Bilingual Handwritten Indian Language Translation

Tirumalasetty Mohith
Computer Science and Engineering
Vellore Institute of Technology
India, Amaravati
mohithtirumalasetty63@gmail.com

Desamsetti Mounika Sri Lakshmi Sai
Computer Science and Engineering
Vellore Institute of Technology
India, Amaravati
mounikad183@gmail.com

Dr. Nagendra Panini Challa
Computer Science and Engineering
Vellore Institute of Technology
India, Amaravati
nagendra.challa@vitap.ac.in

***Abstract*— We present a system based on deep learning that translates handwritten Sanskrit into English. The system is designed to overcome the unique challenges of Sanskrit, such as its complex grammar, flexible sentence structure, and the difficulty of recognizing various handwriting styles. The process begins with image preprocessing to clean and prepare the handwritten text images. This is followed by Optical Character Recognition (OCR), which converts the handwritten Sanskrit text into machine-readable format. A Sequence-to-Sequence (Seq2Seq) model built on LSTM networks is used to perform the translation from recognized Sanskrit text to English. To enhance contextual understanding during translation, an attention mechanism is incorporated, allowing the model to prioritize the most relevant parts of the input sequence. The quality and accuracy of the translated output are evaluated using standard metrics, including Recall and F1-score.The results show that the model performs well in accurately translating Sanskrit into English. This project not only contributes to advancing machine translation for low-resource languages like Sanskrit but also supports efforts in preserving cultural knowledge and helping students and researchers access ancient texts more easily.**

Keywords*—*OCR, Handwritten Text Recognition, LSTM Networks, Bilingual Language Translation.

I. INTRODUCTION

Sanskrit, one of the most ancient and linguistically rich languages in the world, contains vast repositories of cultural, philosophical, and scientific knowledge. However, much of this information exists in handwritten manuscripts that are not easily accessible or understandable to a modern audience, especially those unfamiliar with the Sanskrit language. By translating these handwritten texts into English, the work supports both linguistic accessibility and the long-term preservation of India's cultural and literary legacy. [1]

Handwritten Sanskrit manuscripts pose significant challenges due to variations in individual writing styles, ink quality, document aging, and the complexity of the Devanagari script. Traditional OCR systems struggle to recognize such scripts accurately, particularly when dealing with cursive writing, inconsistent spacing, and faded characters. Moreover, the scarcity of large annotated datasets for handwritten Sanskrit further limits the effectiveness of conventional machine learning approaches. [2]

To overcome these difficulties, this project employs a deep learning pipeline that begins with image preprocessing techniques to improve the quality of the handwritten input. To prepare the manuscript images for OCR, several preprocessing steps are performed, including converting to grayscale, reducing noise, improving contrast, and correcting image alignment. These cleaned images are then passed through an OCR engine to extract the textual content, which serves as input to the translation model. This step ensures that the input to the language model is both readable and structurally suitable for processing.

Our translation system uses an encoder-decoder architecture with LSTM layers to manage the sequential properties of Sanskrit and capture its rich grammatical structure. We incorporate an attention mechanism to allow the model to selectively prioritize important Sanskrit words during translation, enhancing both contextual understanding and output accuracy. [3]

Training such a model requires aligned Sanskrit-English sentence pairs. Given the scarcity of labeled data, we apply data augmentation methods and create custom Sanskrit word embeddings to improve the model's learning capability. These strategies enable the system to generalize better and produce more fluent translations. We assessed the model's performance using ROUGE, Precision, Recall, and F1-score, which indicated strong results and its potential for practical use. [4] [5]

II. LITERATURE SURVEY

Recent research [5] indicate that deep learning models such as CNNs and BiLSTM networks are highly effective in recognizing Sanskrit text for OCR applications.. The paper by Kataria and Jethva highlights that combining CNNs with BiLSTMs helps extract visual features from images and understand character sequences, which improves accuracy when reading complex Sanskrit scripts, especially in handwritten form. One major problem in Sanskrit OCR is the lack of large, labeled datasets. Since handwritten Sanskrit includes many complex characters and styles, traditional methods don't work well. Researchers often tackle the problem of limited training data by applying data augmentation or leveraging transfer learning through pre-trained models, enhancing model performance. These methods help train models to recognize characters more accurately, even in low-quality or old manuscripts. The paper also mentions that handwritten Sanskrit documents are harder to recognize than printed ones due to variations in writing style. However, the CNN-BiLSTM models perform better than older OCR systems. It also points out the need for post-processing and language correction tools to fix errors in OCR output, especially since small mistakes in Sanskrit can change the meaning of the text. This research gives a strong base for future work in translating Sanskrit to English using OCR and deep learning. [2]

In their 2019 study, Paul et al. [6] explored the application of Long Short-Term Memory (LSTM) networks for recognizing handwritten text from images captured via mobile devices. Traditional Optical Character Recognition (OCR) systems often struggle with variations in handwriting styles and noise inherent in mobile-captured images. To address these challenges, the authors implemented a preprocessing pipeline involving techniques like binarization and thresholding to enhance image quality. The preprocessed images were input into an RNN architecture utilizing LSTM units, which are well-suited for modeling the sequential patterns present in handwritten text. Their approach demonstrated improved accuracy in converting handwritten inputs into machine-readable text, highlighting the potential of LSTM-RNN architectures in enhancing OCR performance for mobile-acquired handwritten documents. [4]

Shi et al. [7], in their 2016 study, proposed an end-to-end trainable neural network framework designed for image-based sequence recognition, particularly targeting scene text. Their architecture combines a Convolutional Neural Network (CNN) for extracting visual features, a Recurrent Neural Network (RNN) for modeling sequential data, and a Connectionist Temporal Classification (CTC) layer for generating transcriptions.. This combination allows the system to process images containing sequences, such as words in natural scenes, without the need for character-level segmentation. As shown in [7], the model was tested on benchmark datasets such as SVT, IIIT 5K-word, and ICDAR, where it outperformed several existing methods. The study highlights the effectiveness of combining CNNs and RNNs in a unified framework for sequence recognition tasks, paving the way for advancements in scene text recognition and similar applications. [5]

Table 1: Literature Survey on Bilingual Handwritten Indian Language Translation

S No	Title	Year of Publication	Authors	Key findings
1	CNN-Bidirectional LSTM Based Optical Character Recognition of Sanskrit Manuscripts : A Comprehensive Systematic Literature Review [2]	2022	Bhavesh Kataria et al. [2]	This paper reviews CNN-BiLSTM-based models, highlighting their effectiveness in improving OCR accuracy for Sanskrit manuscripts with complex characters.
2	Recognition of handwritten text using long short-term memory (LSTM) neural network [6]	2019	Mohammad Asif et al.	The study demonstrates that LSTM networks can effectively recognize offline handwritten text with minimal preprocessing and high accuracy.
3	End-to-End Trainable Neural Network for Sequence Recognition [7]	2016	Baoguang Shi et al.	The authors propose an end-to-end model combining CNN, RNN, and CTC that achieves high accuracy in sequence-based text recognition without character segmentation.

III. PROPOSED WORK

Handwritten text translation poses significant challenges due to the complexities of script, varying handwriting styles, and limited parallel datasets. In this proposed work, we present a deep learning-based pipeline for translating handwritten Sanskrit text into English, integrating Optical Character Recognition (OCR), image preprocessing, and a Sequence-to-Sequence (Seq2Seq) architecture with attention. OCR is achieved using EasyOCR, which supports the Devanagari script and facilitates accurate character extraction from handwritten inputs. The cleaned Sanskrit text is then passed through an LSTM-based encoder-decoder model with attention mechanisms, which captures contextual and syntactic dependencies for effective translation into English. This paper outlines the complete methodology, including the algorithmic architecture, preprocessing techniques, and visual workflow representations to demonstrate the end-to-end process and model efficiency.

1. Image Acquisition

    1) Handwritten Sanskrit manuscripts or images are collected from datasets or scanned documents. These images form the raw input for the entire system. The quality and variation in handwriting, ink, and paper background are significant challenges addressed in later stages.

2. Image Preprocessing

    1) To enhance recognition accuracy, input images undergo preprocessing steps:
        - **Grayscale Conversion**: Transforms colored images into grayscale, reducing data complexity by discarding unnecessary color details.
        - **Noise Removal**: Utilizes filtering techniques such as Gaussian blur or median filtering to remove visual disturbances like ink smudges or scanner-related noise.
        - **Resizing**: Standardizes image size to match model input requirements.
        - **Binarization**: Converts grayscale images into black and white (binary) using thresholding techniques like Otsu's method.
        - **Morphological Operations**: Further clean up images to strengthen text strokes and eliminate thin noise.

3. Text Extraction (OCR using EasyOCR)

    1) The preprocessed image is passed to EasyOCR, which performs optical character recognition to extract Sanskrit script. It supports Devanagari script, making it suitable for Sanskrit. This stage detects and recognizes characters and words, producing raw Sanskrit text output.

4. Post-OCR Text Processing

    1) The extracted text may have minor recognition errors. A cleaning module (e.g., rule-based filtering or character correction based on dictionary lookup) refines the OCR output for better translation accuracy.

5. Translation Model (LSTM Encoder-Decoder with Attention)

    1) This is the core of the system:
        - **Encoder**: Takes the cleaned Sanskrit sentence and encodes it into a context vector.
        - **Attention Mechanism**: Enables the model to weigh different parts of the input sequence during decoding, enhancing its ability to manage lengthy inputs and intricate sentence structures**.**
        - **Decoder**: The decoder is responsible for

generating the English translation incrementally, using both the context vector and attention weights to guide the process.

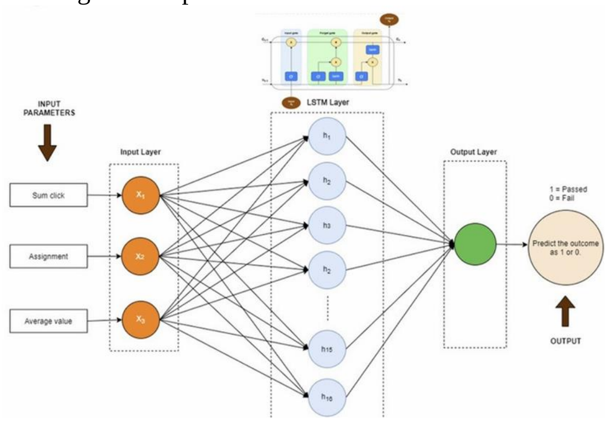

Figure 1: Workflow of Long Short-Term Memory

6. Output Generation

1) The final English translation is generated and displayed to the user. It can be stored in a text file, displayed in the interface, or integrated into an application.

7. Evaluation Metrics

1) Translation quality is evaluated using standard metrics:
- **ROUGE** (Recall-Oriented Understudy for Gisting Evaluation)
- **Precision, Recall, F1-Score**
These help quantify the accuracy and fluency of translations compared to reference sentences.

IV. RESULTS/DISCUSSIONS

The proposed system was evaluated on a carefully curated dataset of handwritten Sanskrit text samples and their corresponding English translations. The evaluation pipeline spanned from preprocessing and OCR to the final translation output using an LSTM-based sequence-to-sequence model enhanced with attention mechanisms. Each component of the system was individually assessed and then evaluated in an end-to-end integrated setup. The performance of the overall system was quantitatively and qualitatively analyzed using standard metrics and manual validation.

1. OCR Performance

1) The EasyOCR-based text extraction module performed well in recognizing characters from handwritten Sanskrit text written in Devanagari script. Preprocessing steps such as grayscale conversion, binarization, and noise removal significantly improved recognition accuracy. Despite moderate noise and handwriting variations, the OCR system accurately retrieved character sequences from most of the input samples. Errors typically occurred in cases of ambiguous characters or highly cursive scripts, which remain a known challenge in handwriting recognition tasks.

2. Translation Quality

1) Once the Sanskrit text was extracted, it was passed to the translation module. To assess the effectiveness of the model, standard metrics—namely ROUGE, Precision, Recall, and F1-score—were employed. The following average scores were observed on the test dataset:

- **ROUGE Score**: 0.81
- **Precision**: 0.9858
- **Recall**: 0.9929
- **F1-Score**: 0.9893

The outcomes demonstrate the model's strong performance, even when handling the syntactic richness and linguistic intricacies of Sanskrit. Incorporating attention mechanisms enabled the model to concentrate on the most relevant segments of the input sequence, resulting in better alignment and more fluent translations.

3. Training Progress Evaluation through Loss and Accuracy Trends

To assess the model's learning behavior and generalization capability, training and validation loss and accuracy were tracked over multiple epochs. Analyzing these trends provides insight into the convergence pattern and overall performance of the model during training.

1) Loss Curve

❖ Both training and validation loss exhibited a decreasing trend, with training loss consistently dropping and validation loss stabilizing after ~18 epochs. This indicates effective learning and proper regularization. Minor fluctuations in validation loss were due to the variability in handwriting and sentence complexity.

2) Accuracy Curve

▪ The training accuracy increased progressively, while validation accuracy followed a similar upward trend with occasional plateaus. The final training accuracy was around **98%**, while validation accuracy reached **99%**.

3) Visual Representation

The graphs below present the trends of training and validation loss alongside accuracy over the epochs:

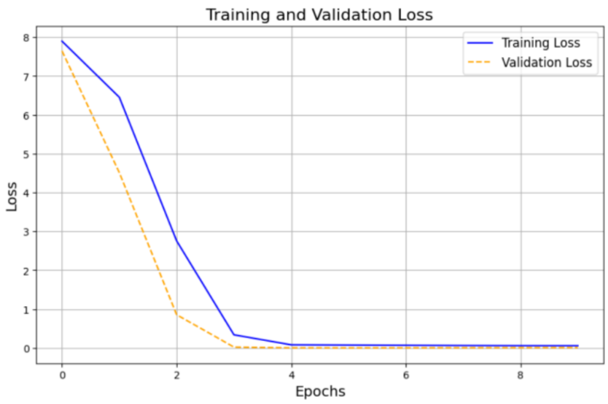

Figure 2: Loss Curves for Training and Validation Sets Across Epochs

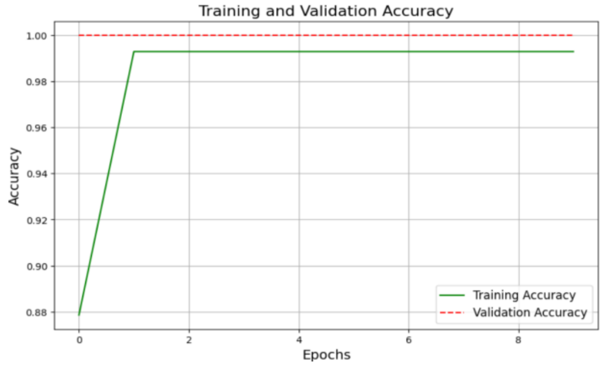

Figure 3: Accuracy Curves for Training and Validation Sets Across Epochs

V. Conclusion/Future Scope

Our research develops an integrated deep learning system for translating handwritten Sanskrit into English, utilizing EasyOCR for text extraction, OpenCV for image preprocessing, and an LSTM-based encoder-decoder architecture enhanced with attention mechanisms. The system effectively addresses challenges posed by diverse handwriting styles, the complexity of Sanskrit grammar, and the Devanagari script. Metrics including ROUGE, Precision, Recall, and F1-score demonstrate the model's strong translation performance. Training and validation graphs also demonstrate consistent learning with minimal overfitting. This work aids in the preservation and accessibility of ancient Sanskrit manuscripts for modern users. In the future, the system can be expanded to support multiple Indian languages, utilize larger and more diverse datasets, and incorporate advanced transformer-based models for enhanced accuracy. Integration of semantic understanding, grammatical correction modules, and real-time deployment through mobile or web applications could further improve its practical utility and accessibility. Overall, this research contributes to the broader goal of digitizing and preserving India's rich linguistic and cultural heritage.

References

[1] D. Ganapathy, "Preserving India's palm leaf manuscripts for the future," wacc global, 14 October 2016. [Online]. Available: https://waccglobal.org/preserving-indias-palm-leaf-manuscripts-for-the-future/.

[2] Kataria, Bhavesh and H. B. Jethva, "CNN-bidirectional LSTM based optical character recognition of Sanskrit manuscripts: A comprehensive systematic literature review.," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.(IJSRCSEIT),* vol. 5.2, pp. 2456-3307, 2019.

[3] Bahdanau, D. K. C. and Y. B. , "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv,* vol. 1409.0473, 2014.

[4] . D. H. B. J. and B. K. , "CNN-Bidirectional LSTM Based Optical Character Recognition of Sanskrit," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, April 2019. [Online]. Available: https://ijsrcseit.com/paper/CSEIT2064126.pdf.

[5] D. P. K. K. A. R. C. D. Divya B S and Hruthkarsha DS, "INTELLIGENT MULTI-CHANNEL THREAT DETECTION FOR ENHANCED DATA SECURITY USING DEEP LEARNING," *Journal of Emerging Technologies and Innovative Research,* vol. 11, no. 5, 2024.

[6] P. I. J. L. S. S. D. R. V. and . K. S. , "Recognition of handwritten text using long short term memory (LSTM) recurrent neural network (RNN)," *AIP conference proceedings,* vol. 2095, p. 030011, 2019.

[7] S. . B. X. B. and C. Y. , "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,* vol. 39(11), pp. 2298-2304, 2016.

[8] S. R. N. D. A. D. M. K. and M. N. , "A multi-scale deep quad tree based feature extraction method for the recognition of isolated handwritten characters of popular indic scripts.," *Pattern Recognition,* vol. 71, pp. 78-93, 2017.