Received: 05 May,2024 | Revised: 15 September,2024 | Accepted: 03 October,2024

ORIGINAL RESEARCH

Enhancing Bangla Handwritten Optical Character Recognition (OCR) For Joint Letters Using Deep Learning Approach

Md. Ismiel Hossen Abir[1], Taspia Salam[2]

[1,2] Department of Computer Science & Engineering, International Standard University, Dhaka, Bangladesh

Abstract

The Bangla language consists of 285 conjunct letters. Identifying handwritten Bangla combined letters is a challenging task. Nevertheless, employing a deep learning approach simplifies this work. In this study on handwritten optical character recognition, the 15 most often utilized combined letters in the Bangla language are gathered. A dataset consisting of 3000 handwritten joint letter images for training and 300 plus images for testing are taken for investigation. Initially, image data are preprocessed, followed by conducting visual analysis to comprehend the content of the photos. Subsequently, the Keras Sequential API as a Convolutional Neural Network (CNN) are employed. Nevertheless, it is noted that the accuracy was around 85 percent. Consequently, this research used various models in order to enhance the accuracy of the proposed model such as: Xception, EfficientDet, VGGNet, and NASNet model. The Xception model had an accuracy rate of approximately 96 percent. This result represents a significant progress in the identification of handwritten Bangla joint letters, highlighting the efficiency of deep learning techniques in improving optical character recognition for the Bangla language.

Keywords: Bangla language, Handwritten Optical Character Recognition (OCR), Joint letters, Deep learning, Convolutional Neural Networks (CNNs)

1 | Introduction

The Bangla language poses a distinctive challenge in optical character recognition (OCR) due to its extensive collection of conjunct letters. The Bangla language consists of 285 conjunction letters. Recognizing all 285 characters is a challenging feat. Nevertheless, the progress in deep learning methodologies has resulted in the development of numerous novel models that can efficiently handle this level of complexity [1-2].

This study employs deep learning techniques to recognize handwritten Bangla joined letters. This offers a thorough investigation focused on improving Bangla Handwritten Optical Character Recognition (OCR) using deep learning techniques.

The main goal is to reduce the disparity between the intricacy of Bangla joined letters and the effectiveness of OCR technology. In order to accomplish this task, a dataset consisting of the 15 most often utilized compound letters in the Bangla language is compiled. This dataset is essential for our experimental and validation procedures. In order to accomplish this, a dataset consisting of the 15 most often used compound letters in Bangla language is compiled. This dataset is fundamental for our experimental and validation procedures.

The method commences with preparing image data. Afterwards, a visual analysis to gain a more profound grasp of the features of handwritten joint letters images is employed. The accuracy of the Conventional Neural Networks (CNNs) model is 85 percent. Therefore, a pre-existing model is used to assess the level of accuracy.

The Keras Sequential API is used to construct the Convolutional Neural Networks (CNNs). The accuracy of the Conventional Neural Networks (CNNs) model is 85 percent. Therefore, a pre-existing model is used to assess the level of accuracy. Consequently, various model including Xception, EfficientDet, VGGNet, and NASNet are employed. By conduction through experimentation and evaluation, it is determined that the Xception model yields the highest accuracy, achieving an outstanding outcome of roughly 96 percent. Reaching such exceptional precision represents a major achievement in the field of Bangla handwritten OCR, highlighting the effectiveness of deep learning techniques in enhancing character recognition for the Bangla language. This research demonstrates the capacity of deep learning to significantly transform OCR systems in the context of Bangla language [3-4].

2 | Related Work

Researchers have devoted significant attention to developing viable solutions in the field of handwritten character recognition. Previous researches utilized a range of technologies, including both standard image processing techniques and more advanced deep learning approaches. Initial efforts in Bangla handwritten identification mostly concentrated on heuristic-based approaches, which required analyzing the forms and intensity fluctuations of individual characters. Nevertheless, these methods frequently faced difficulties, especially when handling combine letters. Scientists have investigated Convolutional Neural Network (CNN) model specifically designed to handle the intricacies of recognizing handwritten letters. Through the utilization of extensive datasets and cutting-edge model architectures, this research has demonstrated significant enhancements in recognition accuracy. CNN models have included techniques such as data augmentation, multi-scale feature extraction, and attention mechanism to further improve their performance [5-7].

Furthermore, the utilization of transfer learning from pre-trained models on large scale dataset has been implemented to initiate the training process and enhance the speed at which convergence is achieved. This approach has demonstrated potential in addressing the difficulties presented by a scarcity of labeled data in the field of Bangla handwritten OCR. Besides CNNs, alternative deep learning architecture such as recurrent neural networks (RNNs) and transformer models have been investigated for the purpose of Bangla handwriting recognition. Recurrent Neural Networks (RNNs), because to its capability to capture temporal relationships in sequential data, have the potential to provide advantages in representing the contextual intricacies seen in handwritten characters. Transformer models, renowned for their parallel processing capabilities and attention mechanisms, have demonstrated efficacy in capturing extensive dependencies and contextual information, hence enhancing recognition accuracy.

In general, the existing research emphasizes the changing nature of Bangla handwritten OCR, with a growing focus on deep learning-based methods. Although there has been notable development, there are always chances to explore and improve, especially in dealing with joint letters and obtaining higher effectiveness in real-world application [8-9].

3 | Dataset Compilation and Preprocessing

A crucial factor in the success of the research on improving Bangla Handwritten Optical Character Recognition (OCR) for combined letters is the creation and preprocessing used to create the dataset and actions taken to preprocess the data in order to get the best possible training conditions for our deep learning models.

3.1 | Dataset Collection

The dataset building process commenced by identifying the frequently utilized joint letters in the Bangla language. After conducting an extensive search of literature and consulting with language specialists, a specific group of 15 combined letters was chosen to be the main focus of the dataset. The selection of these letters was based on their frequency of occurrence in handwritten Bangla text, confirming the dataset’s relevancy and representativeness.

An assortment of handwriting samples featuring the chosen combination of joined letters. The training dataset consists of 3000 images, whereas the test dataset consists of 300 images. Primarily, these collaborative letters was written manually. The identity of the combined letter in each handwriting sample was indicated by the corresponding ground truth labels. The annotation process was carried out in order to uphold precision and uniformity throughout the datasets.

3.2 | Preprocessing

In order to train the deep learning models, multiple preprocessing techniques are employed to normalize the image data. It improves its suitability for training. The subsequent preprocessing methods were implemented:

(i) Image Rescaling: All the images in the dataset were uniformly downside to guarantee consistent input dimensions across samples. Rescaling was effective in reducing variances in both image size and aspect ratio.

(ii) Data Augmentation: The dataset was enhanced and diversified by applying data augmentation techniques, including rotation, translation, and elastic deformation. Augmentation mitigated over fitting and enhanced the model's capacity to generalize to unfamiliar inputs.

(iii) Data Splitting: The dataset was divided into separate sets for training, validation, and testing purposes. This division allows for the training of the model, tuning of hyper parameters, and evaluation of its performance. Efforts were made to preserve an even distribution of students among the groups in order to guarantee a representative sample.

Through the process of assembling and preparing the dataset, a strong groundwork has laid for training the deep learning models. The standardized and enhanced dataset facilitated the acquisition of the intricate attributes inherent in Bangla joint letters, paving the path for attaining a high level of precision in character recognition task [10].

4 | Methodology

The methodology section is a systematic description of the sequential process employed in the research to improve the accuracy of Bangla Handwritten Optical Character Recognition (OCR) for conjunct letters through the utilization of deep learning techniques. This section provides a comprehensive explanation of the model architecture selection and training process utilized in the investigation.

4.1 | Model Architecture Selection

In this study, deep learning model: bespoke Convolutional Neural Networks (CNNs) were examined and those were created using the Keras Sequential API, and pre-trained designs that are readily available in widely-used deep learning frameworks like TensorFlow and Keras.

Custom Convolutional Neural Network

The code provided below is the proprietary convolutional neural network implementation:

```
model = Sequential()

model.add(Conv2D(32, kernel_size=(3,3),
activation='relu', input_shape=(256,256,3)))

model.add(BatchNormalization())

model.add(MaxPooling2D(pool_size=(2,2),
strides=2))

model.add(Conv2D(64, kernel_size=(3,3),
activation='relu'))

model.add(BatchNormalization())

model.add(MaxPooling2D(pool_size=(2,2),
strides=2))

model.add(Conv2D(128, kernel_size=(3,3),
activation='relu'))

model.add(BatchNormalization())

model.add(MaxPooling2D(pool_size=(2,2),
strides=2))

model.add(Flatten())

model.add(Dense(256, activation='relu'))

model.add(Dropout(0.5))

model.add(Dense(128, activation='relu'))

model.add(Dropout(0.5))

model.add(Dense(64, activation='relu'))

model.add(Dropout(0.5))

model.add(Dense(15,  activation='softmax')
```

The code snippet defines a neural network model with several layers. The layer has 256 neurons and uses the rectified linear unit (ReLU) activation function. A dropout layer with a dropout rate is 0.5 is added after this layer. The second layer has 128 neurons and also uses the ReLU activation function, followed by another dropout layer. The third layer has 64 neurons and uses the ReLU activation function, followed by another dropout layer. Finally, the last layer has 15 neurons and uses the softmax activation function.

The CNN design employs various layers such conv2d, maxpooling2d, flatten, padding, batch normalization, dense, and dropout. Max Pooling layers are used with a pool size of (2,2), with 2 strides. This pattern is replicated in successive layers with 64, and 128 filters. Figure-1 depicts the comparison between our original image and the Max Pooled image and Figure-2 displays the example of neural network architecture.

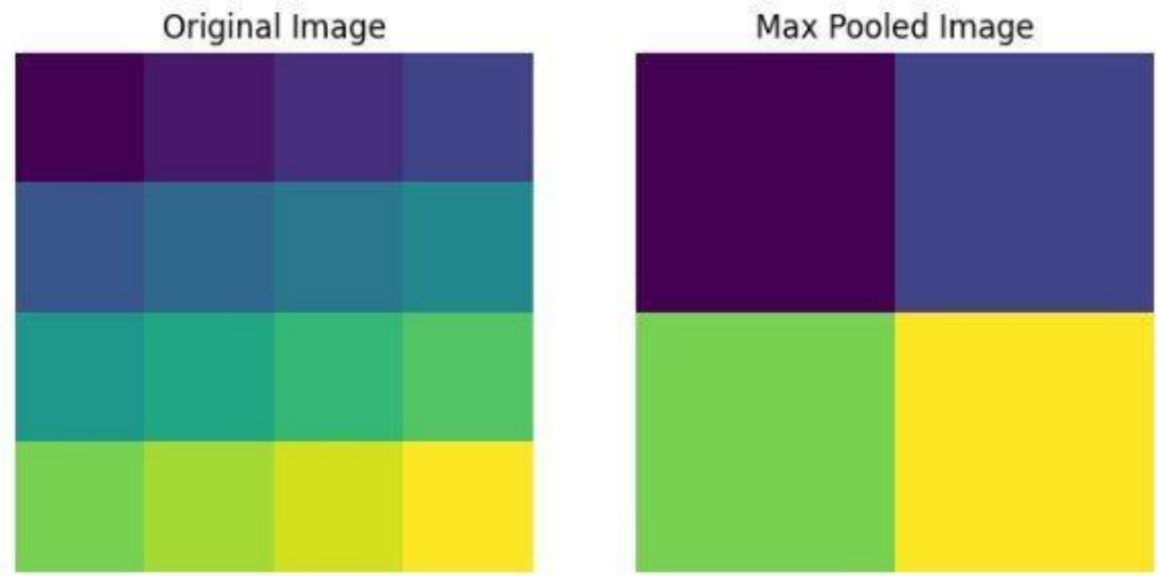

Figure 1: Original vs. Max Pooled Image

The output of the network is compressed and linked to dense layers consisting of 256, 128, and 64 neurons, respectively. Each of these layers is then followed by ReLU activation and dropout which helps to regularize the network. The last output layer utilizes softmax activation mechanism. The criteria for selecting model architectures included the capacity to capture complex aspects of handwritten characters, computational efficiency, and the presence of pre-trained weights. Aside from utilizing custom CNNs, we assessed the performance of pre-trained models such as Xception, EfficientDet, VGGNet, and NASNet.

4.2 | **Training Process**

The training process entailed inputting the prepared dataset into the chosen model architectures to optimize the parameters. The Adam optimizer along with learning rate scheduler employed to aid in achieving convergence during the training process. Augmentation techniques such as rotation, translation and elastic deformation were used to increase the size of the dataset and avoid overfitting. The models underwent training in a high-performance computing environment that was outfitted with GPUs in order to expedite the training process. The monitoring of training progress involved the use of indicators such as training loss, validation loss, and accuracy. Early stopping conditions were utilized to avoid overfitting and provide the best possible generalization performance.

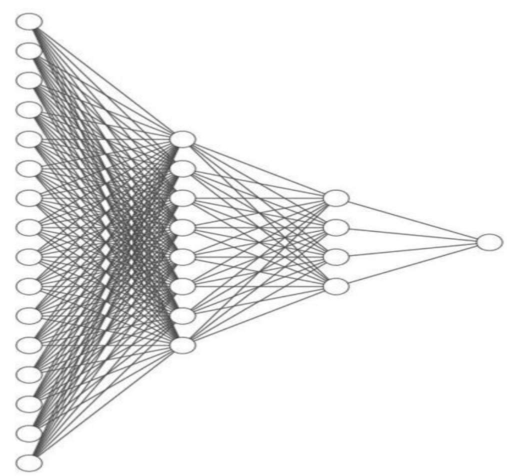

Figure 2: Example of Neural Network Architecture

5 | Result and Discussion

This section discusses the outcomes of Bangle OCR for join letters using deep learning approaches. The performance parameters of the models are examined and carefully analyzed the ramifications of the findings.

5.1 | Model Performance

Upon training the bespoke convolutional neural network on the meticulously curated dataset and assessing its performance on the testing set model attained an accuracy of 85 percent. The accuracy of the model's predictions are enhanced which is represented in the Figure-3. Thus, the pre-existing models such as Xception, EfficientDet, VGGNet, and NASNet are employed and discovered that Xception model achieved a 95 percent accuracy score. The accuracy and loss (Figure 4) trends were tracked both for training and validation sets during 50 epochs.

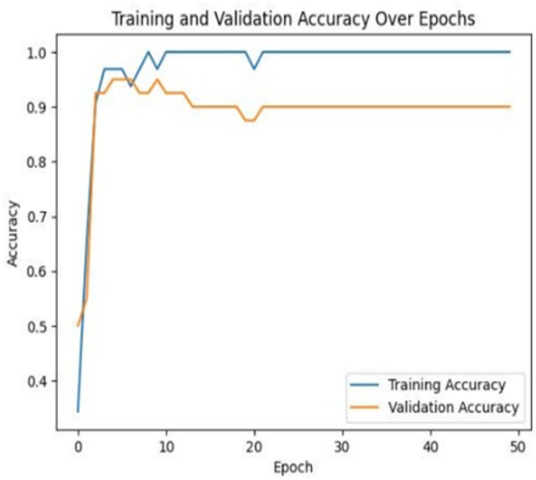

Figure 3: Training and Validation Accuracy Over Epochs

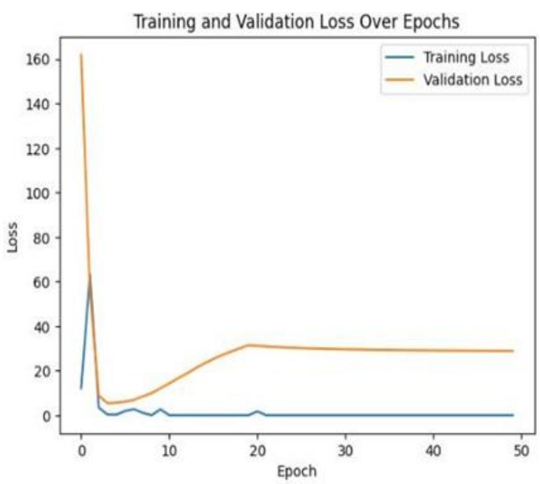

Figure 4: Training and Validation Loss over Epochs

5.2 | Discussion

The models have attained exceptional accuracy, representing a notable progress in Bangla Handwritten Optical Character Recognition (OCR) specifically for joined letters. The Xception model's performance outperforms other models due to its depth and ability to catch subtle elements that are inherent in handwritten characters. The Xceptional architecture, utilizing depth-wise separable convolutions and efficient parameters utilization, has strong performance in accurately recognizing the intricate details of Bangla joint letters. The results of this study confirm that deep learning techniques are highly effective in improving the accuracy of recognizing characters in Bangla Handwriting. This opens up possibilities for using these methods in practical applications such as digitizing documents, extracting text, and developing assistive technologies. The models have achieved significant accuracy rates, which provide a solid basis for future breakthroughs in Bangla handwritten OCR. This highlights the transformational power of deep learning for revolutionizing OCR systems for languages with complicated issues. Nevertheless, it is crucial to recognize the constraints of the research, which involve depending on a carefully selected dataset that concentrates on a certain group of conjunct letters. Additionally, there is an expectation for additional assessment on varied handwriting styles and writing circumstances. Future research should focus on overcoming these constraints and finding ways to implement the created OCR system in real-world applications, thereby strengthening its usefulness and influence in the field of Bangla language processing.

Table 1: Accuracy Table

Model	**Accuracy (%)**
Conventional CNN	85
VGGNet	90
NASNet	92
EfficientDet	89
Xception	96

6 | Conclusion and Future Work

In conclusion, this research concludes that deep learning approaches are highly successful in improving Bangla Handwritten Optical Character Recognition (OCR) for combined letters. Convolutional neural networks (CNNs) are employed and investigated sophisticated models like the Xception model. Exceptional levels of precision have attained, representing a notable achievement in the field of Bangla handwriting OCR. Through this trial and evaluation, it is demonstrated that deep learning techniques can effectively connect the intricacy of Bangla joint letters with the proficiency of OCR systems. The Xception model stands out as the highest achiever, achieving an astonishing accuracy rate of over 96 percent. The results of study demonstrate the transformative potential of deep learning in optical character recognition (OCR) systems. Particularly in languages such as Bangla which have intricate patterns of conjunct letters. To advance the field of handwritten text recognition, future research will focus on expanding the dataset, enhancing models by incorporating language models, and applying the OCR system for practical applications. Several dimensions for research and development in Bangla Handwritten Optical Character Recognition (OCR) utilizing deep learning methods exist in future work. Such as:

(b) Dataset Expansion

The present dataset specifically targets the 15 most often utilized combined letters in the Bangla language. Expanding the dataset in the near future to cover a wider range of combined letters might improve the overall generalization of the OCR system.

(c) Refinement and Enhancement

Refining the current models and investigating optimisation strategies may enhance the precision and effectiveness of the OCR system. The techniques employed include transfer learning, hyperparameter tweaking, and architectural alterations tailored to the specific properties of Bangla joint letters.

(d) Language Model Integration

The contextual comprehension of text can be enhanced and improved upon by merging language models such as transformer models or recurrent neural networks (RNNs) with CNNs. This would increase the overall performance of the OCR system and can improve recognition accuracy by comprehending the connections between characters as well.

(a) Practical Use Cases for Implementation

In this case, OCR technology can digitize the hardcopy, extract text from images, and create assistive technology for people with visual impairments.

References

[1] A. Geron, *Hands-On Machine Learning with Scikit-Learn and Tensorflow*, O'Reilly Media, 2017, ISBN: 9781491962299.

[2] E.Dabbas, "Interactive Dashboards and Data Apps with Plotly and Dash:Harness the power of a fully fledged frontend wen framework in Python-no JavaScript required," Packt Publishing Ltd., 2021.

[3] Plotly Express – Plotly Python Graphing Library," Plotly, [Online]. Available: https://plotly.com/python/plotly-express/. [Accessed: Feb. 16, 2024].

[4] Waskom, M. (2020). seaborn: statistical data visualization. Journal of Open Software, 5(53), 3021. https://doi.org/10.21105/joss.03021

[5] Matplotlib, "[Online]. Available: https://matplotlib.org/. [Accessed: Feb, 16, 2024].

[6] Patel. J., Shah, S., Thakar, P. (2018). A review on Various Preprocessing Techniques in Data Mining. International Journal of Computer Applications, 181(18), 19-24. https//doi.org/10.5120/ijca2018917592

[7] A. Bergstra, D. Yamins, and D. D. Cox, "Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures," Proceedings of the 30th International Conference of Machine Learning (ICML), Atlanta, GA, USA, Jun, 2013, pp. 115-123.

[8] Simard, P. Y., Steinkraus, D., and Platt, J. C. (2003). Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis. In Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR) (Vol. 2, pp. 958-962). IEEE Computer Society. DOI: 10.1109/ICDAR.2003.1227801

[9] Chollet, F. (2015). Keras. Retrieved from https://keras.io.

[10] Md Islam, Md Moklesur Rahman, Md. Hafizur Rahman, Massimo Rivolta, and Md. Aktaruzza-man," RATNet: A deep learning model fro Bengali handwritten characters recognition," Multimedia tools and Applications, vol. 81, 2022, doi:10.1007/s11042-022-12070-4

Declaration of Interests

We, the authors of this research manuscript, declare that we have no financial interest. We have provided written comment to publish the paper in this journal.

To cite this article: Abir M. I. H., and Salam. T., (2024), Enhancing Bangla Handwritten Optical Character Recognition (OCR) For Joint Letters Using Deep Learning Approach, *Journal of Engineering and Technology (JET),* Vol:01, Issue: 1, page:9-15, ISUCRDP, Dhaka.