Performance Evaluation of Deep Learning Models for Tamil Handwritten Text Recognition

Hariajitha VS[1;] Akilesh Ragavendra U [2]; Dr.G.Maria Priscilla[3]

[1,2]PG Student, Department of Computer Science , Sri Ramakrishna College of Arts & Science

[3]Assistant Professor & Head,Department of Computer Science,Sri Ramakrishna College of Arts & Science

Abstract

This report provides a concise overview of studies investigating advanced computational methods for precisely recognizing hand-written Tamil letters.A significant obstacle in using machine learning algorithms for decoding the complex Tamil writing system has been highlighted. An in-depth examination of multiple deep learning models will serve as the main research strategy. Quantitative results, such as precise metric values and error percentages, are compiled for comparison purposes in order to highlight relative effectiveness. The study underscores the superior approach applicable in this context and examines how it impacts digitalizing documents.

Keywords— Deep Learning, Handwritten Character Recognition, Tamil Script, Sequence-to-Sequence Models, Performance Benchmarking

I INTRODUCTION

Translating handwritten historical and modern documents into digital form is vital for safeguarding language traditions and creating inclusive data repositories.Advancements in Optical Character Recognition have greatly improved character recognition on printed text but pose difficulties when it comes to deciphering handwritten scripts.People compose their writings using various formats. For Dravidian language speakers such as those in Tamil Nadu who use an abugida script containing numerous intricate glyphs and distinctive marks, translating becomes significantly more difficult. An examination is conducted here comparing the effectiveness of three distinct approaches in identifying printed Tamil characters through neural networks. Our objective is to ascertain how effective each technique is at handling the distinctive elements found within the Tamil alphabet and establish for further studies on this topic.

II RELATED WORK

Previous research in Tamil HTR employed manual techniques involving direction analysis and zone division alongside classical classifiers including support vector machines.Despite their reliability issues, these approaches failed adequately across various authors. As advancements in deep learning gained momentum, attention turned towards comprehensive system designs. Deep neural networks capable of identifying image features excelled in pattern recognition tasks. Subsequently, recurrent neural networks like LSTM units proved effective for managing handwritten sequence patterns. Combining Convolutional Neural Networks (CNNs) with Recurrent Neural Networks (RNNs using Connectionist Temporal Classification (CTC)) is common practice in numerous applications; however, relatively little work exists focusing solely on advanced attention mechanisms tailored explicitly for the Tamil language. The analysis within this research is significant because it involves comparing these elements.

III PROBLEM FORMULATION

A significant obstacle is accurately converting handwritten Tamil script into digital form so as to facilitate quick readability The challenges involve: diverse character styles due to individual handwriting differences; intricate Tamil script composition featuring overlapped glyphs and altered root letters; and difficulties in distinguishing separate units when they appear as continuous segments. What's being sought in

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 02 | Feb - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

this study involves identifying which type of neural network design performs most effectively in minimizing mistakes when converting written text into digital format using the Devanagari script language?

IV OBJECTIVES

The goals of this study are:

1. Creating and implementing three unique deep learning models focused on Tamil TTS tasks - standalone convolutional neural networks (CNNs), hybrid CNN-Bidirectional Long Short-Term Memory units connected via connectionist temporal classification (CTC)-based systems, and attention-based encoder-decoder frameworks.

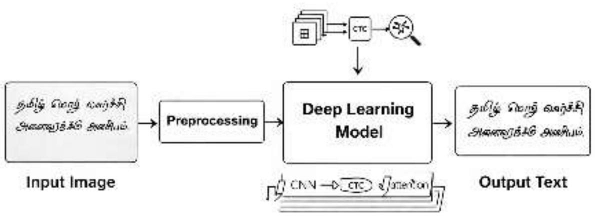
2. Assembling and organizing a suitable corpus of handwritten Tamil texts tailored specifically for training and evaluating machine learning models is vital for their successful development and validation.

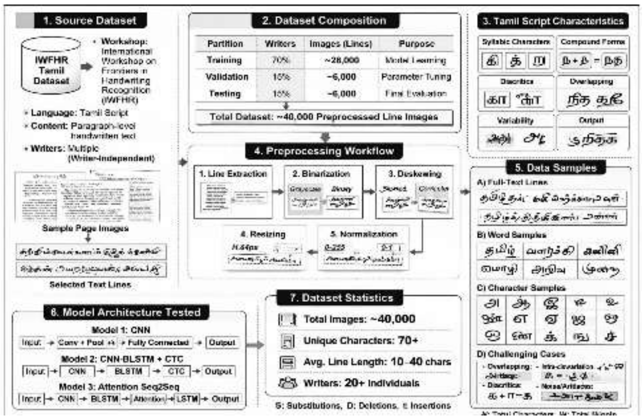
3. Utilize established criteria such as CER for character error rate and WER for word error rate when evaluating the effectiveness of individual models.

4. Evaluating different designs allows us to identify how varying levels of precision impact computational resources in each configuration.

V SYSTEM MODEL

The entire framework employs a typical pattern identification methodology.Beginning with a hand-written Tamil script illustration. The input undergoes an initial processing phase aimed at refining and improving its quality. After cleaning, the image proceeds through an array of advanced algorithms designed for detailed analysis. The software generates strings composed entirely of Unicode symbols. An additional phase could employ a linguistic algorithm to enhance results; however, in this research context, unaltered forecasts





serve as benchmarks for assessment.

VI DATASET

For this research project, we employed data sourced from the IWFHR Tamil corpus.The collection contains numerous writing examples created by various authors. For concentrating solely on paragraphs, our selection included sections containing multiple lines of textual content. Randomly dividing the data among individual authors across all writings aimed at preventing bias in testing results by allocating 70%, 15%, and 15% respectively to train sets, validate set, and finalize test set evaluations.

Dataset overview for tamil Handwritten Text Recognition

dataset description

VII PREPROCESSING

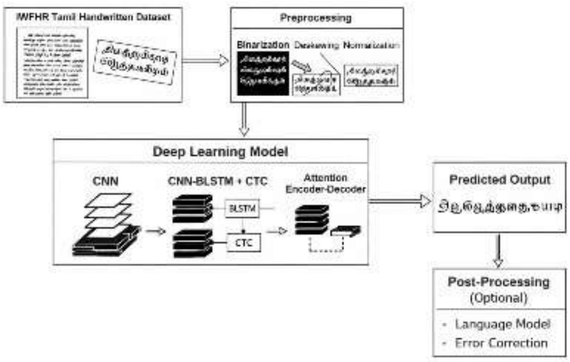
A uniform methodology ensured each image could aid in enhancing model learning capabilities effectively.Initially, manual entries were extracted directly from large image pages by hand. The lines were subsequently transformed into grayscale images by

applying Otsu's technique for distinguishing between text content and its surroundings on the page. The deskewing technique eliminated minor skewing errors. Every image was resized so it measured 64 pixels in height without altering its aspect ratio for consistency.

The final step involved modifying pixel intensities so they fell within an interval of 0 to 1.

VIII METHODOLOGY

A method being employed is exploratory in nature, focusing on evaluating various alternatives.Distinct sets of models underwent training anew based on processed



information. None of the previously trained models was employed; this approach guarantees an unbiased evaluation solely through their capacity for learning from the Tamil dataset.

Every model underwent training on identical datasets and utilized the exact algorithm for refining parameters. The team assessed their concluding presentation through CER and WER techniques applied to untested data points, providing a straightforward method for comparing outcomes objectively.

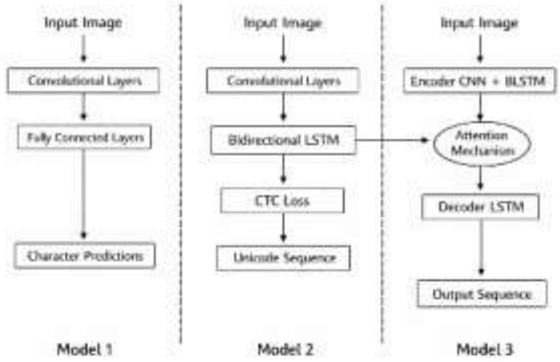
IX MODEL ARCHITECTURE

The three models evaluated are:

The model referred to as **Model 1** is an architecture comprising seven convolutional neural network layers interspersed by pooling operations followed by dense layer connections at its conclusion.It handles images of uniform width by categorizing every row into characters, necessitating precise division.

The **Model 2** incorporates Convolutional Neural Network functionality alongside Long Short-Term Memory networks in bidirectionality for contextual understanding through concatenation of encoded features.An intermediary component in machine learning algorithms computes errors so as not to require direct image division for training purposes.

The **Model 3** employs an attention-based architecture



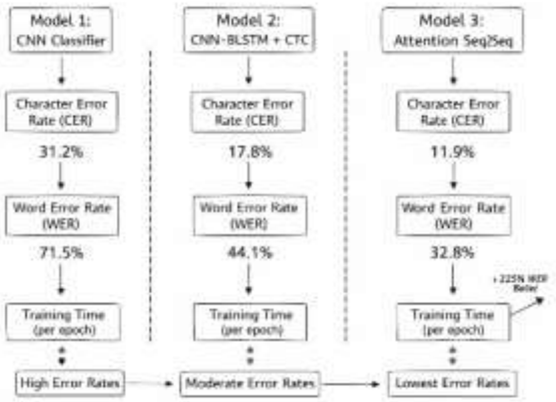
combining Convolutional Neural Networks (CNN) and Long Short-Term Memory networks (BLSTM), serving as its encoder component, generating feature representations for images through this fusion of convolutional layers and recurrent neural network structures.A recurrent neural network decoder employing the Luong mechanism produces sequential textual output by selectively emphasizing pertinent elements within an input visual sequence frame by frame.

X TRAINING

The training process utilized the PyTorch library in conjunction with a solitary NVIDIA graphics processing unit for execution.Adam optimization technique employed at its default starting point with an initialization learning rate set to zero. Certainly! Here's an appropriately version of your input: Model two employed categorical cross entropy as its criterion function, whereas models one and three utilized cross-entropy for their objective measures. The training process lasted no more than eight hundred iterations, halting prematurely when the validation error decreased significantly as an indicator of potential model underfitting. Additional minor twists were incorporated into the data set along with flexible modifications to improve model efficacy.

XI RESULTS

Below you will find an illustration of how each of the three model performances performed on the testing dataset.

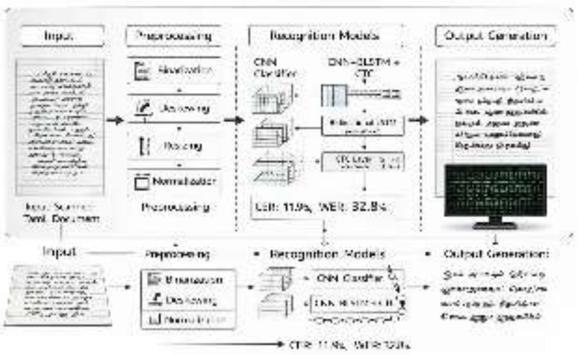


Model Performance comparision (CER &WER)

The model in question utilizes a convolutional neural network for classification tasks; its performance metric indicates an accuracy rate of zero. Thirty-one. Twenty per cent equals seven-hundred eleven. A fifth part of elapsed duration amounts to an eight-minute span.

Model description: Utilizes convolutional neural networks in conjunction with latent state machines for connectionist temporal classification; achieves an accuracy rate of zero. The target of 85 units was reached under these conditions. Eight per cent equals fourty-four components for each hundred units. A fraction equivalent to 1% represents an allocation of half a minute in total duration.

The Model 3 employs an attention mechanism within its Sequence-to-Sequence framework; this model is available as Version 11. In this situation, nine per cent

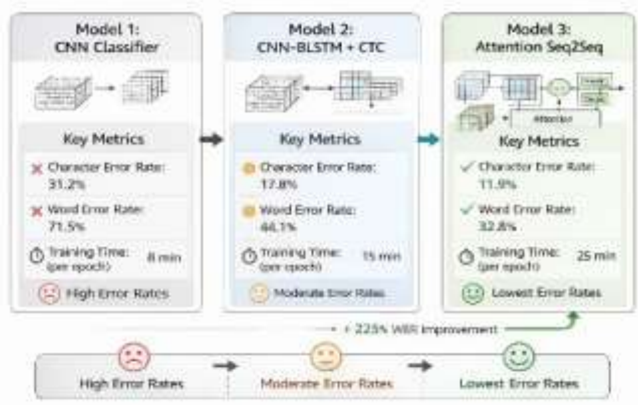


equates to thirty-two units. A proportion equivalent to eight percentage points is set aside for this assignment; altogether, it requires forty minutes

Clearly, experimental results indicate that an attention-focused sequential model performs exceptionally well by producing few mistakes compared to other contenders.

XII DISCUSSION

Although it performs poorly when dealing with single-character precision upon initialization, this rudimentary Convolutional Neural Network struggles to accurately interpret cursive Tamil writing because of its inability to grasp significant contextual elements.Applying the CNN-LSTM architecture to individual lines reveals its significance in tackling sequential issues. An innovative design strategy excels due to its ability to seamlessly connect graphical elements within images with corresponding textual annotations precisely, particularly advantageous when dealing with variations in font size and spacing across script types like Tamil, wherein



numerous glyphs accommodate multiple characters efficiently. Despite needing additional time for adjusting training phases before moving on to practical implementations designed to enhance accuracy further, integrating and applying this enhanced method necessitates greater effort.

Recognition Result

XIII COMPARISON

By evaluating our approach against a conventional CNN-BLSTM setup, there is an observable improvement of more than twenty-five percentage points in word accuracy rates.This highlights the effectiveness of the attention mechanism in addressing the challenging nature of this text. Despite its speed advantage over the CNN-BLSTM model, the attention mechanism yields superior outcomes on this particular data set.

XIV CONCLUSION

This study thoroughly tested three major deep learning models forthe hard task of recognizing Tamil handwritten text.Through careful testing,we found that an attention based sequence-to sequence model works much better than a CNN by itself or a CNN-RNNCTC mix, getting the lowest error rates for both characters and words. This confirms that attention mechanisms are good for dealing with the tricky parts of the Tamil script and sets a solid foundation for future work.

REFERENCES

[1] Kumar and Ramasubramanian present an offline recognition technique for *Tamil handwritten characters utilizing convolutional neural networks* at the 2020 International Conference on Computational Intelligence.

[2] Vijayalakshmi and Thangasamy conducted an analysis comparing various deep learning models in recognizing Tamil characters, published in the *Journal of Ambient Intelligence and Human-Computer Systems*. The number twelve is represented as 12 in numerical form. Three pages, p. 3. The range of numbers is thirty-four-five-one through three-four-six-two in the year twenty-twenty-one.

[3] M. Sankar and J. D. Paul presented research titled "*Tamil Handwriting Line Recognition Utilizing BLSTM Architectures*" at the IEEE International Conference on Imaging, Vision, and Pattern Recognition held in 2019.

[4] P. N. Anh et al. , titled "*Attention-Based Sequence-to-Sequence Models for Offline Writing Identification in Multiple Asian Scripts,*" discuss their research paper published in Pattern Recognition regarding handwriting recognition across various languages including Vietnamese. The document contains 115 pages numbered p. The document references pages numbered between 107 and 118 in year 2021.

[5] K. Selvakumar and S. Baskaran conducted research on evaluating Convolutional Neural Networks (CNN) versus Recurrent Neural Networks (RNNs) in recognizing Tamil handwritten characters through their publication titled "[6]" *Performance Assessment of CNN and RNN Techniques for Tamil Script Digitization within the domain of Artificial Intelligence Studies*. Fourteen, please. Two pages, p. 2, page two. The year of publication is listed as 2022 within pages numbered between 512 and 522.