IEEE - 49239

A New State of Art Deep Learning Approach for Bangla Handwritten Digit Recognition using SVM Classifier

Md. Jahid Hasan
Department of Electrical and Electronic Engineering
Hajee Mohammad Danesh Science and Technology University
Dinajpur, Bangladesh
jahidnoyon36@gmail.com

Md. Ferdous Wahid
Department of Electrical and Electronic Engineering
Hajee Mohammad Danesh Science and Technology University
Dinajpur, Bangladesh
mfwahid26@gmail.com

Md. Shahin Alom
Department of Electrical and Electronic Engineering
Hajee Mohammad Danesh Science and Technology University
Dinajpur, Bangladesh
ashahin200@gmail.com

Mohammad Mahmudul Alam Mia
Department of Electronics and Communication Engineering
Sylhet International University
Sylhet, Bangladesh
mahmud_ece_ku@yahoo.com

***Abstract*—Deep Convolutional Neural Network (CNN) has earned remarkable success in computer vision technology, particularly in image classification task. With the growing interest, it has been also employed to recognize Bangla handwritten digits. In this paper, we have presented a deep convolutional neural network with support vector machine (CNN-SVM) in order to recognize Bangla handwritten digits. Here, we have designed and implemented a 14 layered deep CNN architecture as a feature extractor to acquire remarkable features automatically from Bangla handwritten digits image and SVM was used as a classifier to recognize Bangla handwritten digits. We have trained and tested the proposed model with two different Bangla handwritten digit dataset namely NumtaDB and BanglaLekha-Isolated. The accuracy was found 99.53% on NumtaDB dataset and 99.59% on BanglaLekha-Isolated dataset. We also used our prepared dataset which is completely unseen to proposed CNN-SVM model in order to evaluate the generalization capability of the proposed model. The CNN-SVM model achieved a test accuracy of around 99% on our prepared dataset that indicates that the proposed model is reliable and convincing.**

Keywords*—Bangla handwritten digit recognition, deep CNN, computer vision technology, image classification, support vector machine.

I. Introduction

Handwritten digit recognition has become a great interest of research with the improvement of computer vision technology due to its wide application in different fields including, ZIP code recognition [1] postal system automation [2], automatic processing of bank checks [3], reading ID card, license plate recognition, automatic passport reading and so on. Many researches have been done on recognizing handwritten digit in English [4], Urdu [5], Chinese [6], French [7], Oriya, Tamil, Telegu [8]. But researches on Bangla handwritten digit recognition are inadequate compared to other languages due to lack of big dataset with lots of variance and complex nature of Bangla numerals. Although currently over 250 million people speaks Bengali natively in the world [9]. It is the seventh most spoken native language in the world by population [10]. Even this day the number is increasing rapidly. So, it is important to recognize Bangla handwritten digit in an automatic way to eases the human effort by utilizing an effective computer vision and artificial intelligence system.

Recent development in machine learning due to bloom of convolutional neural network has created a buzz word in computer vision technology and automatic image recognition and classification task. LeCun et al. [3] proposed a gradient based approach of CNN to recognize numeral and characters from a document. Several researchers have worked to recognize Bangla handwritten digit, but yet still there is no significant development. Wen, Y et al. [11] proposed two approaches for recognizing handwritten Bangla numerals. One is the image reconstruction recognition approach and the other one is the direct feature extraction approach combined with PCA and SVM. Bashar et al. [12] proposed a digit recognition system based on windowing and histogram techniques. Windowing technique is used to extract uniform features from scanned image followed by generating histogram from extracted features. M. R. Mamun et al. [13] proposed an ensemble method of three Xception networks for recognizing Bangla handwritten digit and evaluated the performance of 96.69% on NumtaDB dataset. Another work has been done on recognizing Bangla handwritten numeral recognition by R. Noor et. al [14], where an ensemble method of two CNN model where one is 15 layered and another one is 12 layer and achieve 96.79% on NumtaDB dataset. In another work by O. Paul [15] on NumtaDB dataset, the author has applied some pre-processing like resize, grayscale conversion, thresholding, cropping etc and best accuracy has found by using CNN which is 91%. H. Zunair et.al [16] used a pre-trained VGG16 model with transfer learning method on NumtaDB dataset and achieved 97.06% by 50 epochs. S. M. A. Hakim et. al [17] proposed a nine-layer DCNN architecture for classifying Bangla numerals on BanglaLekha-Isolated dataset and has achieved 99.44% accuracy. All the methods mentioned above have used CNN architecture for extracting features and softmax for classification. This CNN-softmax combination requires a significantly more time to train as well as testing than CNN-SVM model.

To improve the performance of Bangla Handwritten Digit Recognition (BHDR), we present a new hybrid approach based on deep convolutional neural networks for extracting features from input images and an SVM classifier for classification. This approach has recently shown excellent performance in many pattern recognition tasks and deep learning applications, such as bacteria classification [18], recognizing handwritten digit[19] etc. The model significantly increases the recognition accuracy as well as decreases misclassificattion. To evaluate this model, we have trained the proposed model with two Bangla Handwritten digit dataset NumtaDB and Bangla-Lekha Isolated. The model has achieved 99.53% accuracy on NumtaDB and 99.59% accuracy on Bangla-Lekha Isolated dataset while testing. The model has also shown remarkable generalization performance as it achieved around 99% test accuracy on prepared dataset which is completely unseen to it.

II. Dataset Preparation

The task of dataset preparation consists of two parts: A. Dataset Description & B. Data Pre-processing.

A. *Dataset Description*

NumtaDB[20] and BanglaLekha-Isolated[21] are two well-known customizable handwritten digit dataset. NumtaDB has a collection of more than 85000 Bangla handwritten digit images. The dataset contains representation from diverse regions of Bangladesh as well as unbiased in terms of geographic location and age. The dataset is divided into six different directories with codename 'a', 'b', 'c', 'd', 'e', and 'f'. Each dataset is split into 85:15 for training and testing purpose except 'f'. The dataset of codename 'f' and two heavily augmented datasets from codename 'a' and 'c' along with test set of codename 'a', 'b', 'c', 'd', 'e' was used for testing purpose and rest of the image is used for training. Then finally 72044 images were used for training and 13552 images for testing.

BanglaLekha-Isolated dataset consist of children handwritten including Bangla character, digit and combined character samples. From there we have separated only 10-digit classes where each class contains average 1972 images and total 19748 images. We split this dataset into 80:20, where 80 % data is used for training and 20 % for testing.

Our prepared dataset is comprised of 10-digit classes where each class contained 100 images. The dataset was acquired from student of Hajee Mohammad Danesh Science and Technology University, Dinajpur, Bangladesh. A data collection form which had a regular grid pattern was provided to the student to write down the Bangla numerals at their natural writing style. We maintained equal gender ratio among the participating students. Then, the filled-in form was collected and scanned using a CanoScan LiDE 300 scanner with a resolution of 300 dpi. The scanned image is converted to grayscale before applying canny edge detection algorithm to extract the edges. After that, the images were segmented by grid line to get the individual Bangla numeral image and finally, separated in folders by their image classes.

B. *Data Pre-processing*

NumtaDB dataset and BanglaLekha-Isolated dataset both comprised of wide variety of images which makes them very difficult to work with. The images of the datasets also contained noise in various forms such as variation in size and color, salt and pepper noise, Gaussian noise etc. Thus, Preprocessing techniques have been applied to remove these noises in order to avoid misclassification. At the beginning of preprocessing, all images were resized to a particular size 32 × 32 pixel and then, all images were converted to grayscale image. This method of preprocessing was used to reduce the computational cost greatly as well as to maintain the uniformity of samples. Then, we have applied Gaussian filter to remove Gaussian noise which is caused by poor illumination or high temperature of an image and salt-and-pepper noise was removed using median blurring filter. Finally, an augmentation process has been applied to create variance among samples that can occur when someone else writing the digits. This also increase the number of image samples that helps the model to avoid over-fitting. This data augmentation involves random flipping, horizontal or vertical translation up to 10px and zooming up to 25%.

III. Deep Convolutional Neural Network

Nowadays, Deep Convolutional Neural Network (CNN) is one of the most influential state-of-art methods for image recognition. A CNN mainly comprised of three layers namely convolution layer, pooling layer and fully connected layer. These three layers can be repeatedly used to form a deep CNN architecture shown in Fig. 1. A particular organization of these layers starts with input layer and ends up with output layer. Following subsections contains description of these layers.

A. *Input Layer*

Feature is an important and unique property of an image. The first layer in every CNN is input layer. It holds the raw pixel values of the input image. Input image can be fed to the input layer of the CNN by applying pre-processing technique in order to achieve better accuracy. The shape of the input for an image would be (image height) x (image width) x (image depth). For RGB image, the image depth is equal to 3 but in the case of gray scale images the image depth is 1.

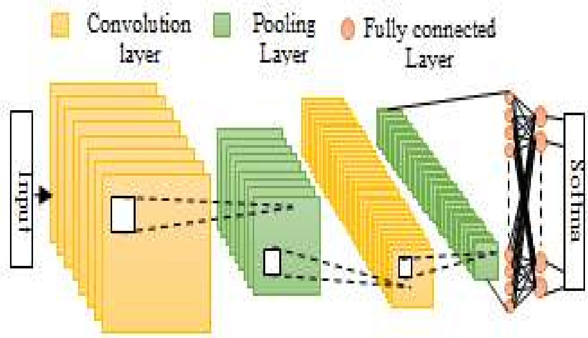

Fig. 1. Architecture of the complete CNN model

B. *Convolutional Layer*

Convolutional layer is the major building block of CNN. It is consisting of a set of learnable convolution filters or kernels. These filters can learn feature representations of the input image to create a feature map. Each neuron in a feature map is connected via set of trainable weights to a neighborhood of neurons in the previous layer. In order to obtain a new feature, the input feature maps are first convolved with a learned kernel and then the results are passed into a nonlinear activation function. The execution of convolution operation is performed by sliding the filter over the input. The convolved operation performed a matrix multiplication at every location of an image and sums the result onto the feature map. We will get different feature maps by applying different kernels. The typical activation functions are sigmoid, tanh and ReLU. Several features can be extracted at each location by applying different feature maps within the same convolutional layer.

C. Pooling Layer

Pooling layer is responsible for extracting dominant features by reducing the spatial resolution of the feature maps. This helps to achieve spatial invariance to input distortions and translations as well as increased the computational performance. It is usually placed between convolutional layers. The size of feature maps in pooling layer is determined according to the moving step of kernels. The typical pooling operations are average pooling and max pooling. We can extract the high-level characteristics of inputs by stacking several convolutional layers and pooling layer.

D. Fully Connected Layer

Fully connected layer converts the multidimensional input feature map matrix into a one-dimensional feature vector and performs the function of high-level reasoning. Each neuron of the fully connected layer is bind with every other neuron of the previous layer and each link carries its own weight.

E. Output Layer

The last layer of CNN is output layer. It is responsible for producing the output probability of each given input class. A softmax unit is used to obtain the output probability. Softmax is commonly used because of it generate a well-performed probability distribution. The probability of each output classes sums up to 1. The class containing the largest value will be the correct class.

IV. Proposed CNN Architecture

We have proposed a convolutional neural network with ReLU nonlinear function to recognize Bangla Handwritten Digit. The proposed network has eight convolutional layers and four pooling layers. The kernel size of each filter of the first two convolutional layers is 5x5 and the remaining six convolutional layers contain each filter of kernel size 3x3. Pooling layers are placed after every two convolutional layers in the proposed CNN architecture to decrease the number of parameters. All the pooling layers are of the max-pooling type and have the dimension 2x2. Finally, in order to determine the final classes of the digits, two fully connected layers are used. The first fully connected layer contains 64 neurons and the second fully connected layer contains 10 neurons with softmax activation function, equal to the total number of output classes. The organization of these three types of layers is shown in Table I.

The ReLU activation function is used in each of the convolutional layers to introduce non-linearity [22]. We have eliminated over-fitting problem using Dropout [23]. Dropout basically removes some of the weight connections between two layers that are selected randomly. We also used batch normalization to normalize our input data before each maxpooling operation that increase the stability of a neural network by subtracting the batch mean and dividing by the batch standard deviation [24].We have mainly analyzed different methods of setting learning rate and different optimization algorithm for solving the optimal parameters of influence. Here, Adam optimizer makes the value of loss function smallest by finding an optimal parameter set [25].

TABLE I. Formation Of Proposed Model

Layer	No of Neurons /Filters	Activation Function	Kernel Size
Conv1	32	ReLU	5x5
Conv2	32	ReLU	5x5
MaxPool	-	-	2x2
Conv3	128	ReLU	3x3
Conv4	128	ReLU	3x3
MaxPool	-	-	2x2
Conv5	256	ReLU	3x3
Conv6	256	ReLU	3x3
MaxPool	-	-	2x2
Conv7	512	ReLU	3x3
Conv8	512	ReLU	3x3
MaxPool	-	-	2x2
Flatten_1	1024	-	-
FullyConnected1	64	ReLU	-
FullyConnected2	10	Softmax	-

V. Methodology

In this paper, we have presented a Bangla Handwritten Digit recognition scheme based on CNN and SVM. Here, we have developed a 14 layered CNN architecture which was mainly used as feature extractor to extract effective features automatically. After completing the feature extraction using CNN, we used SVM as the classifier for digit recognition because it provides high accuracy by tuning very few parameters.

A. Feature Extraction using Proposed CNN

In this work, the proposed 14 layered CNN architecture is used as feature extractor in order to achieve state-of-the-art accuracy in handwritten digit recognition. In order to extract mid-level to high-level representative and hierarchical features automatically, the input images of size 32 × 32 pixel are feed to the proposed architecture. The architecture automatically learns representations of input image with multiple levels of abstraction as it propagates through the multiple processing layers of CNN architecture. The architecture of the proposed 14 layered CNN is depicted on Fig. 2.

The parameters of the proposed CNN are optimized using Adam optimizer for effective extraction of features. It takes the previous value of parameter into consideration for making the learning process smooth. We have used different combination of batch-size and number of epochs to get minimum training time with optimal training accuracy. The final values of all training-parameters of CNN architecture are represented in Table II.

TABLE II. Final Values Of All-User Defined Parameters Of Proposed CNN Model

Proposed CNN	
Parameter	***Value***
Learning-rate of active layers	0.0001
Batch-size	64
Number of epochs	30

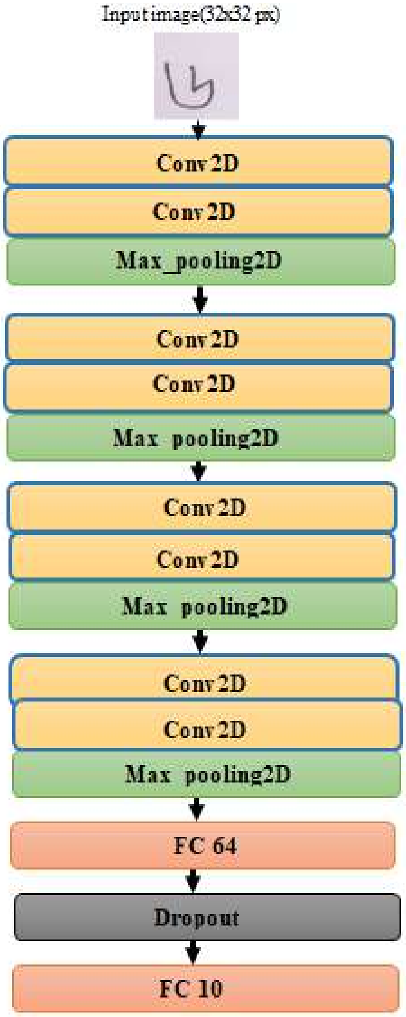

Fig. 2. DCNN architecture for feature extraction

B. *Classification*

We have chosen SVM with Radial basis function (RBF) kernel as classifier for handwritten digit recognition due to its excellent performance in solving linear inseparable problem. It finds the best separating hyperplane by mapping input in low dimensions into a higher dimension feature space and performs a good generalization. Here, all the features extracted using CNN was flattened by 'Global average-pooling' function and used to train the classifier keeping one-fifth extracted features separated for future usage. In order to achieve good classification accuracy, hyper-parameters of SVM was tuned using Grid-search algorithm. Finally, the classification accuracy of the classifier was predicted using the previously separated features. The final value of the SVM hyper-parameters was presented in Table III.

TABLE III. FINAL VALUES OF ALL-USER DEFINED PARAMETERS OF SVM CLASSIFIER

SVM	
Parameter	***Value***
Kernel	RBF
C	50
Gamma	0.0001

VI. RESULT AND DISCUSSION

In this section, we carried out experiments on two datasets namely NumtaDB and BanglaLekha-Isolated. To adequately asses the performance of CNN-SVM model, we have split the datasets into two parts, i.e. training set and test set in the ration of 80:20. In order to validate the performance of the model, 20% of the training images were used. We evaluated the proposed method using processor Intel core i7 @ 3.6GHz 64-bit, 8.00 GB RAM and free-space on SSD: 100 GB with 'Linux' operating system.

We train the proposed model to recognize Bangla handwritten digit with optimized hyperparameter and parameter values which achieves accuracy of 99.53% on the NumtaDB and accuracy of 99.59% on the BanglaLekha-Isolated dataset. Then, the performance of model was tested on test dataset of NumtaDB and BanglaLekha-Isolated which yield accuracy above 99% on both datasets. Fig. 3 and Fig. 4 showed the graphs of training and validation accuracy for training data with respect to number of epochs for both NumtaDB and BanglaLekha-Isolated respectively. From the figure it can be seen that, the training and validation accuracy keeps increasing with each epoch as the model is trained and during last 8 epochs it is above 99% for both the dataset.

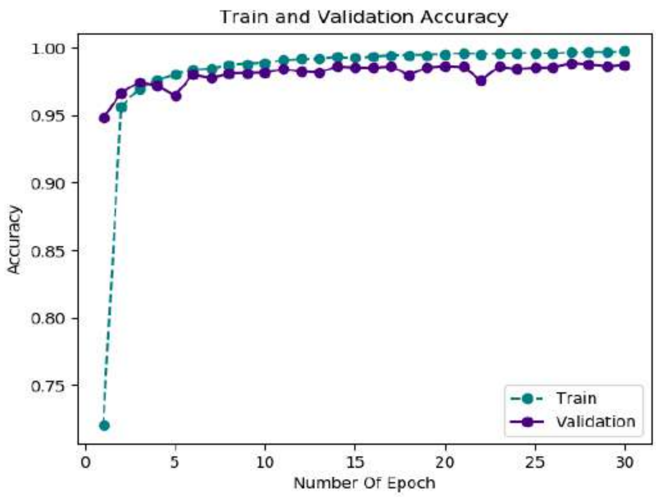

Fig. 3. NumtaDB dataset training and validation accuracy

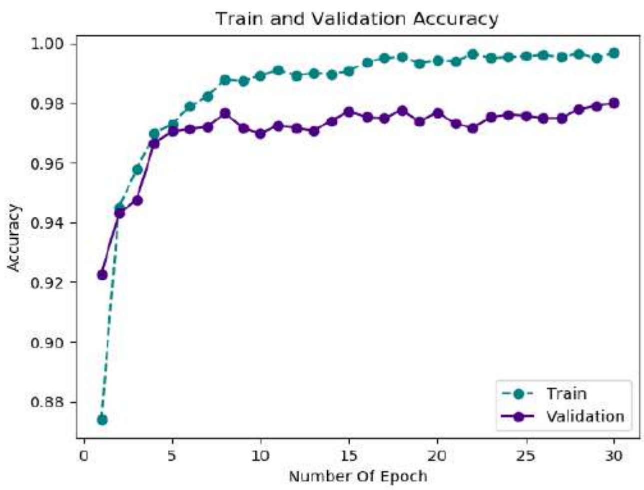

Fig. 4. BanglaLekha-Isolated dataset training and validation accuracy

The graphs of training and validation loss were shown in Fig. 5 and Fig. 6 for NumtaDB and BanglaLekha-Isolated respectively. It is observed that the training and validation loss for both the dataset were below 0.2 during the last 8 epochs.

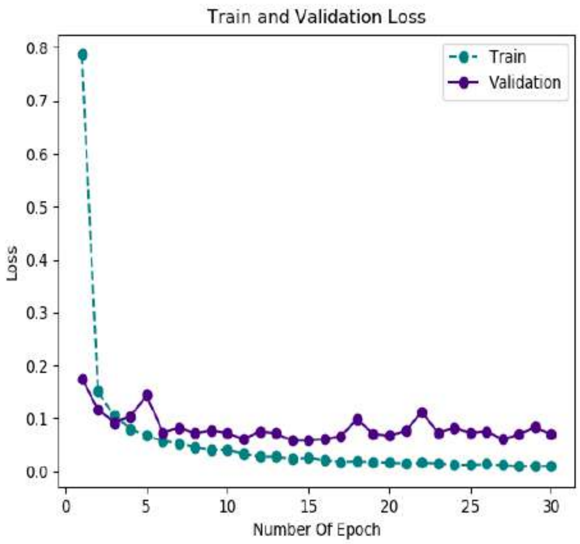

Fig. 5. NumtaDB dataset training and validation loss

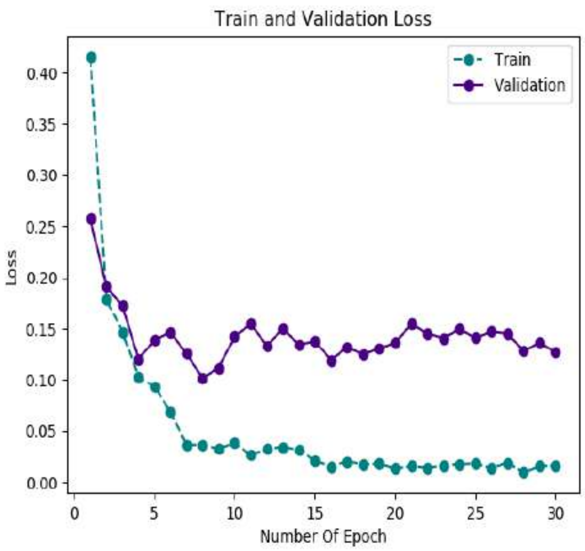

Fig. 6. BanglaLekha-Isolated dataset training and validation loss

The generalization of the proposed model was also evaluated in this work. In order to evaluate the generalization capability of the CNN-SVM model, the performance of the model have been tested by a completely unseen prepared test dataset. The CNN-SVM model trained with NumtaDB dataset has been achieved a score of 98.66% test accuracy on the prepared test dataset and when the model trained with BanglaLekha-Isolated dataset, it attained 99.10% test accuracy on the prepared test dataset.

Table IV shows a comparison of accuracy between previous known methods of Bangla handwritten digit recognition with the proposed model on these two datasets. We clearly observed from Table IV that the proposed model achieved state-of-art performance than others on both NumtaDB and BanglaLekha-Isolated dataset.

TABLE IV. COMPARISION OF PROPOSED MODEL WITH PREVIOUS WORKS FOR BANGLA HANDWRITTEN DIGIT RECOGNITION

Reference	Dataset	Methodology	Accuracy (%)
[13]	NumtaDB	Xcepption + transfer Learning	97.06
[14]	NumtaDB	Ensembling CNN	96.79
[17]	BanglaLekha-Isolated	DCNN	99.44
Proposed	NumtaDB	CNN+SVM	99.53
	BanglaLekha-Isolated	CNN+SVM	99.59

VII. CONCLUSION

In this paper, we have proposed a fourteen layered deep CNN with SVM model to recognize Bangla handwritten digits which provides both training and test accuracy above 99% on the two datasets. The proposed CNN-SVM model have also shown a good generalization capability. Therefore, it is noteworthy that the proposed method shows significant improvement over already devised classification methods for recognizing Bangla handwritten digit. However, we will intend to implement the proposed model for recognizing all the Bangla characters in the future.

REFERENCES

[1] Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, and L. D. Jackel, “Handwritten digit recognition with a backpropagation network,” in Advances in neural information processing systems, pp. 396–404, 1990.

[2] M. K. Nasir and M. S. Uddin, “Hand Written Bangla Numerals Recognition for Automated Postal System”, IOSR Journal of Computer Engineering (IOSR-JCE),vol 8, Issue 6, pp 43-48, 2013.

[3] 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.

[4] C. Viard-Gaudin, P. M. Lallican, S. Knerr, and P. Binter, “The ireste on/off (ironoff) dual handwriting database,” in Document Analysis and Recognition, 1999. ICDAR’99. Proceedings of the Fifth International Conference on. IEEE, pp. 455–458, 1999.

[5] M. W. Sagheer, C. L. He, N. Nobile, and C. Y. Suen, “A new large urdu database for off-line handwriting recognition,” in International Conference on Image Analysis and Processing. Springer, pp. 538–546, 2009.

[6] L. Jin, Y. Gao, G. Liu, Y. Li, and K. Ding, “Scut-couch2009—a comprehensive online unconstrained chinese handwriting database and benchmark evaluation,” International Journal on Document Analysis and Recognition (IJDAR), vol. 14, no. 1, pp. 53–64, 2011.

[7] R. Kumar and K. K. Ravulakollu, “Handwritten devnagari digit recognition: Benchmarking on new dataset.” Journal of theoretical & applied information technology, vol. 60, no. 3, 2014.

[8] U. Pal, N. Sharma, T. Wakabayashi, and F. Kimura, “Handwritten numeral recognition of six popular indian scripts,” in Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on, vol. 2. IEEE, pp. 749–753, 2007.

[9] (2020) “Ethonologue”, [Online]. Available : https://www.ethnologue.com/19/language/ben/.[Accessed: 10 April 2020].

[10] (2020) “The world factbook”, [Online]. Available: https://www.cia.gov/library/publications/ the-world factbook/geos/xx.html.[Accessed: 10 April 2020].

[11] Y. Wen, Y. Lu, and P. Shi, “Handwritten bangla numeral recognition system and its application to postal automation,” Pattern recognition, vol. 40, no. 1, pp. 99–107, 2007.

[12] M. R. Bashar, M. A. F. M. R. Hasan, M. A. Hossain, D. Das, “Handwritten Bangla Numerical Digit Recognition using Histogram Technique.,” Asian J. Inf. Technol., vol. 3, pp. 611–615, 2004.

[13] M. R. Mamun, Z. A. Nazi, S. U. Yusuf,”Bangla Handwritten Digit Recognition Approach with an Ensemble of Deep Residual Networks”,

International Conference on Bangla Speech and Language Processing (ICBSLP), 21-22 September, 2018.

[14] R. Noor, K. Mejbaul Islam and M. J. Rahimi, "Handwritten Bangla Numeral Recognition Using Ensembling of Convolutional Neural Network," 2018 21st International Conference of Computer and Information Technology (ICCIT), Dhaka, Bangladesh, pp. 1-6, 2018.

[15] O. Paul, "Image Pre-processing on NumtaDB for Bengali Handwritten Digit Recognition," 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), Sylhet, pp. 1-6, 2018.

[16] H. Zunair, N. Mohammed and S. Momen, "Unconventional Wisdom: A New Transfer Learning Approach Applied to Bengali Numeral Classification," 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), Sylhet, pp. 1-6, 2018.

[17] S. M. A. Hakim and Asaduzzaman, "Handwritten Bangla Numeral and Basic Character Recognition Using Deep Convolutional Neural Network," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox'sBazar, Bangladesh, pp. 1-6, 2019.

[18] T. Ahmed, M. F. Wahid, M. J. Hasan, “Combining Deep Convolutional Neural Network with Support Vector Machine to Classify Microscopic Bacteria Images”, 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 7-9 February, 2019.

[19] X. X. Niu, C. Y. suen, “A novel hybrid CNN–SVM classifier for recognizing handwritten digits”, Pattern Recognition, vol. 45, 2012, pp. 1318–1325.

[20] S. Alam, T. Reasat, R. Doha, and A. Humayun, (2020)“NumtaDB - Assembled Bengali Handwritten Digits. [online] arXiv.org. Available at: https://arxiv.org/abs/1806.02452.

[21] Data.mendeley.com. (2020). BanglaLekha-Isolated. [online] Available at: https://data.mendeley.com/datasets/hf6sf8zrkc/2 [Accessed 29 May 2020].

[22] V. Nair, and G. E. Hinton. "Rectified linear units improve restricted boltzmann machines." In Proceedings of the 27th international conference on machine learning (ICML-10), pp. 807-814.

[23] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. "Dropout: a simple way to prevent neural networks from overfitting." The Journal of Machine Learning Research vol 15, no. 1, pp 1929-1958, 2014.

[24] S. Ioffe, and C. Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." arXiv preprint arXiv:1502.03167, 2015.

[25] D. P. Kingma, and J. Ba. “Adam: A Method for Stochastic Optimization.”, [online] arXiv.org. Available at: https://arxiv.org/abs/1412.6980 [Accessed 29 May 2020].