Bangla Digit Recognition across Different Variations of CNN Models

ATIF RONAN, Brac University, Bangladesh
NAFIUN AL AMIN, Brac University, Bangladesh
NUSAIBA ZAMAN, Brac University, Bangladesh
AYAN HAIDER, Brac University, Bangladesh
ASIF ALI, Brac University, Bangladesh

Abstract

With so many literature about Bangla character or digit recognition, it was important to see if we could find a better solution to detect the Bangla handwritten digits. This research recognizes the Bangla handwritten digits with several models, and compares them to find the model that gives the best results.The dataset used in this research is NumtaDB[3], which consists of about 85,000 images of bangla handwritten digits. The dataset was trained under several versions of the Convolutional Neural Network (CNN). Firstly, the Simple Convolutional Neural Network was tested, followed by a merged CNN model with Logistic Classifer, a merged CNN model with Random Tree Classification, and an Ensemble approach of two variations of Simple CNN. When comparing the accuracy of the models, it was found that the most accurate of them all was the Ensemble model, which gives an accuracy of 94.34%, followed by the Simple CNN model, which gives an accuracy of 93.06%, CNN with Random Forest Classification gave 70.69%, and CNN with Logistic Regression Classifier gave 66.21%. This research focused on training the merged and Ensemble model as a first step in building a literature that might help build more efficient and accurate models in the future. The research, along with the findings, also recognises the Limitations it had, and proposes ideas for Future Research, which can be implemented to get better results.

Keywords: Bangla Handwritten Digit Recognition; NumtaDB; Merged CNN Models; Ensemble Method

ACM Reference Format:

Atif Ronan, Nafiun Al Amin, Nusaiba Zaman, Ayan Haider, and Asif Ali. 2024. Bangla Digit Recognition across Different Variations of CNN Models. 1, 1 (May 2024), 9 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

In the era of digital transformation, the ability to recognize and interpret visual data plays a pivotal role in automated document processing to image-based authentication systems. In this paper, we have focused on the evaluation of Convolutional Neural Network (CNN) for accurately classifying handwritten Bengali digits, which poses significant challenges for traditional recognition algorithms. Furthermore, this paper explores the integration of ensemble techniques with CNN architectures to harness the collective intelligence of multiple models, thereby enhancing the classification accuracy and resilience to variations. By combining the strengths of individual CNN models within an ensemble framework, the proposed approach seeks to mitigate the limitations associated with single-model approaches and

Authors' Contact Information: Atif Ronan, Brac University, Dhaka, Bangladesh, atif.ronan@g.bracu.ac.bd; Nafiun Al Amin, Brac University, Dhaka, Bangladesh, NafiunAlAmin@g.bracu.ac.bd; Nusaiba Zaman, Brac University, Dhaka, Bangladesh, nusaiba.zaman@g.bracu.ac.bd; Ayan Haider, Brac University, Dhaka, Bangladesh, ayan.haider@g.bracu.ac.bd; Asif Ali, Brac University, Dhaka, Bangladesh, asif.ali@g.bracu.ac.bd.

Manuscript submitted to ACM

achieve superior performance in handwritten digit recognition tasks. Thus, by leveraging insights from applied database principles and machine learning methodologies, this research facilitates the development of more robust and inclusive digital solutions.

2 RELATED WORKS

2.1 Handwriting Recognition based on CNN Models

This study [2] presents an improved and unique CNN-based approach for recognizing Bangla handwritten number characters. It uses updated pre-processing and trains three CNNs on various datasets, and a final decision has been made by combining their decisions. In this paper, a multiple Convolutional Neural Network (MCNN) approach is proposed, which focuses mainly on 2 steps. Firstly, the preprocessing of raw bangla handwritten images of numerical characters and the production of three training sets. Secondly, training 3 CNN models and combining their data. The dataset consists of scanned pictures of handwritten Bengali numeral characters acquired from postal mail pieces from people of all demographics and educational levels. Then the dataset is split into train and test sets, where the training set consists of 19392 images and the test set consists of 4000 images. After that, the software Matlab R2015a is used to transform the images into binary images, adjusting the foreground -background colors, cropping and resizing dimensions, and generating patterns using double type matrices to enhance quality. Using the MCNN method and its unique preprocessing and rotation techniques, superior performance in recognising handwritten numerical characters can be achieved compared to the previous method, as can be seen through the high accuracy and benchmark dataset.

The paper[1] seeks to fulfill the need for a text classifier capable of searching for handwriting to convert them digitally and identifying who is the author. The paper is concerned with developing the offline mode of a Handwritten text classifier which is image based. The paper identified a model using CNN as an effective tool to use for the task of HTR and that such a model was superior to the SVM model overall.The paper also proved that CNN model was capable to performing high accuracy identification of the author of a handwritten text image after training with a person's handwriting samples. Handwriting samples, including paragraphs of text, were collected from different people before being digitized into images of fixed dimensions, which was pre-processed to remove noise and empty data. Two models were prepared using CNN and SVM respectively. The models were then fed training data consisting of 75 images from 3 different individuals with equal distribution, and then testing was conducted with 30 images from the same 3 individuals. The model attempted to predict and identify which image corresponded to which person, and the classification results were then recorded. After 26 epoch CNN reached 100% accuracy and after 50 epoch the training loss was near 0. CNN and SVM both had good validation accuracy. The model was able to correctly identify a person's handwriting with 96.3% accuracy.

The motivation behind this research [7] is to address the challenges posed by Bangla characters such as digits containing distinct shapes and strokes, through a proposed model containing a seven layered D-CNN architecture. This paper contributes to the development of a seven layered D-CNN model on a large dataset containing handwritten bangla characters in the form of digits unlike previous research using printed documents or traditional machine learning algorithms. The large dataset containing 6000 images of handwritten Bangla digits used in the paper was the CMATERdb 3.1.1 dataset. The dataset was divided into 10 categories for each Bangla digit containing 600 images. The CMATERdb

3.1.1 dataset was divided into 4000 training images and 2000 testing images. The D-CNN architecture contains three main layers, the convolution, pooling and fully connected layer, however for this model, it starts with an input layer and ends with an output layer. The LeNet-5 architecture inspired the proposed seven layered D-CNN model. Moreover, to optimize and make learning smoother for the model the Adam optimiser was used. The CMATERdb 3.1.1 dataset was trained and tested on the LeNet-5 architecture producing 96.80% accuracy. The results of the proposed seven layered D-CNN model had a 97.6% accuracy on testing data and 99.9% accuracy on training data.

The paper [5] addresses the lack of organized databases for Bengali Transcribed Digits and the need for improved Optical Character Recognition (OCR) in the Bengali language, highlighting the importance of advancing technology in the context of Bengali language processing, specially for obscure words in Bengali books, novels, and newspapers etc. The primary contribution of this paper is proposing a novel approach for extracting Bengali words from digital images by implementing image classification and Convolutional Neural Network (CNN) models, ultimately verifying the accuracy of the identified words. This paper utilizes CNN models, including VGG16, for text recognition and verification, since CNN is well-suited for capturing spatial hierarchies in data. Then characters are separated from the Bengali words in digital images using the trained CNN models, this process involves identifying characters and reconstructing them into complete words. The converted digital font is compared against a dataset to determine the correctness of the identified words, this comparison with existing works and models provides insights into the effectiveness of the approach.The research achieves a high accuracy rate of 98% word verification, the proposed model demonstrates its potential for enhancing OCR system and language specific tools for Bengali language.

This paper [8] addresses the importance of recognizing bangla digits and the challenges faced in identifying handwritten digits. In reality, this paper aimed to overcome those challenges by developing a deep CNN model that gives robust performance and high accuracy for the NumtaDB dataset, which consists of large, unbiased and unprocessed data. Similar research has been conducted before this paper; however, most of the studies used biased datasets, since unbiased datasets like NumtaDB were not available back then. The NumtaDB dataset consists of about 85,000 Bangla handwritten digits, and it is said to be unbiased in respect to geographic location, age, and gender, making it a suitable choice for the authors to check the performance of their proposed model. Their proposed model consists of two parts - preprocessing of the images and deep CNN. Preprocessing of the images includes resizing and grayscaling, interpolation, removing blur from images, sharpening images, and removing noise from images. Secondly, the custom fully functional deep convolutional neural network consists of 6 convolutional layers and 2 fully connected dense layers, along with a Retrified Linear Unit (ReLu), which is used to activate all the layers. Lastly, the NumtaDB dataset consists of training and testing sets, which are split into 85% and 15%, respectively, and the training set is later split into an 80:20 ratio for training and validation sets, respectively. The proposed model achieved an accuracy of 92.72%, which is a good number considering the large and unbiased NumtaDB dataset used. Additionally, the calculated precision, recall, and F1 scores are 94.28%, 94.17%, and 94.19%, respectively.

2.2 Ensemble Approach to CNN Models

The paper[4] aims to enhance the accuracy of recognizing handwritten characters in the Bangla language using an ensemble approach with pre-trained CNN models. The datasets that were used were BanglaLekha-isolated, Ekush and MatrrivaSha, all of which contain handwritten characters in the Bangla language. The proposed model employed an

ensemble technique that combined predictions from 4 pre-trained CNN architectures- ResNet50, DenseNet121, Xception, and EfficientNetB0. The model achieved a high accuracy of 97.83% on BanglaLekha, 97.78% on Ekush, and 97.01% on MatrrivaSha, showcasing superior performance compared to existing works in Bangla handwritten character recognition.

2.3 CNN Models for Feature Extraction

The paper [6] aims to improve the existing Optical Character Recognition (OCR) technology for handwritten text recognition. The motivation behind the paper is the lack of efficiency behind handwritten text recognition and analysis such as language constraints and accuracy issues. The dataset used is the IAM dataset, consisting of handwritten texts in both English and German. The dataset is annotated and includes variety in the form of different handwriting styles, page layouts and quality of writing.The dataset is pre-processed and then it undergoes feature extraction using popular models such as CNNs, HOG or SIFT. Secondly, the handwriting recognition is done by models such as RNNs, HMMs and SVMs. The paper researches various combination of models and datasets, but the highest accuracy at 96% was achieved using the combination of CNN for feature extraction, RNN for handwriting recognition and CTC for word decoding.

3 METHODOLOGY

3.1 Dataset Description

The dataset used on the different CNN models is the NumtaDB [3] dataset, consisting of over 85,000 images of handwritten Bangla digits, each approximately 180x180 pixels in dimension. The NumtaDB dataset was used due to its unbiased nature against geographical location, gender and age. It has been collected from six different sources at different times to increase variation and thus each source has been differentiated within folders labeled from a to f, each containing differentiated training and testing sets for all labeled sources with an exception for source f. For example, source a will contain training a and testing a. This is done to ensure that handwriting samples from the same contributor are not present in both test and train sets. Additionally, less information was available on the contributors for source f and thus it has been labeled as testing f, without separate training and testing folders. Moreover, two more testing folders containing augmented datasets from a and c are present. These images are augmented through spatial transformations, superimposition, occlusion and by changing or shifting brightness, contrast, saturation, noise and hues, to create more challenging test sets.

3.2 Initial Exploratory Data Analysis

From the NumtaDB dataset, data from 3 of the included training databases, training-a, training-b, and training-c, were merged together to create a dataset of 24298 images, and this merged dataset was used for the model. In order to get the dataset ready a combination of data pre-processing and exploratory data analysis was performed based on initial viewing and previous knowledge of the data.

- **Data diversity of initially suspected redundant columns**
  The unique values present in columns “database name original”, “contributing team”, “database name” were determined.

- **Correlation between target variable and other suspected redundant columns**
  The correlation between “scanid” and the target variable “digit” was determined.

- **Distribution of values for the target variable “digit”**
  The distribution of images present in the dataset for each digit value [0-9] was determined.

```
Unique values in "database name original": ['BHDDB' 'B101DB' 'OngkoDB']

Unique values in "contributing team": ['Buet_Broncos' 'Shongborton' 'Buet_Backpropers']

Unique values in "database name": ['training-a' 'training-b' 'training-c']
```

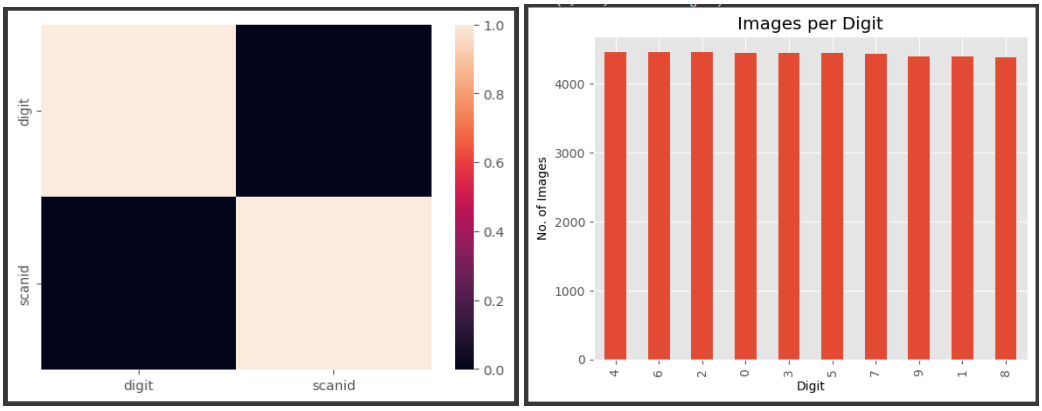


Fig. 2. Initial EDA Results

From the results of the initial EDA, the dataset was refined while the redundant columns were discarded. Columns “database name original” and “contributing team” were considered redundant due to low diversity, as our end goal is to be able to check images from any source, not just the handful present in the dataset. Additionally, column “scanid” was considered redundant as it essentially acts as an identity column and has no correlation with our target variable, making it useless for our required task. Furthermore, as there was no imbalance in our target variable it was concluded there would be no bias towards any digit during the training phase of our model.

3.3 Data Pre-processing

Based on the initial EDA the following operations were performed on the dataset to remove the redundant columns, and ready the dataset for modelling.

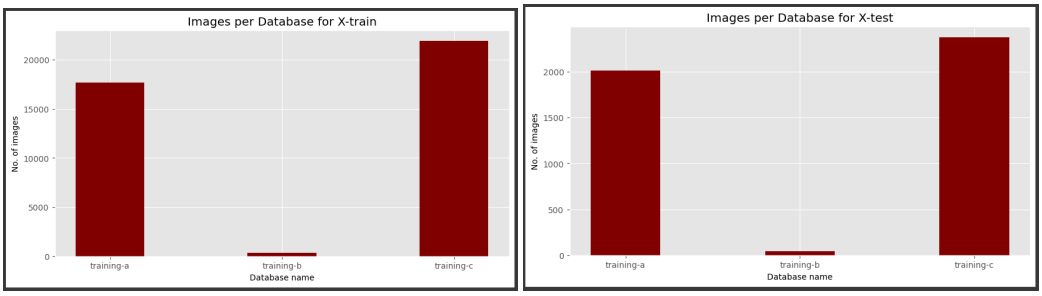
- Redundant columns “database name original”, “contributing team”, “scanid” were removed and discarded from the dataset

- Null and Duplicate values were checked and found to be absent from the dataset, requiring no action
- The dataset was split into train and test sets, with a train set size of 85 percent
- The images of the train sets were resized into 28X28 pixels
- The images of the train sets were scaled to fit the new size
- The images of the train sets were augmented via rotating and shearing random images at random amounts
- The test set labels were converted to binary labelling, in order for smoother classification

3.4 Continued Exploratory Data Analysis

After the dataset was split into train and test sets, and then processed, further analysis were performed on the following factors:

- **Distribution of values from the 3 databases in X train and X test**
  The amount of images present in X train and X test from each of the 3 databases were determined.
- **Distribution of values for the target variable "digit" in Y train and Y test**
  The distribution of images present in Y train and Y test for each digit value [0-9] was determined.



Despite the seemingly large imbalance in the X train and X test sets, balancing was not required as having prior knowledge of and examining the images of databases training-a, training-b, and training-c shows that there are no images with significant uniqueness to a single database. Additionally, due to the size imbalance of 3 databases, balancing would require the loss of an extreme amount of data for there to be equal distribution. Additionally, there was no imbalance in our target variable in Y train set, and the minor imbalance in Y test set was negligible compared to large volume of images present for each digit.

3.5 Model Description

- **CNN**
  Similar to [8], which uses 6 convolution layers and 2 dense layers along with ReLu for activation, our simple

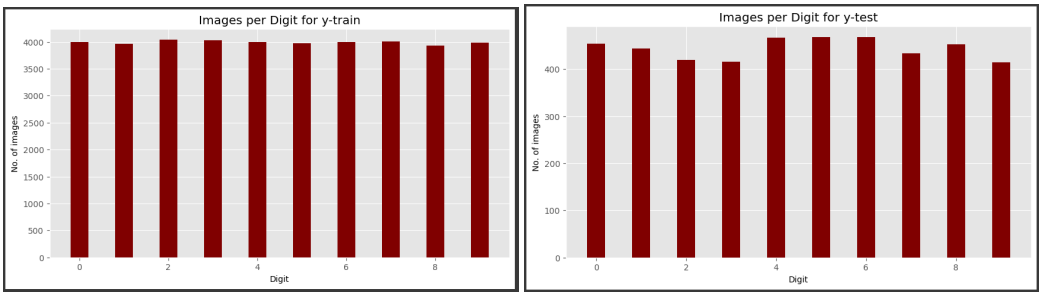


Fig. 4. Continued EDA Results

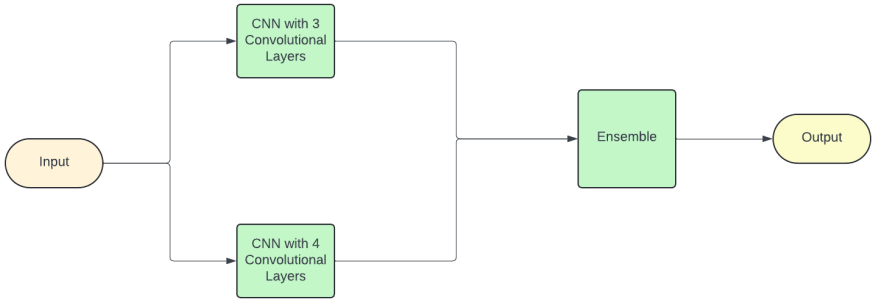


Fig. 5. Ensemble CNN Architecture

CNN consists of three convolution layers and one dense layer to do the classification tasks.

- **CNN with Logistic Regression**
  In this case, CNN using three convolution layers is used to extract features from the image. The extracted features are then passed on to the Logistic regression classifier for prediction.

- **CNN with Random Forest Classifier**
  Similarly, CNN with three convolutional layers is used to extract the features out of the bangla digit images and then the extracted features are passed on to the Random Forest classifier for classification.

- **Ensemble CNN model**
  In our proposed model, we tried to ensemble two different CNN models. Variations within the CNN were made by using three convolutional layers on one model and four convolutional layers on the other. According to [4], accuracy of the models can be significantly improved with the help of an ensemble mode. In our case, the mean of the two different CNN architectures are taken to combine the predictions and determine the accuracy of the model. Figure 5 shows the architecture for the proposed model for better understanding.

4 RESULTS AND ANALYSIS

Table 1. Accuracy and Models

Model	Accuracy
CNN	93.06%
CNN with Logistic Regression Classifier	69.21%
CNN with Random Forest Classifier	70.69%
Ensemble CNN model	94.34%

Table 2. Precision and Models

Model	Precision
CNN	93.26%
CNN with Logistic Regression Classifier	68.90%
CNN with Random Forest Classifier	70.49%
Ensemble CNN model	94.34%

Table 3. Recall and Models

Model	Recall
CNN	93.00%
CNN with Logistic Regression Classifier	69.14%
CNN with Random Forest Classifier	70.71%
Ensemble CNN model	94.34%

Here, we can see that the accuracy of both CNN as a feature extractor using logistic regression as a classifier and Random Forest as a classifier yields a very moderate accuracy of 69.21 % and 70.69 %, respectively. Logistic Regression is a linear classifier, which is why it can struggle to handle complex and non-linear patterns in the features extracted by CNN. On the other hand, although random forest is capable of handling non-linear relationships between features and target variables, it still lacks compared to deep neural networks. We can see that the simple CNN model had a large increase in accuracy, which is 93.06% compared to logistic regression and random forest, where it was used as both a feature extractor and classifier. Lastly, we can see that ensemble CNN has slightly better accuracy than the simple CNN model, which is 94.34%. Ensembling models can help prevent the risk of overfitting, which is done by combining multiple models that have different biases and patterns. Thus giving better accuracy than base models.

5 LIMITATIONS

The study will address some limitations that, if addressed, may have improved the results for the detection of Bangla digits through different CNN models. Firstly, the research requires a computer with high computational complexity to run the various CNN models with efficient speed. Thus, due to a lack of access to such a computer, the study had to limit the size of the dataset, for the models to run on the currently available computer. Secondly, due to time constraints,

the dataset did not undergo complex augmentation methods such as changing the ink colors or inversing the image colors within the dataset. Thus, the dataset did not contain more complicated enhancements that would challenge the implemented models. Lastly, there was a lack of research papers or literature on ensemble models. Thus, the implementation of different ensemble models was not conducted for this study. The improvement of these limitations that were encountered during this study, would greatly improve the reliability of the results obtained.

6 CONCLUSION AND FUTURE WORK

Overall this study has attempted to implement multiple different Convolutional Neural Network models to find the model that gives the best results for accurately recognizing Bangla handwritten digits. In order to test these models we utilised the NumtaDB dataset, developed from six different sources. From our findings, the best performance shown in this study was from an Ensemble model of two variations of Simple CNN with 94.34% accuracy, followed by the Simple CNN model with 93.06% accuracy, CNN with Random Forest Classification model with 70.69% accuracy, and CNN with Logistic Regression Classification model with 60.21%.

Further improvements can be performed by building on the limitations of this study, and improvements such as including more augmented photos from other datasets to train the models being a good starting point that we hope to expand upon in future studies. In addition, researchers are open to investigate every potential ensemble technique in the future to find the best performing strategy for combining CNN models. Furthermore, different CNN models and models other than CNN could be used in future research to broaden the view on the best method for Bangla digit recognition.

REFERENCES

[1] Sheikh Abujar, S.M. Saiful Islam Badhon, and Prakash Duraisamy. 2021. Handwritten Text Recognition for Non-Latin Languages using Deep Learning - Bangla. In *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. 1–6. https://doi.org/10.1109/ICCCNT51525.2021.9579870

[2] M.A.H. Akhand, Mahtab Ahmed, and M.M. Hafizur Rahman. 2016. Multiple Convolutional Neural Network Training for Bangla Handwritten Numeral Recognition. In *2016 International Conference on Computer and Communication Engineering (ICCCE)*. 311–315. https://doi.org/10.1109/ICCCE.2016.73

[3] Samiul Alam, Tahsin Reasat, Rashed Mohammad Doha, and Ahmed Imtiaz Humayun. 2018. NumtaDB-Assembled Bengali Handwritten Digits. *arXiv preprint arXiv:1806.02452* (2018).

[4] Kazi Fuad Bin Akhter and Tanvir Ahmed. 2023. An Ensemble approach of Pretrained CNN models for Recognition of Handwritten Characters in Bangla. 1–7. https://doi.org/10.1109/eSmarTA59349.2023.10293702

[5] Shakibul Hasan, Mohammad Abu Nadif, Nasim Bin Rahman, and Masud Rana. 2023. A Bengali Word Identification and Verification Using Machine Learning Approach. In *2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*. 1–5. https://doi.org/10.1109/ICAECT57570.2023.10117745

[6] L. Navya, MD.Farhan Ali, K.Pavan Sai, K. Shyam, and Alabazar Ramesh. 2023. Handwritten Text Recognition Using Deep Learning Techniques. In *2023 Annual International Conference on Emerging Research Areas: International Conference on Intelligent Systems (AICERA/ICIS)*. 1–5. https://doi.org/10.1109/AICERA/ICIS59538.2023.10420040

[7] Chandrika Saha, Rahat Hossain Faisal, and Md. Mostafijur Rahman. 2019. Bangla Handwritten Digit Recognition Using an Improved Deep Convolutional Neural Network Architecture. In *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*. 1–6. https://doi.org/10.1109/ECACE.2019.8679309

[8] Ashadullah Shawon, Md. Jamil-Ur Rahman, Firoz Mahmud, and M.M Arefin Zaman. 2018. Bangla Handwritten Digit Recognition Using Deep CNN for Large and Unbiased Dataset. In *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*. 1–6. https://doi.org/10.1109/ICBSLP.2018.8554900