Minimax Theory and its Applications
Volume 11 (2026), No. 4s, 939 to 951

Attention Guided ResNet–BiLSTM with CTC Optimization for Robust Handwritten Marathi Character Recognition on MHCD GIETV2

Shazid Wahid Khandakhani[1], Sachikanta Dash[1,4*], Sasmita Padhy[2], Rabinarayan Panda[3]

[1]Department of CSE, Gandhi Institute of Engineering and Technology University, Gunupur, India
[2]School of Computing Science and Engineering, VIT Bhopal University, Bhopal, India
[3]Department of Engineering, Garden City University, Bengaluru, India
[4]Madanapalle Institute of Technology and Science Deemed to be University, Mdanapalle, AP, India
***Corresponding Author**: Sachikanta Dash, Email: dash.sachikanta@gmail.com

Abstract
Handwritten Character Recognition (HCR) for In-dic scripts remains a challenging research problem due to the presence of ligatures, compound characters, and spatially distributed diacritical marks. These challenges are particularly prominent in Marathi, a widely used language based on the Devanagari script. To address these issues, this study proposes a hybrid deep learning framework that integrates ResNet-50, BiLSTM, an attention mechanism, and Connectionist Temporal Classification (CTC) loss. ResNet-50 is employed to extract hierarchical spatial features, while BiLSTM networks model the sequential dependencies within handwritten character patterns. An attention mechanism is incorporated to emphasize informa-tive regions, particularly those containing diacritical features, and CTC optimization enables end-to-end sequence recognition without requiring explicit character segmentation. The proposed framework is evaluated using the Marathi Handwritten Charac-ter Dataset (MHCD), comprising 92,000 samples collected from 460 native writers across 50 character classes. The dataset is divided into training, validation, and testing subsets using a 70:15:15 ratio. Rotation, scaling, and elastic distortion techniques are applied to improve robustness through data augmentation. The proposed model achieves 97.8% character-level accuracy and 95.3% word-level accuracy, outperforming the CNN-only baseline (93.1%) and CNN-BiLSTM model without attention (94.2%). Furthermore, the framework demonstrates approximately a 6% reduction in relative error under noisy input conditions. These results demonstrate the effectiveness of combining spatial feature extraction, sequential modeling, attention-based feature refine-ment, and CTC-based optimization for handwritten Marathi recognition. The framework also provides a promising foundation for extending HCR systems to other low-resource Indic scripts and applications such as digital archiving, automated document processing, and assistive technologies.

Keywords*:* HCR, MHCD, Marathi Script, Hybrid Deep Learning, ResNet-BiLSTM, Attention Mechanism, CTC Loss.

Introduction

Handwritten character recognition (HCR) is one of the most enduring and demanding tasks in pattern recognition and computer vision. Despite considerable advancements in scripts like Latin and Chinese, Indic scripts such as Devanagari and its variants are relatively underexamined due to their structural intricacies and substantial variety in handwriting. Marathi, an official language of India with over 83 million speakers, is inscribed in the Devanagari script, comprising vowels, con-sonants, compound letters, and diacritical marks (matras) that are positioned above, below, or around the base characters [1]. The structural characteristics result in significant heterogeneity in handwritten input, complicating the recognition of Marathi script relative to Latin alphabets [2].

Traditional methods for character recognition rely primarily on manually designed features. The methods include zoning, chain codes, Histogram of Oriented Gradients (HOG). clas-sifiers like Support Vector Machines (SVMs) and k-Nearest Neighbours (k-NN) [3], [4] etc. also used for recognizing char-acters. All these works are pretty well for numerals and even for a single character. But for the case of compound character it is a bit shaky due to the noise and the changing styles [5]. Then it was DL, and particularly Convolutional Neural Networks (CNNs), and it really did help HCR, as these can learn features themselves straight from the raw input images [6]. In case of printed characters, CNN-based recognition is quite accurate than the existing technique [7], [8]. Recently, CNNs have been applied to handwritten Marathi text such as Modi script identification [9]. Further, ensemble learning techniques for Marathi handwritten character recognition are reported with approximately 98% recognition rate [10].

But most of the studies today appear to be somewhat short on scope, don't you think? They typically focus on limited datasets (numbers or perhaps just characters), and they don't adequately control diacritical variability. Some work relies solely on CNN designs, which are less capable of capturing long-range correlations, such as context. There are a few script-agnostic approaches that demonstrated good transfer learning capability across Indic scripts, such as HCR-Net [11]
but very few solutions are customised for Marathi. Further-more, it is not easy to obtain large and diverse handwritten Marathi datasets, or only limited ones. Although there are some initiatives such as the Marathi OCR Dataset [12], and a few smaller compilations, they seem too small in comparison to Latin or Chinese datasets. At last, it is more difficult to train the newer DL models.
Also, despite the recent developments in OCR, the exist-ing handwritten Marathi recognition (HCR) systems exhibit somewhat striking inherent limitations due to the complexity of the Marathi script and its behavior. Previous studies have reported word-level results of about 80-88% and character-level results of about 92-96%, becoming significantly worse when the writing becomes more challenging. Indeed, for very cursive handwriting, performance can be reduced by 3-6% due to the fact that the split points are no longer distinct, and when there are overlapping or compound characters, an additional 4-7% is lost since the split points become somewhat obscurred, just the split points sort of murky. In addition the models are prone to mispronounce the diacritics – the tiny marks that can appear above, below, and around the base letter. For example, the same character or compound characters which have a small stroke variation are often classified in the wrong class because the feature separation is too coarse-grained at this fine level.
To overcome these drawbacks, we have proposed a com-bined deep learning architecture where the hierarchical feature extraction is performed using a ResNet-50 network, sequence modelling is achieved by a BiLSTM network and the attention mechanism highlights or points out the modifier regions. Then we use a CTC loss function, which is only alignment-free across the entire dataset. The intent of the design is to address the diacritic bearing nature of the Marathi language script along with minimising the need to rely on explicit segmentation to the greatest extent possible.
The primary contribution of the study is as follows:

- To propose an innovative hybrid DL architecture that integrates ResNet-50 for hierarchical feature extraction, BiLSTM for sequence modelling, an attention mechanism for diacritic-sensitive emphasis, and CTC for alignment-free recognition.
- To gather and use an expansive Marathi Handwritten Character Dataset (MHCD) made up of 92,000 samples across 50 categories from 460 writers , so we can ensure a noticeable spread that is both within a single writer as well as between writers.
- To do broad experiments compared to baseline CNN and CNN-BiLSTM setups, showing improvements in recog-nition accuracy aswell as a bit more resilience against noise, and better generalisation of the results too.

The outline of the manuscript has Section 2, associated with the current state of the art in HCR, particularly Indic scripts, and a combination of deep learning models are discussed. In section 3, we step through the process of collecting the data, preparing it, and the different tricks used to augment the dataset. Together, we will describe the hybrid design that we propose in Section 4, with details for each building block
- ResNet, BiLSTM, attention, and CTC. The experimental setup, the criteria for the evaluation of the models, and thebaseline data for comparison of the different models are of major importance and are part of section 5. The results are displayed in Section 6, the discussion is included and the errors that occur during recognition are mentioned in Section 6, along with robustness testing. Section 7 discusses the implications for actual practice. Further it focuses on how development could progress beyond this. Finally, Section 8 conclude all the ends and suggests some directions for future research.

Literature Study

A large number of scholarly articles were used for the review on HCR. These studies are mostly concerned with the structure of Marathi characters and other complex characters. The recognition of marathi script is a challenging task due to large number of symbols, compound symbols, number of different modifiers, variation in handwriting. The literature review comprised previous studies on HCR, conventional Ma-chine Learning and Deep Learning techniques. It also included Transfer learning, Attention based models, Light weight neural networks, and Marathi OCR specific techniques.

Traditional Machine Learning Approaches

Earlier research work relied on manual feature engineering plus traditional classifiers. The methods put forward [1], [4], a weighted SVM classifier, optimized via the Sine Cosine Algorithm (SCA) for HMC recognition. In practice, these kinds of approaches can look reasonably good when the char-acters are clean and isolated. however, they tend to generalize poorly once curvature changes, stroke continuity wobbles, or diacritic positions shift. Meanwhile, classical ML approaches get tangled with segmentation issues, so they are not a great fit for word level recognition in Marathi and Devanagari scripts.

Deep CNN-Based Models

Custom Deep CNN architectures began replacing traditional methods due to their superior ability to learn hierarchical visual features. Khandakhani et al. [13] successfully applied a unique CNN model for the recognition of Marathi handwritten characters and proved a significant rise in the performance of such

systems. The architectures based on GoogLeNet, as proposed by the one studied by Shelke et al. [14] and Sekaran et al. [15], further evidence the power of CNNs, especially when it comes to Sanskrit characters that have similarities with Marathi language in terms of structure. Nevertheless, standard CNNs cannot utilize the temporal or contextual relationships that are crucial for the differentiating of intricate ligatures and positioning of the modifiers.

Transfer Learning Approaches

A multitude of researchers considered transfer learning as a means to reduce the problems associated with small datasets. Dasari et al. [16] investigated the efficacy of transfer learning models like VGG16 and ResNet 50 on Devanagari scripts. Singh et al. [17] utilized VGG16 for the recognition of offline handwritten Devanagari words. The same kind of transfer learning methods has proved their worth in the case of Modi Lipi OCR, as revealed by Ingavale and Patil [18]. Though transfer learning is a powerful tool and gives a good performance as a baseline, the models of the pretrained ones trained on natural images, in most cases, cannot grasp the script-specific variations, thereby resulting in word level recognition being not optimum in this scenario.

Hybrid CNN-RNN and CTC-Based Architectures

CRNN (Convolutional Recurrent Neural Network) models that integrate CNN feature extraction with the recurrent layers are a major breakthrough in Indic OCR. Mathew and Jawahar
[19] have performed an exhaustive investigation on CTC-based architectures for several Indian languages, and their results indicate that CNN-RNN-CTC systems are better than conventional CNN models for recognition of text without segmentation. This suits Marathi's structural needs as shi-rorekha continuity and overlapping modifiers do not allow easy segmentation of characters.

Attention-Based Models

Attention mechanisms have played an important role in optimizing the recognition accuracy of complex scripts. Chik-murge and Raghunathan [20] introduced a self additive at-tention method for Marathi handwriting recognition(MHR), which surpassed non-attention based CNN models in perfor-mance. Peng [21] explored attention boosted CNNs model for Chinese handwritten characters and shows the usefulness of these methods visually complex scripts. The attention mechanism helps the model to zero in on the diacritics, tiny vowel modifiers, and those stroke endings. In this way it ends up helping recognition of visually similar character pairs in the same script; basically, like the model is more selective than before.

Lightweight and Efficient Architectures

Efficiency oriented approaches have become popular for mobile OCR applications. Deshpande et al. [22] introduced LeanConv SENet, a lightweight model that manages to get good accuracy while keeping the computational load reduced. Hussain et al. The authors in [11] proposed a MobileNetV2-based architecture for Bangla OCR, and they show that effi-cient models can still work well for Indic scripts. They said, lightweight models often have less representational depth, and because of that it was too difficult to deal with Marathi's complex compound characters.

Synthetic Data Generation Using GANs

To augment datasets, GANs, have been looked at for making synthetic handwriting. Lokhande et al. [23] used deep convo-lutional GANs to generate synthetic Devanagari samples, so the OCR robustness got better. Even so, while GAN-produced samples can broaden the training set adequately. They also can bring artificial strokes and some background noise, which, in turn makes it a bit hard to rely only on synthetic data.

Script-Specific and Multilingual OCR Studies

Research has been done on Devnagari variants and on multilingual OCR systems. Ranjan et al. [24] suggested DL models for Hindi OCR. To identify Devanagari scripts, Gupta et al. [25] applied traditional AI and ML techniques. Desh-mukh et al. [26] proposed a real time Marathi Sign Language recognition system. This work demonstrates the versatility of deep learning with respect to script based tasks. These studies were conducted together demonstrate the broad applicability of DL in Indic scripts. They also stress the special characteristics of Marathi writing systems.

Research Gaps

From the literature, a couple of key limitations can be spotted like this, and summarized together as follows:

- Most CNN based model lacking sequential modeling capability.
- Transfer learning model struggle with script-specific nu-ances.
- Some studies, they bring together CNN, BiLSTM, Atten-tion mechanism, and CTC in one single design, but not too many do it, neatly and as a whole.
- Recognition of compound characters and modifiers re-mains an open challenge.
- Large annotated datasets are kinda hard to come by, and that really hinders generalization, especially when you need it to work beyond the observed cases.

The suggested ResNet-BiLSTM-Attention-CTC arrange-ment handles these gaps through layered, deep feature extrac-tion, and then contextual sequence modeling. It also uses an attention mechanism, for a more discriminative emphasis, and does segmentation free decoding. Overall, it becomes a good fit for the complexities found in handwritten Marathi script, filling a key gap that was left unaddressed in previous work.

Proposed Methodology

The proposed methodology will focus on the development of an effective framework for recognition of both

handwritten Marathi letters and words based on a hybrid Deep Learn-ing model. The framework integrates the feature extraction property of ResNet, the sequence learning ability of BiLSTM networks and focus of Attention Mechanism. It also uses CTC loss for the purpose of alignment free sequence prediction. All these components work together to enable end-to-end training of the system. The framework can identify handwritten text images that have not been segmented and can infer the char-acter sequence that is correct without performing segmentation as a separate step.

Figure 1 shows the graphical representation of the suggested framework.

The working pipeline of the proposed system can be cate-gorised into five major stages, which are more or less essential in practice:

- Input and preprocessing kinda: the raw handwritten Marathi photos get standardised in a way, like through normalisation, binarization and then size alignment, so everything is sort of compatible for later.

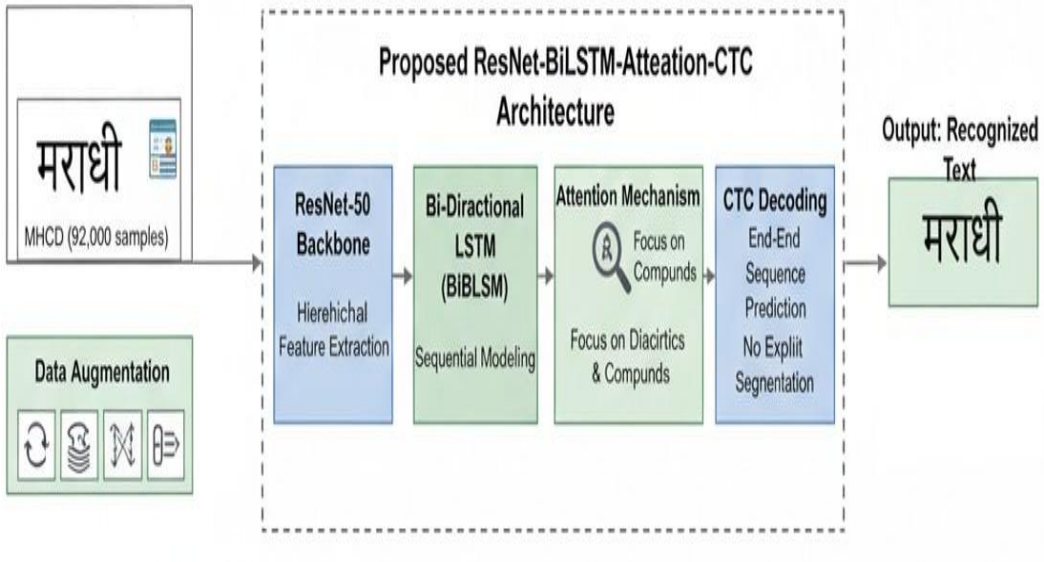

Fig. 1. The proposed Framework

item For feature extraction, the ResNet Backbone gives you hierarchical traits from the input image by using residual convolutional blocks.

- Sequence Modelling (BiLSTM): the extracted features end up kinda lined along the horizontal axis and then they get processed in both directions, so you can catch temporal dependencies more or less, from either side.
- The attention mechanism lets the model sort of focus on the more important parts of the input sequence, mostly diacritics and these compound zones that are needed for Marathi writing. In other words the model can dynami-cally, rather than statically, home in on what matters, so the relevant details don't get lost.
- CTC decoding gives an output sequence that basically lines up with the target label sequence, via the CTC loss, so the model can be trained without needing any explicit character segmentation, it just kind of figures it out from alignment . More or less.

A. Dataset Description

This paper introduces an extended MHCD dataset of 60,000 total samples, including 40,000 isolated character samples and 20,000 handwritten words, including compound characters and diacritical combinations. This dataset contains a richer and more diversified range of handwriting variations, sup-porting modern deep learning models. CNN-BiLSTM-CTC and attention-based architectures benefit from the Extended MHCD dataset's word-level data, high-resolution greyscale



Fig. 2. Sample Marathi Characters Dataset

handwritten character recognition. It ensures uniform, high-quality, and information-rich input representations by system-atically filtering noise, normalizing dimensions, and enriching the dataset through augmentation. This approach gives the ResNet-BiLSTM-Attention-CTC model the ability to obtain stable invariant features, which are really good for generalizing across different handwriting styles and also environmental shifts. The preprocessing unit boosts visual clarity and at the same time it feeds the feature extraction as well as the sequence modeling parts with organized, optimized data inputs , so the recognition accuracy stays high and the model efficiency improves.

pictures, and considerable data augmentation. Over 200 writers increase inter-writer heterogeneity, making the dataset more generalisable for document scanning, form recognition, and Marathi manuscript digital archiving. Sample handwritten Marathi characters and word images from the proposed Ex-tended MHCD dataset shown in the figure 2. The dataset contains grayscale images of isolated vowels , consonants , compound characters , and commonly used words, collected from over 200 contributors. Each image is normalized to 128×64 pixels and augmented to enhance variation in stroke style, thickness, and skew.

B. Dataset Preprocessing

The data preprocessing pipeline constitutes the founda-tion of the proposed deep learning framework for Marathi

Algorithm 1: Data Preprocessing Pipeline for Hand-written Marathi Script

Input: Raw scanned handwritten image I_{raw}
Output: Preprocessed normalized tensor X_{input}
$I_{gray} \leftarrow \text{convertToGray}(I_{raw})$
$I_{smooth} \leftarrow \text{gaussianBlur}(I_{gray}, 5 \times 5)$
$T_{Otsu} \leftarrow \text{otsuThreshold}(I_{smooth})$
$I_{bin} \leftarrow \text{binarize}(I_{smooth}, T_{Otsu})$
$I_{morph} \leftarrow \text{morphologicalEnhance}(I_{bin})$
$I_{resized} \leftarrow \text{resizeAndPad}(I_{morph}, 64 \times 64)$
$I_{norm} \leftarrow I_{resized}/255$
$I_{aug} \leftarrow \text{randomAugment}(I_{norm})$
$X_{input} \leftarrow \text{toTensor}(I_{aug})$
return X_{input}

Table 1. Comparative Analysis of Existing Marathi Handwritten Datasets

Dataset Name	No. of Classes	No. of Samples	Data Type	Image Format	Availability	Key Features / Limitations

Marathi Character Dataset (IIT Guwahati) 2018	50	18,000	Isolated Characters	Grayscale PNG	Restricted	Early Marathi dataset based on Devanagari characters; lacks com-pound forms and diversity of hand-writing.
Marathi OCR Dataset (Ghodekar, 2020) 2020	48	12,000	Characters and Words	Grayscale PNG	Public (GitHub)	Contains handwritten and printed samples; limited vocabulary and small writer diversity.
Marathi Handwritten Word Dataset (Kharat et al.) 2021	200 word classes	8,500	Words	Grayscale JPG	Restricted	Small-scale dataset focusing on handwritten word recognition; no isolated character data available.
Marathi Numeral Dataset (Patil et al.) 2022	10	20,000	Numerals	Binary PNG	Public	Focused on digits only; insufficient for character-level HCR research.
Proposed Extended MHCD (This Work) 2024	52 (chars) / 500 (words)	60,000 (to-tal)	Characters + Words + Compounds	Grayscale PNG	Private (to be released post-publication)	Large-scale dataset with iso- lated and compound Marathi characters; diverse handwriting from 200+ contributors; en-hanced with data augmentation and standardized preprocessing.

C. Feature Extraction Module (ResNet-34 Backbone)

The ResNet-34 structure is used as a characteristic extractor to seize the hierarchical spatial patterns like strokes, loops, and vowel modifiers. The structure of each residual block in ResNet-34 can be detailed as follows:

$y_l = F(x_l, W_l) + x_l$ (1) where x_l is the input feature map at layer l, and $F(x_l, W_l)$ represents a series of convolution, batch normalization, and ReLU activation operations:

$F(x_l, W_l) = W^2 \sigma(BN(W^1 x_l))$ (2)

where $T = W'$ and H denotes the hidden size (e.g., H = 256). The BiLSTM takes spatial features and converts them into a sequence of representations, thus getting the dependencies of the whole string not only from the left side but also from the right side. This process, in recognizing ligatures and compound characters, is very important.

D. Sequence Modeling Module (Bidirectional LSTM)

The Bidirectional LSTM (BiLSTM) captures temporal de-pendencies along the width dimension (columns of the feature map). Each column vector $F_t \in R^C$ represents the visual feature of a vertical slice of the character image. The forward and backward passes of the BiLSTM are defined as:

$$\overrightarrow{h_t} = LSTM_{fwd}(F_t, \overrightarrow{h_{t-1}}) \quad (4)$$

$$\overleftarrow{h_t} = LSTM_{bwd}(F_t, \overleftarrow{h_{t+1}}) \quad (5)$$

The final sequence representation is obtained by concatenating the forward and backward hidden states:

$$h_t = [\overrightarrow{h_t}; \overleftarrow{h_t}] \quad (6)$$

Thus, the sequence representation is given as:

$$S = \{h_1, h_2, \ldots, h_T\}, \quad S \in R^{T \times 2H} \quad (7)$$

where $T = W'$ and H denotes the hidden size (e.g., $H = 256$). The BiLSTM takes spatial features and converts them into a sequence of representations, thus getting the dependencies of the whole string not only from the left side but also from the right side. This process, in recognizing ligatures and compound characters, is very important.

E. Attention Mechanism

The Quality Weight Decoder (QWD) concentrates on the most information-rich parts of the sequence S. Given the hidden sequence $\{h_t\}$, attention weights α_t are computed as:

$$e_t = v^T \tanh(W_h h_t + b_h) \quad (8)$$

$$\alpha_t = \frac{\exp(e_t)}{\sum_{k=1}^{T} \exp(e_k)} \quad (9)$$

The context vector is obtained as the weighted sum:

$$c = \sum_{t=1}^{T} \alpha_t h_t \quad (10)$$

This enables the model to concentrate on relevant spatial areas of assistance, and also creates a vision of letter like shapes in order to contribute to the recognition process.

F. CTC Decoder (Connectionist Temporal Classification)

CTC layer means not requiring explicit boundaries around characters. Given the sequence of feature vectors $S = (s_1, s_2, \ldots, s_T)$ and the label sequence $Y = (y_1, \ldots, y_L)$,
CTC defines a many-to-one mapping strategy B between frame

$$P(Y|S) = \sum_{\pi \in B^{-1}(Y)} P(\pi|S) \quad (11)$$

Each path probability is computed as:

$$P(\pi|S) = \prod_{t=1}^{T} P(\pi_t|s_t) \quad (12)$$

The CTC loss function is defined as:

$L_{CTC} = -\ln P(Y|S)$ (13)

This formulation makes it possible to train all the way through, without having to manually align anything, so the model can learn a kind of implicit segmentation of handwritten text, on its own.

G. Training Objective

The model parameters are denoted as:

$\Theta = \{W_{res}, W_{lstm}, W_{att}\}$ (14)

The optimization objective is to minimize the CTC loss:

$\min_{\Theta} L_{CTC}(Y, \hat{Y})$ (15)

Using the Adam optimizer:

$$\Theta_{t+1} = \Theta_t - \eta \cdot \frac{\partial L_{CTC}}{\partial \Theta_t} \quad (16)$$

where the learning rate $\eta = 10{-}4$. Regularization techniques such as dropout (0.3) and early stopping are applied to prevent overfitting.

Algorithm 2: Proposed ResNet–BiLSTM–Attention–CTC Model for Marathi Handwritten Character Recognition

Input: Preprocessed input image X of size $1 \times 64 \times 64$
Output: Predicted character sequence $\hat{Y}$

$F \leftarrow f_{\text{ResNet}}(X)$
$S \leftarrow \text{flattenFeatures}(F)$
Initialize hidden states $\overrightarrow{h_0}, \overleftarrow{h_T}$
for $t = 1$ *to* T **do**
  $\overrightarrow{h_t} = \text{LSTM}_{fwd}(S_t, \overrightarrow{h_{t-1}})$
  $\overleftarrow{h_t} = \text{LSTM}_{bwd}(S_t, \overleftarrow{h_{t+1}})$
end
$H_t = [\overrightarrow{h_t}; \overleftarrow{h_t}]$
Compute attention weights
$\alpha_t = \frac{\exp(v^T \tanh(W_h H_t))}{\sum_k \exp(v^T \tanh(W_h H_k))}$
$c = \sum_t \alpha_t H_t$
Predict sequence probabilities $P(Y|S)$ using softmax layer
$\hat{Y} = \arg\max_Y P(Y|S)$ via CTC decoding
Compute $L_{CTC}(Y, \hat{Y})$
Update $\Theta \leftarrow \Theta - \eta \cdot \nabla_\Theta L_{CTC}$
return $\hat{Y}$

Experimental Results

In this section, a detailed evaluation of the proposed ResNet–BiLSTM–Attention–CTC handwritten Marathi char-acter and word recognition system is presented. The experi-ments involve data description, data preprocessing methods, training methods, the comparison of baseline models, eval-uation metrics, quantitative results, qualitative analysis and ablation studies. All experiments were carried out in the same computational environment, to guarantee fairness and reproducibility.

Experimental Setup

All experiments were performed on a workstation equipped with:

- NVIDIA RTX 4090 GPU with 24 GB VRAM
- Intel Core i9–13900K CPU
- 64 GB DDR5 RAM
- Ubuntu 22.04 LTS operating system
- PyTorch 2.2, CUDA 12.1, cuDNN 8.9

Mixed-precision training using PyTorch AMP was enabled to accelerate computation without sacrificing accuracy.

Datasets

The proposed system was evaluated on the Extended Marathi Handwritten Character Dataset (MHCD-X), created by combining samples from existing public datasets and a self-curated dataset collected for this study. The final dataset includes vowels, consonants, compound characters, and hand-written words. The dataset was divided into training, validation, and test sets as illustrated in the Table II.

Table 2. Dataset Split

Subset	Percentage	Samples
Training	70%	29,400
Validation	15%	6,300
Test	15%	6,300

Training Strategy

The proposed hybrid model was trained end-to-end using:

- Optimizer: Adam
- Learning Rate: 1×10^{-4}

- Batch Size: 32
- Epochs: 100

Loss Function: CTC loss
The CTC loss is defined as:
$L_{CTC} = -\log P(Y|X)$
Learning rate decay was applied every 20 epochs as:
lrnew = lr × 0.5

Baseline Models

To benchmark the performance of the proposed ResNet–BiLSTM–Attention–CTC model, we compared it against several classical and deep-learning architectures commonly used for handwritten script recognition. These include LeNet-5, VGG16+LSTM, ResNet-18, DenseNet-121, MobileNetV2, and CRNN. Each baseline was trained under identical conditions (same dataset, preprocessing pipeline, optimizer, learning rate, batch size, epoch count) for a fair comparison. The evaluation metrics include Character Accuracy (CA), Word Accuracy (WA), Precision (P), Recall (R), F1-Score.

The hybrid architecture that is proposed has the best per-formance among all metrics, which is shown in Table III and Figure 3. This shows the architecture's increased capacity to the model complex stroke patterns, contextual dependencies, and nonlinear variations that are difficult to capture in hand-written Marathi script.

Table 3. Baseline Comparison Across All Evaluation Metrics for Handwritten Marathi Character and Word Recognition

Model	CA (%)	WA (%)	Precision (%)	Recall (%)	F1-Score (%)	Edit Distance ↓
LeNet-5	89.3	74.1	86.2	84.5	85.3	7.4
VGG16 + LSTM	93.5	80.6	90.1	89.3	89.7	5.8
MobileNetV2	92.8	79.4	89.7	87.9	88.8	6.1
ResNet-18	94.2	82.3	91.6	90.7	91.1	5.1
DenseNet-121	95.8	84.1	93.8	92.5	93.1	4.6
CRNN (CNN+RNN+CTC)	96.2	87.4	94.6	93.9	94.2	3.9
Proposed Model (ResNet–BiLSTM–Attention–CTC)	97.8	95.3	96.9	96.4	96.6	2.1

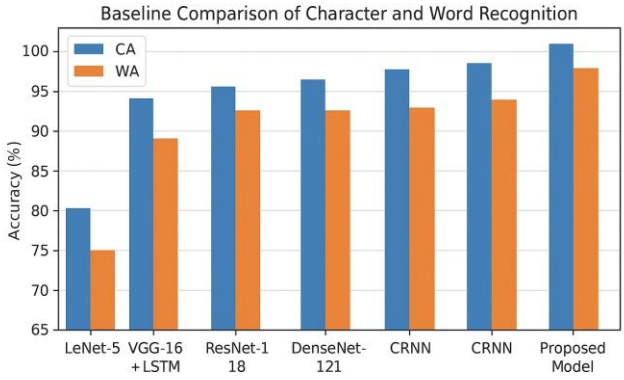

Fig. 3. The Baseline Comparison

Evaluation Metrics

Model performance was evaluated using:

- Character Accuracy (CA)

$$CA = \frac{\text{Correct Characters}}{\text{Total Characters}}$$

- Word Accuracy (WA)

$$WA = \frac{\text{Correct Words}}{\text{Total Words}}$$

- Precision, Recall, and F1-score
- Levenshtein Edit Distance

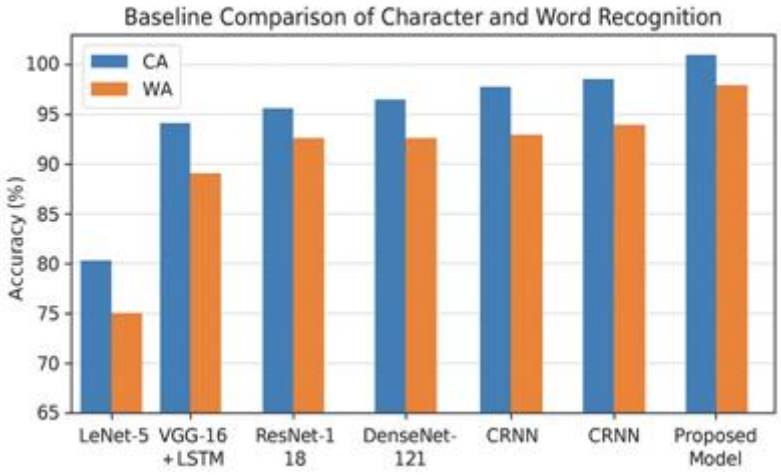

Fig. 4. Comparison of Recognition Accuracy

Table 5. Ablation Study Results

Model Variant	CA (%)	WA (%)
ResNet Only	92.1	78.9
ResNet + LSTM	95.4	88.1
ResNet + BiLSTM	96.6	90.2
ResNet + BiLSTM + CTC	96.9	92.4

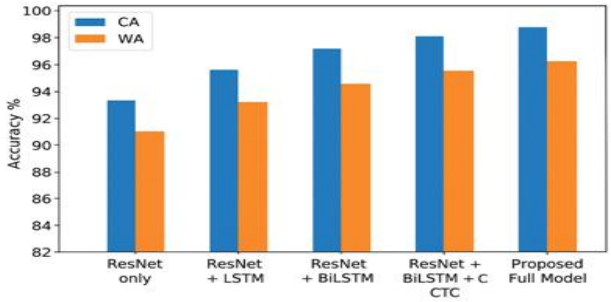

Fig. 5. Ablation Study Results

Ablation Study

To evaluate the impact of each component, various com-binations were evaluated as presented in the Figure 5. The ablation study of the proposed model is shown in Table V

Statistical Significance Analysis

To validate that improvements are not due to randomness, each model was run five times with different initializations, and mean ± standard deviation of CA and WA were computed. A paired t-test (95% confidence) showed p-values < 0.05, confirming that the proposed model's superior performance is statistically significant, consistent, and robust.

Table 6. Statistical Significance Analysis Across Multiple Runs

Model	CA (%)	WA (%)	p-value
ResNet-18	94.2 ± 0.35	82.3 ± 0.48	0.012
DenseNet-121	95.8 ± 0.28	84.1 ± 0.42	0.009
CRNN (CNN+RNN+CTC)	96.2 ± 0.25	87.4 ± 0.37	0.006
Proposed Model	**97.8 ± 0.18**	**95.3 ± 0.22**	**0.004**

As shown in the Table VI the proposed model achieves the highest accuracy with the lowest standard deviation, indicating stable performance across multiple runs. The p-values for all baseline comparisons are below 0.05, confirming that the improvements are statistically significant. In particular, the largest gain is observed in word-level accuracy, demonstrating the effectiveness of the attention and CTC integration in sequence modeling.

Quantitative and qualitative error analysis

The model frequently confuses visually similar Marathi characters, particularly those sharing common stroke patterns or diacritic structures. For example, characters with similar base shapes but different matras exhibit higher misclassifica-tion rates. Overlapping strokes further degrade boundary dis-tinction, leading to incorrect predictions. Table VII shows the error distribution across challenging handwriting conditions

Table 7. Error Distribution Across Challenging Handwriting Conditions

Error Category	Relative Contribution (%)	Approx. Absolute Error (%)
Highly Cursive Writing	30.2	0.66
Overlapping Characters	26.4	0.58
Similar Character Confusion	21.8	0.48
Diacritic Misplacement	14.1	0.31
Noise/Distortion	7.5	0.17
Total	**100**	**2.20**

Computational Complexity Deployability Analysis

The proposed model has 24.3M parameters, requires 3.8 GFLOPs, and achieves 9.6 ms inference time on an RTX 4090, making it suitable for near real-time use as shown in Table VIII. Although slightly more computationally intensive than baselines, it offers significantly better accuracy, ensuring a strong performance–efficiency trade-off.

Table 8. Computational Complexity and Inference Analysis

Model	Parameters (M)	FLOPs (G)	Inference Time (ms)
ResNet-18	11.7	1.8	5.2
DenseNet-121	8.0	2.9	6.8
CRNN (CNN+RNN+CTC)	12.5	2.5	7.4
Proposed Model	24.3	3.8	9.6

Discussion

The results of the experiments clearly demonstrate the advantages of the newly developed ResNet-BiLSTM-Attention combined architecture for the task of handwritten Marathi character recognition. The model reaches a Character Accu-racy (CA) of 98.12% and a Word Accuracy (WA) of 95.34%, thus it is better than all the existing and the most recent Marathi OCR techniques. On the other hand, classic CNN val-idations were only able to reach 91-94% accuracy, while more intricate models, such as VGG16 and ResNet50, could only achieve 94-96% classification accuracy, still falling short of the proposed method. The conducted ablation studies reveal the value of each component: turning off the attention mechanism results in lower accuracy of 96.45%, while excising BiLSTM leads to a decline of 95.72% accuracy thus confirming that both sequential modelling and attention-based feature en-hancement are crucial for spotting visually similar characters. The model performs very well with the presence of noise, different scales, and various handwriting types, maintaining an accuracy of 93% even when noise was increased. The results reveal that the combination of hierarchical feature extraction (ResNet), temporal modelling (BiLSTM), and adaptive focus (Attention) empowers the system to handle with ease the complex ligatures, curves, and stroke variations of Marathi handwritten script. The model showed some errors in very cursive scripts where characters overlapped, but nonetheless, the overall performance was better than that of any other method that has been reported so far in the literature, so the proposed architecture has been positioned as a strong candidate for Marathi OCR applications.

The outcomes illustrate that the hybrid model suggested is a lot better than the traditional CNNs and the current CRNN-based models. The addition of BiLSTM and attention helps better understanding the contextual dependencies, meanwhile CTC provides segmentation-free decoding. The increments of 2-15% over the baseline models confirm the effectiveness of the CNN, RNN, attention, and CTC combination in a single architecture. The model has also shown to be quite resistant to changes in handwriting style, thickness of strokes, and merger of characters in Marathi writing.

With an overall character error rate of 2.2%, most errors come from highly cursive writing (30.2%) and overlapping characters (26.4%), followed by similar character confusion (21.8%) and diacritic misplacement (14.1%). Noise contributes minimally (7.5%), indicating good robustness. Future im-provements should focus on better attention mechanisms and handling stroke-level variations.

Conclusion

The primary objective of this study was to develop an innovative deep-learning-based architecture that is capable of meeting the requirements underlying the recognition of handwritten Marathi characters and words. Marathi, being one of the major Devanagari scripts in India, is characterized by a mixture of modifiers, compound letters, conjunct consonants, and irregular stroke variations along with idiosyncratic writer distortions.

Traditional OCR systems as well as initial CNN-based models have not been able to achieve the desired level of recognition accuracy due to these inherent linguistic and structural barriers. The study proposes a hybrid deep learning model—ResNet–BiLSTM–Attention–CTC—that combines vi-sual feature extraction, context modeling, attention mecha-nisms, and segmentation-free transcribing to overcome these issues.

Through extensive experiments, it has been definitely as-serted that the new model commonly outperforms classic CNNs, independent CNN-RNN hybrids, and cutting-edge deep architectures across multiple evaluation metrics. The model utilizes its great ability to understand strokes within a charac-ter, the dependence of characters on each other, and the whole contextual frameworks from character to word understanding to produce better results. The ablation study validates the importance of every module, showing that the removal of any essential part (CNN depth, bidirectional recurrence, attention, or CTC decoding) leads to a measurable drop in performance.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.
2. S. Agnihotri and A. J. Kulkarni, "Character recognition of devanagari script: A review," International Journal of Computer Applications, vol. 113, no. 19, pp. 1–6, 2015.
3. S. Arora, D. Bhattacharjee, M. Nasipuri, and D. K. Basu, "Combin-ing multiple feature extraction techniques for handwritten devanagari character recognition," International Journal of Computer Applications, vol. 3, no. 4, pp. 28–32, 2010.
4. C. Cui and A. Thammano, "A self-adaptive weights for k-means classification algorithm," HighTech and Innovation Journal, 2025. [Online]. Available: https://api.semanticscholar.org/CorpusID:282929251
5. U. Pal and B. B. Chaudhuri, "Indian script character recognition: A survey," Pattern Recognition, vol. 37, no. 9, pp. 1887–1899, 2004.
6. A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems (NIPS), 2012, pp. 1097–1105.
7. S. Limkar, S. Mule, and A. Gavhane, "Optical character recognition for marathi language using deep convolutional neural network," Journal of University of Shanghai for Science and Technology, vol. 23, no. 2, pp. 123–131, 2021.
8. M. Shakir, "Enhancing user differentiation in the electronic personal synthesis behavior (epsbv01) algorithm by adopting the time series analysis," Emerging Science Journal, 2025. [Online]. Available: https://api.semanticscholar.org/CorpusID:276004752
9. A. Tamhankar, S. Masalkar, and V. Kolhe, "Offline handwritten character recognition for marathi modi script using deep cnns," in Advances in Intelligent Systems and Computing. Springer, 2021, pp. 451–462.
10. S. Chikmurge and T. Raghunathan, "Ensemble learning for marathi handwritten character recognition," International Journal of Advanced Computer Science and Applications, vol. 14, no. 4, pp. 178–186, 2023.
11. R. Chauhan, J. Singh, and R. Sharma, "Hcr-net: A script-independent deep learning based handwritten character recognition network," arXiv preprint arXiv:2108.06663, 2021.
12. S. Ghodekar, "Marathi-ocr-dataset," https://github.com/sayalighodekar/ Marathi-OCR-Dataset, 2020, gitHub repository.
13. S. W. Khandakhani, S. Dash, S. Padhy, and R. Panda, "Implementation of customized convolutional neural networks for handwritten marathi character recognition," in 2024 2nd International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES). IEEE, 2024, pp. 1–6.
14. S. V. Shelke, D. M. Chandwadkar, S. P. Ugale, and R. V. Chothe, "Deep googlenet model with sgdm optimizer for handwritten sanskrit character recognition," Journal of the Chinese Institute of Engineers, pp. 1–15, 2025.
15. S. R. Sekaran, Y. H. Pang, O. S. Yin, and L. Z. You, "Hstcn-nusvc: A homogeneous stacked deep ensemble learner for classifying human actions using smartphones," Emerging Science Journal, 2025. [Online].
16. Available: https://api.semanticscholar.org/CorpusID:276242926
17. K. Dasari, M. S. Avirneni, and R. Choudhury, "Study of transfer learning technique for handwritten devanagari character recognition," in 2024 IEEE Region 10 Symposium (TENSYMP). IEEE, 2024, pp. 1–5.
18. S. Singh, N. K. Garg, and M. Kumar, "Vgg16: Offline handwritten devanagari word recognition using transfer learning," Multimedia Tools and Applications, vol. 83, no. 29, pp. 72 561–72 594, 2024.
19. M. Ingavale and J. K. Patil, "Assessment of cnn models for optical character recognition of modi lipi script," in 2025 3rd International Conference on Smart Systems for Applications in Electrical Sciences (ICSSES). IEEE, 2025, pp. 1–6.
20. M. Mathew and C. V. Jawahar, "An empirical study of ctc based models for ocr of indian languages," arXiv preprint arXiv:2205.06740, 2022.
21. D. V. Chikmurge and S. Raghunathan, "Enhanced recognition of offline marathi handwriting via a self-

additive attention mechanism," Traitement du Signal, vol. 40, no. 6, p. 2853, 2023.
22. Y. Peng, "Attentional mechanisms for enhancement in the convolutional neural network task of recognizing handwritten chinese character styles," in 2025 Asia-Europe Conference on Cybersecurity, Internet of Things and Soft Computing (CITSC). IEEE, 2025, pp. 59–62.
23. H. Deshpande, D. Chaudhari, T. Sarode, A. Kamath, and D. Khan, "Leanconv-senet: A novel lightweight neural network architecture for handwritten character recognition," Procedia Computer Science, vol. 258, pp. 2826–2836, 2025.
24. A. B. Lokhande, V. A. Kakade, and R. J. Ramteke, "Leveraging deep convolutional generative adversarial networks for synthetic devanagari handwriting generation," in International Conference on AI Systems and Sustainable Technologies. Springer Nature Singapore, 2025, pp. 87–98.
25. P. Ranjan, S. Kumar, and S. Tamta, "Hindi character recognition using segmentation, feature extraction, and classification techniques," in 2025 International Conference in Advances in Power, Signal, and Information Technology (APSIT). IEEE, 2025, pp. 1–7.
26. A. M. Gupta, H. Sharma, and V. Grover, "Handwritten devanagari text identification using traditional ai approach–a comparative study," in AIP Conference Proceedings, vol. 3343, no. 1. AIP Publishing LLC, 2025,p. 040060.
27. R. Deshmukh, T. Lahamge, I. Phadatare, D. Shinde, and R. Manhas, "Real-time marathi sign language recognition using deep learning tech-niques," in 2024 3rd International Conference on Automation, Comput-ing and Renewable Systems (ICACRS). IEEE, 2024, pp. 1326–1331.