohammad Heidari Kapourchali, Masoumeh Heidari Kapourchali, "Day-Ahead Residential Electric Load Forecasting Using Bi-Directional LSTM Recurrent Neural Networks", 2025 IEEE Electrical Power and Energy Conference (EPEC), pp.354-358, 2025

[14]. Kalchbrenner, N.; Grefenstette, E.; Blunsom, P. A convolutional neural network for modelling sentences. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Baltimore, MD, USA, 24–27 June 2014; pp. 655–665.

[15]. Liu, Y.; Chen, H.; Shen, C.; He, T.; Jin, L.; Wang, L. Real-time scene text detection with differentiable binarization. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 11474–11481. T.-G. Marchitan, C. Creanga, and L. P. Dinu, "Transformer and Hybrid Deep Learning Based Models for Machine-Generated Text Detection," arXiv:2405.17964 [cs.CL], May 2024.

[16]. J. Puigcerver, "Are Multidimensional Recurrent Layers Really Necessary for Handwritten Text Recognition?" in Proc. 14th IAPR Int. Conf. on Document Analysis and Recognition (ICDAR), 2017, pp. 67–72.

[17]. A. D et al., "Image Text Detection and Documentation Using OCR," 2024 International Conference on Smart Systems for Electrical, Electronics, Communication and Computer Engineering (ICSSEECC), Coimbatore, India, 2024, pp. 410-414, doi: 10.1109/ICSSEECC61126.2024.10649443.

[18]. R. Avros and Z. Volkovich,"Detection of Computer-Generated Papers Using One-Class SVM and Cluster Approaches," in Proc. Machine Learning and Data Mining in Pattern Recognition (MLDM 2018), New York, NY, USA, Jul. 2018, pp. 42–55. Springer, Lecture Notes in Computer Science, vol. 10935.

[19]. K. M. Abuali, L. Nissirat and A. Al-Samawi, "Advancing Network Security with AI: SVM-Based Deep Learning for Intrusion Detection," Sensors, vol. 23, no. 21, p. 8959, Nov. 2023, doi:10.3390/s23218959.

[20]. T. M. Breuel, A. Ul-Hasan, M. A. Al-Azawi, and F. Shafait, “High-Performance OCR for Printed English and Fraktur Using LSTM Networks,” in Proc. Int. Conf. on Document Analysis and Recognition (ICDAR), 2013.

[21]. K. Vani and D. Gupta, "Using K-means cluster based techniques in external plagiarism detection," 2014 International Conference on Contemporary Computing and Informatics (IC3I), Mysore, India, 2014, pp. 1268-1273, doi: 10.1109/IC3I.2014.7019659.

[22]. J. Sauvola and M. Pietikäinen, “Adaptive Document Image Binarization,” Pattern Recognition, vol. 33, no. 2, pp. 225–236, 2000.

[23]. Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," in Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, Nov. 1998, doi: 10.1109/5.726791.

[24]. Zou, L., et al. (2023). Text Recognition Model Based on Multi-Scale Fusion CRNN. PMC, National Center for Biotechnology Information.

[25]. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 10012–10022).

[26]. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale. arXiv preprint arXiv:2010.11929.

[27]. N. Dalal and B. Triggs, “Histograms of Oriented Gradients for Human Detection,” in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2005, pp. 886–893.

[28]. C. Cortes and V. Vapnik, “Support-Vector Networks,” Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.

[29]. R. Smith, “An Overview of the Tesseract OCR Engine,” in Proc. Int. Conf. on Document Analysis and Recognition (ICDAR), 2007.

[30]. A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, and J. Schmidhuber, “A Novel Connectionist System for Unconstrained Handwriting Recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009.

[31]. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770-778).

[32]. Hochreiter S & Schmidhuber J. Long Short-Term Memory. Neural Computation, 9(8), 1735-1780.

[33]. J. Sauvola, T. Seppanen, S. Haapakoski and M. Pietikainen, "Adaptive document binarization," Proceedings of the Fourth International Conference on Document Analysis and Recognition, Ulm, Germany, 1997, pp. 147-152 vol.1, doi: 10.1109/ICDAR.1997.619831.

[34]. Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural Machine Translation by Jointly Learning to Align and Translate. arXiv preprint arXiv:1409.0473.

[35]. Cui M., Wang W., Zhang J., Wang L. Representation and correlation enhanced encoder-decoder framework for scene text recognition; Proceedings of the Document Analysis and Recognition–ICDAR 2021: 16th International Conference; Lausanne, Switzerland. 5–10 September 2021; pp. 156–170.

[36]. Chollampatt S., Ng H.T. A multilayer convolutional encoder-decoder neural network for grammatical error correction. Proc. AAAI Conf. Artif. Intell. 2018;32:1220–1230. doi: 10.1609/aaai.v32i1.12069.

[37]. Zheng T., Chen Z., Fang S., Xie H., Jiang Y. Cdistnet: Perceiving multi-domain character distance for robust text recognition. arXiv. 20212111.11011.

[38]. A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, and J. Schmidhuber,“A Novel Connectionist System for Unconstrained Handwriting Recognition,”IEEE Transactions on Pattern Analysis and Machine Intelligence,vol. 31, no. 5, pp. 855–868, 2009.

[39]. V. I. Levenshtein,“Binary Codes Capable of Correcting Deletions, Insertions, and Reversals,”Soviet Physics Doklady, vol. 10, no. 8, pp. 707–710, 1966.