# Bangla OCR for Hand-written Texts


Abdullah Rahat
*dept. CSE*
*Brac University*
Dhaka, Bangladesh
abdullah.rahat@g.bracu.ac.bd

Jarin Akter Mou
*dept. of CSE*
*Brac University*
Dhaka, Bangladesh
jarin.akter.mou@g.bracu.ac.bd

Hasim Reza
*dept. of CSE*
*Brac University*
Dhaka, Bangladesh
hasim.reza@g.bracu.ac.bd

Zaki Habib
*dept. of CSE*
*Brac University*
Dhaka, Bangladesh
zaki.habib@g.bracu.ac.bd

Farah Binta Haque
*dept. of CSE*
*Brac University*
Dhaka, Bangladesh
farah.binta.haque@g.bracu.ac.bd

Sania Azhmee Bhuiyan
*dept. of CSE*
*Brac University*
Dhaka, Bangladesh
sania.azhmee.bhuiyan@g.bracu.ac.bd

Adib Muhammad Amit
*dept. of CSE*
*Brac University*
Dhaka, Bangladesh
adib.muhammad.amit@g.bracu.ac.bd

Annajiat Alim Rasel
*dept. of CSE*
*Brac University*
Dhaka, Bangladesh
annajiat@bracu.ac.bd



***Abstract*—Recent years have seen tremendous breakthroughs in optical character recognition (OCR) technology, which is essential for converting non-editable documents into formats that can be searched and read by machines. The development of OCR methods, including character segmentation, picture preprocessing, and recognition algorithms, is reviewed in this work. Convolutional neural networks (CNNs), one of the deep learning models, have helped to drive OCR accuracy to previously unheard-of heights, making it possible to extract text from a variety of sources with ease. The abstract also examines the difficulties of using Bangla OCR for handwritten writings. Information accessibility and processing efficiency have advanced due to the use of OCR in applications ranging from document digitization to automated data extraction.**



***Index Terms*—OCR, Bangla, handwritten, CNN**


## I. INTRODUCTION

A game-changing technology called optical character recognition (OCR) is used to turn handwritten or printed text into machine-encoded text. It acts as a link between the digital and analog realms, enabling textual information to be extracted, recognized, and interpreted from a variety of sources, including scanned documents, photos, and recorded video. OCR technology is essential for streamlining information retrieval, improving document searchability, and automating data entering procedures.

In order to accommodate a variety of language and cultural situations, OCR has recently increased the scope of its capabilities to include the recognition of handwritten text. In particular, OCR for handwritten Bangla text tackles the particular difficulties brought about by the complex script and variety of writing forms found in Bengali. Technological developments in OCR for handwritten Bangla text are a crucial part of the linguistic variety in the field and are intended to aid in the digitization of handwritten documents, provide accessibility, and support the preservation of cultural heritage. This customized method highlights how OCR technology can adjust to linguistic subtleties, making it a useful tool for maximizing the potential of handwritten writings in Bangla.

In the field of Optical Character Recognition (OCR), the use of Convolutional Neural Networks (CNNs) to Bengali Handwritten Text Recognition is a remarkable development. Handwritten text identification poses a challenge, but CNNs, a family of deep learning models, have proven to be highly effective in image-related tasks. CNNs are quite good at picking up on the complex elements and patterns present in handwritten Bengali text. Characters can be correctly identified and classified thanks to the network's capacity to autonomously learn hierarchical representations from input photos. In order for the model to generalize and adjust to the variances found in various writing styles, it is trained on a broad dataset of handwritten samples using CNNs for Bengali Handwritten Text Recognition.

## II. LITERATURE REVIEW

Yi et al[1] explored advancements in Handwritten Chinese Text Recognition, introducing the Separable Multidimensional Recurrent Neural Network (SMDLSTM-RNN). The study addresses challenges inherent in handwritten Chinese text, such as diverse writing styles, character segmentation difficulties, and an extensive character set within an unconstrained language domain. SMDLSTM-RNN is proposed as a solution to efficiently capture contextual information in multiple orientations, reducing computation time and enhancing recognition accuracy. The paper introduces a novel 7-layer architecture incorporating SMDLSTM-RNN and Convolutional Neural Networks (CNN) modules, outperforming a Bidirectional Long Short-Term Memory (BLSTM)-based model on datasets with varied character class counts. Experimental results on the ICDAR-2013 competition dataset demonstrate superior performance compared to earlier LSTM-based techniques, positioning the suggested method as competitive with cutting-edge systems. The literature suggests addressing the context overfitting issue for future system enhancements. This review synthesizes key findings, methodologies, and proposed improvements in Handwritten Chinese Text Recognition using SMDLSTM-RNN, contributing valuable insights to the existing body of literature in the field.

The literature on Bangla Handwritten Character Recognition using Convolutional Neural Network (CNN) with Data Augmentation introduces an innovative approach to recognizing individual Bangla characters studied by Pal et al[2]. Focused on the BanglaLekha-Isolated dataset, the research employs a CNN architecture and highlights the significance of recognizing Bangla characters in handwritten form, underscoring the scarcity of research in this domain compared to English handwriting recognition. The proposed CNN model demonstrates substantial accuracy, particularly after the application of data augmentation techniques, emphasizing the effectiveness of CNN for Bangla handwritten character recognition. The methodology involves training the CNN model with the base dataset and expanding it using data augmentation, proving to be a successful strategy for enhancing model accuracy. The paper conducts a comprehensive analysis of results, including precision, recall, and F1 score metrics. Additionally, the model's versatility is showcased by its application to other datasets such as MNIST, Ekush Bangla, and CMATERdb 3.1.2, revealing competitive accuracy across diverse datasets. All things considered, this body of work offers a promising answer to the problem of handwritten character recognition in Bangla, opening the door for more developments in this area.

A thorough Optical Character Recognition (OCR) system designed specifically for single-font Bangla documents is presented in a different work by Mir et al[3]. The unique approach combines template and feature matching for character recognition, eliminating the need for preprocessing steps and contributing to the system's efficiency. The paper emphasizes the significance of this system for OCR applications in the Bangla script, with potential benefits for text entry automation and assistance to visually impaired individuals. The research employs a feature-based tree classifier for recognizing simple characters, achieving an impressive accuracy of approximately 96%. The overall recognition accuracy of the OCR system is reported as 96.55%, highlighting its effectiveness in accurately recognizing both basic and compound characters in the Bangla language. This work contributes valuable insights into OCR advancements for the Bangla script, showcasing a promising tool for various applications in document processing and accessibility.

Abu et al[4] researched challenges associated with non-ideal capture conditions of printed Bangla text images in the context of Optical Character Recognition (OCR). The proposed algorithm involves various preprocessing steps, including skew correction, image cropping, dewarping, noise reduction, binarization, and segmentation phases for lines, words, and characters. The methodology demonstrates effectiveness even in non-ideal scenarios, such as images with multiple font sizes. Results indicate high accuracy in line (100%), word (99.10%), and character (94.32%) segmentation. The study highlights the importance of each algorithmic step and emphasizes the significance of handling non-ideal cases for a robust OCR system. Challenges remain in handling character overlap and improving matra line detection for further advancements in Bangla OCR.

## III. Dataset

A powerful benchmark image library, the AIBangla dataset is evidence of the enormous diversity and teamwork that come from Bangladesh's dynamic geography. Thanks to the efforts of more than 2,000 different writers from different institutions around the country, this collection contains an astounding 80,403 handwritten graphics that have been carefully created to represent the 50 basic Bangla characters.


**Figure 1:** *Sample image of Bangla written alphabets.*

Beyond just individual characters, though, the dataset contains a remarkable collection of 249,911 handwritten pictures that beautifully display 171 different Bangla compound characters, each distinguished by its own pattern forms.

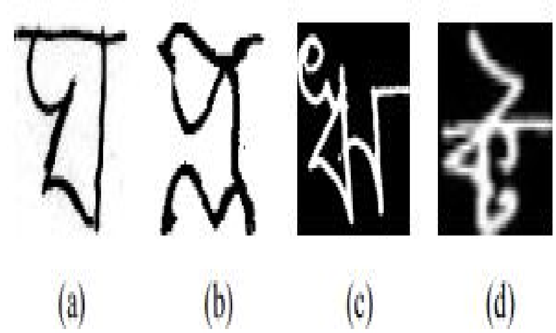


**Figure 2:** *Sample image of isolated characters.*

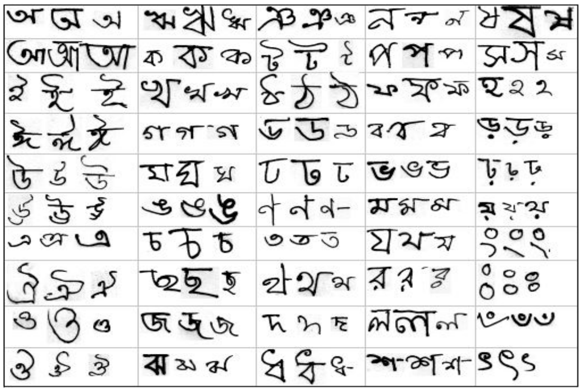

**Figure 2:** *Sample image of same character's multiple way of writing.*

## IV. Methodology

### A. Data pre-process

By Loading the dataset from a CSV file, then extracting the labels for supervised learning. For reshaping the image to appropriate dimensions and performing one hot encoding on the labels. For dataset splitting we can use 80:20 split for the dataset,by using this model 80% is used for training, while the final 20% is used for validation and testing. Additionally, the dataset was split in half by 20%, yielding an 8:1:1 split between the training, testing, and validation sets. The dataset we got has a shape of (5000,4097) indicating 5000 images with 4096 features and the last column is representing the target variable. Understanding the dataset by inspecting the first few rows by gaining insights into the structure. Utilized a countplot for visual representation. Studying the class distribution to identify potential imbalances or patterns in the dataset.The architecture that has been called CNN, We have constructed a convolutional Neural Network (CNN) using Keras. Implemented three Conv2D layers followed by Maxpooling2D layers, a flatten layer and also for the two dense layers. Incorporated flattening and dense layers for classification within the CNN architecture. By utilizing ReLU activation for convolutional layers to introduce non-linearity and applying softmax activation for the output layer for multi-class classification. Convergence can be improved by implementing a learning rate scheduler, which dynamically adjusts the learning rate during training. A rate scheduler can be used to dynamically alter the learning rate during the training. For finding best weights and monitoring the validation loss. The model will give you a summary of the CNN architecture for providing insights into layer configurations and parameter counts. For monitoring the training process by visualizing metrics such as accuracy and loss over epochs offering insights into model performance. Showcased the impact of learning rate scheduling and early stopping on the model's training behavior,demonstrating their roles in optimizing training outcomes.

### B. Model Selection

Imported the information from a CSV file and extracted the labels for supervised learning. Images were resized to the proper proportions and labels were encoded one-hot. 20% of the dataset was used for testing and validation, while the remaining 80% was used for training. After dividing the remaining 20% in half more, the training, testing, and validation sets had a ratio of 8:1:1.The target variable is shown in the last column of the dataset, which has 5000 images with 4096 attributes shaped like (5000, 4097).studied the dataset's initial rows to get a rudimentary idea of its organization.Using a countplot, the distribution of the different classes in the dataset was shown graphically.looked at the distribution of the classes to find any patterns or imbalances in the sample.created a Convolution imported the information from a CSV file and extracted the labels for supervised learning. Labels were encoded one-hot and images were scaled to the appropriate dimensions. About 80% of the dataset was used for training, and the last 20% was used for testing and validation purposes.

After dividing the remaining 20% in half more, the training, testing, and validation sets had a ratio of 8:1:1.The target variable is shown in the last column of the dataset, which has 5000 images with 4096 attributes shaped like (5000, 4097).studied the dataset's initial rows to get a rudimentary idea of its organization.

Using a countplot, the distribution of the different classes in the dataset was shown graphically.searched for any patterns or imbalances in the class distribution by looking at the sample.

### C. Training

We made sure the images were the right size to be loaded into the CNN architecture by resizing them.After extracting the labels from the dataset, the classes were encoded in one go.constructed a CNN with three MaxPooling2D layers, three Conv2D levels, two Dense layers, and a Flatten layer. To achieve non-linearity, we employed Rectified Linear Unit (ReLU) activation after convolutional layers. The output layer used Softmax activation for multi-class categorization.employed the Adam optimizer due to its capacity to modify learning rate and handle sparse gradients in an efficient manner.opted to train the model to classify objects into many categories using categorical cross entropy as the loss function.a learning rate scheduler was put in place to modify the learning rate during instruction in response to trainee advancement.

We Kept an eye on how the learning rate scheduler affected the model's convergence and noted any possible gains in training stability and speed. Early stopping was used while the validation loss was being monitored. ceased training in order to avoid overfitting and restore the optimal weights when the validation loss did not decrease.

Then built the model using accuracy and categorical cross entropy loss as the assessment metrics. Several epochs, or one full run of the whole training dataset, were used to train the model. a set batch size was used to achieve effective gradient modifications.

### D. Illustration

The training process was made easier to see by plotting statistics such as training and validation accuracy and loss over epochs.

## V. Result and Analysis

Our implemented model obtained an accuracy of 85.56% and a test loss of 0.5496 when evaluated on the test set. With a test accuracy of 85.56%, the model appears to have done a good job of correctly predicting values based on data that had not been seen before. The test loss, which stands at 0.5496, represents the degree of error that exists between the actual and predicted values; smaller values are typically preferred. Overall performance is indicated by the accuracy and loss measures, which indicate that the model has successfully generalized to the test set and can produce reliable predictions on newly acquired, unobserved data. To investigate possible areas for improvement or to modify the model to meet particular application needs, more research or tweaking may be done.

Our Implemented model offers an extensive analysis of its effectiveness based on multiple parameters. With an overall precision of 86.13%, precision, which ranges from 65% to 97%, illustrates the model's accuracy in identifying positive events. With an overall recall of 85.56%, the recall, which ranges from 69% to 99%, indicates the model's sensitivity to true positive events. An overall F1 score of 85.60% is obtained by taking the harmonic mean of precision and recall, or F1-score, which reflects a balanced trade-off between the two and ranges from 70% to 95%. With an accuracy of 85.56%, the model makes a large number of accurate predictions.

TABLE I
Result

| **Model Name** | **Result** | | |
|---|---|---|---|
| | ***Accuracy*** | ***Precision*** | ***F1-Score*** |
| CNN | ***85.56%*** | ***86.13%*** | ***85.60%*** |

## VI. Conclusion

It is remarkable that our trained model was able to acquire a decent score, especially given the inherent complexity presented by a variety of writing styles, compound characters, and segmentation difficulties. The CNN-based approach has proven to be quite successful in traversing these complications, demonstrating how well it can handle the inherent complexities in the data. This achievement confirms the model's dependability in real-world applications by highlighting its capacity to handle the subtleties of diverse writing styles and characters.

There is a lot of space for future development and optimization. We can improve our outcomes by using an improved model, especially with regard to processing speed and accuracy. The ongoing development of model architectures and machine learning approaches offers us the chance to further hone our methodology. Subsequent pursuits could entail investigating sophisticated models or refining current ones to achieve more precision while maximizing computational effectiveness. Furthermore, adding additional words or phrases to our training set may enhance the robustness of our recognition technique and help our model perform well in a wider variety of settings. We believe that further development and refinement of our model will lead to even better outcomes in terms of efficiency and accuracy.

## References


[1] Y.-C. Wu, F. Yin, Z. Chen and C.-L. Liu, "Handwritten Chinese Text Recognition Using Separable Multi-Dimensional Recurrent Neural Network," 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, Japan, 2017, pp. 79-84, doi: 10.1109/ICDAR.2017.22.

[2] R. R. Chowdhury, M. S. Hossain, R. ul Islam, K. Andersson and S. Hossain, "Bangla Handwritten Character Recognition using Convolutional Neural Network with Data Augmentation," 2019 Joint 8th International Conference on Informatics, Electronics  Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision  Pattern Recognition (icIVPR), Spokane, WA, USA, 2019, pp. 318-323, doi: 10.1109/ICIEV.2019.8858545.

[3] U. Pal and B. B. Chaudhuri, "OCR in Bangla: an Indo-Bangladeshi language," Proceedings of the 12th IAPR International Conference on Pattern Recognition, Vol. 3 - Conference C: Signal Processing (Cat.No.94CH3440-5), Jerusalem, Israel, 1994, pp. 269-273 vol.2, doi: 10.1109/ICPR.1994.576917.

[4] Abir, A. S. M., Rahman, S., Ellin, S., Farzana, M., Manik, M. H.,  Rahman, C. R. (2020). Confronting the Constraints for Optical Character Segmentation from Printed Bangla Text Image. In Proceedings of the 7th International Conference on Networking, Systems and Security (NSysS '20), December 2020 (pp. 36–44). https://doi.org/10.1145/3428363.3428367

[5] Dipu, N. M., Shohan, S. A., Salam, K. M. A., "Optical Character Recognition for Bengali Language Using Neural Networks: A Comparative Study," presented at the 2021 24th International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 18-20 December 2021. Published by IEEE, DOI: 10.1109/ICCIT54785.2021.9689864.