2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Bangla Handwritten Digit Recognition Using Autoencoder and Deep Convolutional Neural Network

Md Shopon
Department of Computer Science and Engineering
University of Asia Pacific
Email: shopon.uap@gmail.com

Nabeel Mohammed
Department of Computer Science and Engineering
University of Liberal Arts Bangladesh
Email: nabeel.mohammed@ulab.edu.bd

Md Anowarul Abedin
Department of Computer Science and Engineering
University of Liberal Arts Bangladesh
Email: anowarul.abedin@ulab.edu.bd

***Abstract*—Handwritten digit recognition is a typical image classification problem. Convolutional neural networks, also known as ConvNets, are powerful classification models for such tasks. As different languages have different styles and shapes of their numeral digits, accuracy rates of the models vary from each other and from language to language. However, unsupervised pre-training in such situation has shown improved accuracy for classification tasks, though no such work has been found for Bangla digit recognition. This paper presents the use of unsupervised pre-training using autoencoder with deep ConvNet in order to recognize handwritten Bangla digits, i.e., 0 - 9. The datasets that are used in this paper are CMATERDB 3.1.1 and a dataset published by the Indian Statistical Institute (ISI). This paper studies four different combinations of these two datasets - two experiments are done against their own training and testing images, other two experiments are done cross validating the datasets. In one of these four experiments, the proposed approach achieves 99.50% accuracy, which is so far the best for recognizing handwritten Bangla digits. The ConvNet model is trained with 19,313 images of ISI handwritten character dataset and tested with images of CMATERDB dataset.**

***Index Terms*—Autoencoder; Deep Convolutional Neural Network; Handwritten Digit Recognition; Image classification; Supervised Learning; Unsupervised pre-training.**

I. INTRODUCTION

This paper primarily aims at recognizing handwritten Bangla numeral digits. Bangla is the mother language of Bangladesh and the 7th most widely spoken language in the world [1]. There are more than 200 million native Bangla speakers. It is the official language of Bangladesh and several Indian states including West Bengal, Tripura, Assam and Jharkhand [2]. As application of optical character recognition (OCR) is widespread in these regions, recognizing handwritten Bangla digits is becoming more important [3].

Recognizing Bangla handwritten digits is more challenging than recognizing English digits because of their critical shapes and varied sizes. However, this is a classical image classification problem and datasets are important for this purpose. This constitutes another major challenge to the recognition of Bangla digits compared to that of English digits; there are very few datasets available in Bangla. CMATERDB 3.1.1 and ISI handwritten character dataset [4] are the most notable among them. The largest dataset of English handwritten digits is MNIST consisting 60000 images [5], whereas ISI, the largest Bangla dataset, has only 23299 images [6]. The best accuracy achieved on MNIST dataset so far is 99.79% [7], which is very close to how human would recognize. However, such excellent result is unprecedented in Bangla digit recognition.

Many research works were conducted on Bangla handwritten digits recognition [8] - [13]. C. L. Liu and C. Y. Suen proposed a benchmarked model with a reported accuracy rate of 99.4%. They worked on the ISI numeral dataset using advanced normalization techniques and gradient based feature extraction [6]. In [10], a hierarchical bayesian network was proposed to classify the images with an accuracy of 87.5%. This model was based on George and Hawkins' original implementation stated in [14]. A quad tree based feature set was used in [12]; the accuracy rate of this model was 93.38%. In [11], sparse representation classifier produced an accuracy rate of 94%. [13] worked with local binary pattern for classifying Bangla digits resulting an accuracy of 96.7%.

Neural networks are very versatile for the classification problems. [4] is one of the initial studies employing end-to-end neural network training in this context. A Le-net like architecture was used in [4] to classify Bangla digits of ISI numeral dataset. They achieved an accuracy rate of 98.375% on the test set. Recently, [15] used similar neural network on the ISI dataset. As it is a comparatively small dataset, they augmented the training images by rotating each image by (5°, 10°, 20° and 30°). However, their augmentation process was almost equivalent to hand fitting the images, as the change in rotation angles was done individually and then tested for validation accuracy. The best accuracy 98.98% was achieved only when the images were rotated by 10°. By doing so, the images were fine-tuned in such a way that the model learns the specific features which were applicable to the test set.

Convolutional neural networks (CNN) are well recognized models for handwritten digit recognition. [16] surveys different such approaches, but all of them are for English digits recognition [17]. The result discussed in [7] was also

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

achieved using ConvNet. For Bangla digit recognition, this study finds that [15] is the only work that uses ConvNet for supervised learning. However, unsupervised pre-training along with supervised ConvNets has not been tried in any Bangla digit recognition work so far.

Autoencoders are used for unsupervised learning; An autoencoder has an input layer, a set of hidden layer and an output layer. In [19], denoising autoencoders were used for extracting features from images. In [20], content based image retrieval was done with the help of deep autoencoders. [21] presented a detailed explanation of the efficacy of unsupervised pre-training for supervised learning. [22] applied autoencoders as a pre-training method on the MNIST dataset and achieved very competitive error rates. No such attempt can be found for Bangla character recognition. Thus, this paper uses autoencoders and deep ConvNet for classifying images of two widely used standard datasets - CMATERDB and ISI numeral dataset are used for the experiments.

The rest of the paper is organized as follows: Section II contains the background on ConvNet and autoencoders. Section III introduces our proposed approach, details of our used dataset and necessary pre-processing. Section IV describes different experiments we did and the methods followed for them. Section V discusses our results and the impact. Finally, conclusions are made in Section VI.

II. Background

A. ConvNet

Convolutional neural network [23] or ConvNet is a special kind of ANN (Artificial neural network). ConvNets have learnable weights and biases. Just like ANN, ConvNets are trained with the backpropagation algorithm though they have a different architecture from traditional multi layer perceptrons. There are many advantages of using ConvNets as an image classification tool, mainly due to the fact that they can be used as a feature learner and data classifier simultaneously. [16] showed that ConvNet based approaches outperform other more traditional techniques, e.g., SVM, KNN etc.

MNIST is one of the most popular dataset for handwritten digit classification. The best result for this dataset so far is 99.79% which was done by regularizing the neural networks using dropconnect [7]. Other than classification ConvNets are used for many different purposes, such as face recognition, speech recognition, natural language processing etc [23]. LeNet5 is well known ConvNet architecture developed by Yann LeCun in 1998 [18]. Le-Net5 architecture was able to successfully classify digits and recognize the handwritten digits on bank cheques.

The first layer of a ConvNet is a Convolution layer with one or more kernels of fixed size. The input presented to these networks are usually images, and the kernels convolved with the input. The output of the convolution operations are passed through non-linearities, i.e., the Rectified Linear unit (ReLu) [24].

Pooling layers are also applied to outputs of the convolution operations. These are methods to reduce the data size, and can be done through averaging over a fixed region, choosing the maximum from a fixed region, or any other chosen method.

Dense layers are also known as fully connected layers which are used in the last stage of ConvNet. It connects the ConvNet to the output layer and constructs the expected number of outputs. For calculating the spatial dimensions of a ConvNet the following formula is used:

$$W_{out}(i) = 1 + \frac{W_{in} - R + 2P}{S}$$

Here i is the i^{th} input dimension, P is the value of padding, R_i is the receptive field and S is the value of stride.

B. Autoencoder

The functionality of an autoencoder is simply to output the input which it receives, at times, a transformed version of the input. This is not a difficult operation under normal circumstances. The utility of autoencoders derives from the fact that the input data is first transformed into a low dimensional representation (the encoding part). This low dimensional representation is then used to regenerate the input (the decoding part). This scheme enables autoencoders to learn useful representations of the input data without needing any labels a-priori.

Autoencoder can be implemented using various types of neural networks, e.g., ANN, ConvNet etc., depending on the desired representation scheme. When used as a pre-training method, the autoencoder is first trained in an unsupervised manner, enabling the encoding part of the autoencoder to adjust its weights to output useful low dimensional representations of the data. These weights are then combined with further layers and trained in a supervised manner. This approach has yielded good results for multiple classification tasks [21].

III. Proposed Approach and Dataset

A. Proposed Approach

As we discussed in Section I and II, results of recent studies on image classification problems using convolutional neural networks are very promising; traditional MLP is not sufficient for this purpose. So we propose to use Deep ConvNet that consists of more than one hidden layer. Figure 1 shows the proposed approach following a discussion about its parts.

The encoder consists of 3 convolutional layers, each followed by a 2×2 max pooling layer. All three layers of the encoder has 32 3×3 kernels. In between the layers dropout of 25% was used to reduce overfitting. The decoder has a similar architecture with each convolutional layer having 5 neurons, instead of 32. All the layers have RelU activation. Figure 2 shows the architecture of the autoencoder.

Figure 3 shows us the architecture of the convolutional model, where the first two convolutional layers are taken from the encoding part of the autoencoder, thus leveraging the weights learnt during unsupervised pre-training.

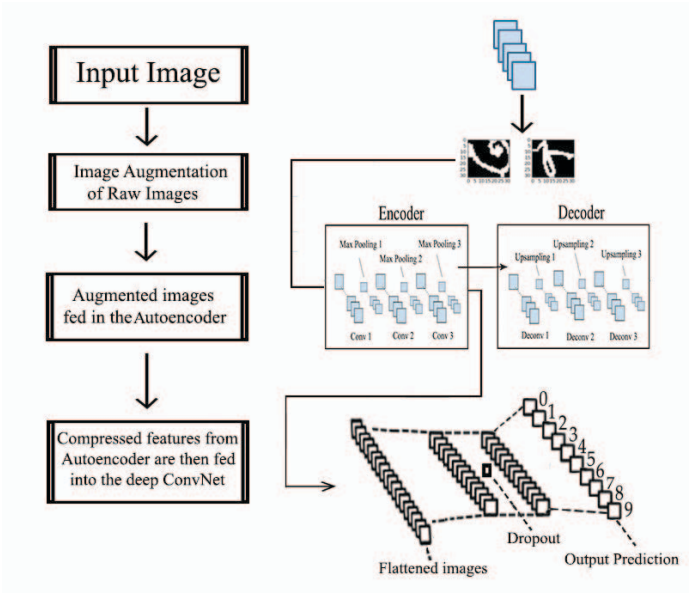

Fig. 1: Diagram of Our Proposed Approach

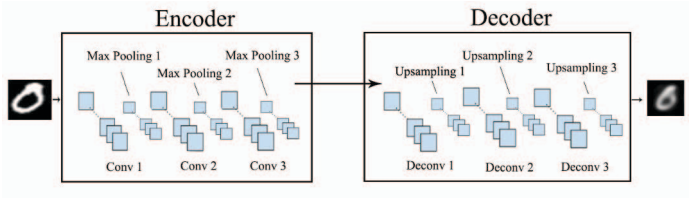

Fig. 2: Architecture of the Autoencoder

B. Dataset and Preprocessing

Deep learning is a data-driven field and requires large data sets. We worked with two datasets with our proposed approach - CMATERDB and ISI dataset. CMATERDB consists a total of 6000 images. We divided the images into 4200 and 1800 for training and testing purpose. The ISI numeral dataset consists a total of 23,299 images. The dataset was divided into 19,313 and 3986 for training and testing purpose respectively.

Images of CMATERDB are 32×32 pixels each. The images were preprocessed before feature extraction in order to normalize the features. They were first converted into grayscale images. The images initially had white(0) background and black(255) foreground. They were converted into black background and white foreground for our experiments.

ISI numeral dataset also had the same properties except their image pixels was arbitrary. They were first reshaped into 32x32 pixels. After that, the same operations were done against this dataset just as done in CMATERDB. Figure 4 Shows the images from both the dataset after pre-processing.

To enhance the training sets, each image was randomly rotated between $0°$ and $50°$. Each image was also shifted vertically by a random amount between 0 and 6 pixels. Horizontal shifts were also done in a similar range. These augmentations increased the size of the training sets significantly, allowing for reduced risk of overfitting. Figure 5 shows some samples

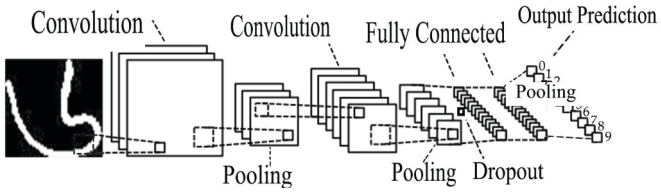

Fig. 3: Architecture of ConvNet Model

Bengali Digit	English Digit	CMATERDB	ISI Numeral
১	1		
২	2		
৩	3		
৪	4		
৫	5		
৬	6		
৭	7		
৮	8		
৯	9		
০	0		

Fig. 4: Digits for CMATERDB and ISI Numerals

of the augmented images

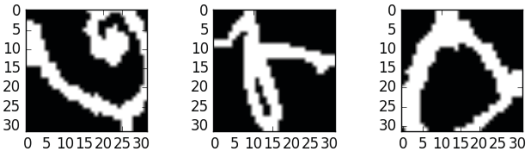

Fig. 5: Augmented images

IV. Experiments

The proposed model was trained in three different configurations: SCM stands for the first model we made which is Simple Convolutional Model, SCMA stands for Simple Convolutional Model with Augmented images and ACMA stands for Autoencoder with Convolutional Model with Augmented images.

All the three configurations were trained and tested against CMATERDB and ISI numeral dataset, and cross-validated with each other. The parameter size is 81053 for both SCM and SCMA; for ACMA, it is 85770.

The autoencoder was trained for 40 epochs with the RMSProp optimiser using binary cross entropy as the cost function. It is also possible to use mean-squared-error as the loss function, however that was not attempted in this study. After the autoencoder is trained, the learnt weights are reused in the ConvNet used for supervised learning. The layers were not frozen at any stage and training continued to update the weights from the first epoch of supervised training. Supervised training was done for 120 epochs, with the RMSProp optimiser using categorical-cross-entropy as the loss function.

V. Results and Analysis

For both datasets, ACMA gave the best accuracy rates, even in cross-dataset testing, demonstrarting the efficacy of unsupervised pre-training for this purpose.

Table I and II show the previously published results on the CMATERDB and ISI Character datasets. Table III summarises the accuracy rates obtained in different experiments. For the ISI character data set, the results achieved by our ACMA configuration is competetive at 98.29% accuracy. Infact, among all the ConvNet-based studies, our reported result is second only to that of [15], and we already have discussed the flaw in their approach in Section I. When an ACMA model is trained on the much smaller CMATERDB and tested on the ISI dataset, the results are still an impressive 97.29%, which is comparable to some previously published results shown in Table II. To reinforce the usefulness of supervised pre-training, we can see that ACMA gives an improvement of $.47\%$ over SCMA with the same dataset augmentation in place.

TABLE I: PAST RESULTS ON CMATERDB

Work	Accuracy
Haider Adnan Khan et al. [11]	94%
Basu et al. [25]	95.1%
Hassan et al. [13]	96.7%
Basu et al. [26]	97.15%
Sarkhel el at. [27]	98.23%
Das et al. [28]	98.55%

TABLE II: PAST RESULTS ON ISI

Work	Accuracy
Nasir and Uddin [29]	96.80%
Wen and He [30]	96.91%
Das et al. [31]	97.70%
Akhnad et al. [32]	97.93%
Bhattacharya and Chaudhuri [4]	98.20%
CNNAP [15]	98.98%

When ACMA is trained on the CMATERDB and tested on CMATERDB, the accuracy rate achieved is 98.61%, which is better than the previously published results on this dataset as shown in Table II. The accuracy rate of 99.50% on CMATERDB is obtained when ACMA is trained on the ISI dataset. As far as we could ascertain, this is the best reported result on the dataset, even better than the 98.61% accuracy achieved in this study. This is particularly interesting because the characteristics of the images in the two datasets are quite different, with the ISI images being smoother and less blocky compared to the CMATERDB images. This leads us to conclude that while the autoencoder used in this study was trained to reproduce images of each dataset individually, the representations learnt by this method generalise across different types of images.

TABLE III: ACCURACY IN DIFFERENT EXPERIMENTS

Dataset	SCM	SCMA	ACMA
Train:CMATERDB, Test:CMATERDB	96.59%	97.29%	98.61%
Train:ISI, Test:ISI	97.02%	97.43%	98.29%
Train:CMATERDB, Test:ISI	94.69%	96.82%	97.29%
TRAIN:ISI, TEST:CMATERDB	98.26%	98.76%	99.50%

VI. Conclusion

This paper presents the implementation of unsupervised pre-training, using an autoencoder, as a pre-cursor to supervised training for Bangla handwritten digit recognition. To demonstrate the effectiveness of this approach, we tested with three different training configurations across two different standard Bangla character datasets. For the test images of the ISI numeral dataset, the proposed approach achieves accuracy rates comparable with previous works. For the CMATERDB, it achieves an accuracy rate of 99.50%, which is the best reported result on this dataset so far. Apart from demonstrating the utility of unsupervised pre-training in the context of Bangla digit recognition, our results also indicate that such pre-training can be useful even when the data sets are independently (and blindly) collected. For every experiment, the model with autoencoder and ConvNet gives better accuracy rate than the model with only ConvNet. The proposed approach achieves state-of-the art results for CMATERDB and very good results for the ISI dataset. It is worth mentioning that previous studies on Bangla handwritten digit recognition rarely reported on these two datasets together. Future studies can explore whether pre-training on larger datasets, which are not specific to Bangla, can help achieve better results to be practically useful.

Acknowledgment

This project is supported by the ICT Division, Ministry of ICT, Government of the People's Republic of Bangladesh (Project ID: 56.00.0000.028.33.066.16-731)

References

[1] Singh, V.K., "Most Spoken Languages in the world", 2012. [Online]. Available: goo.gl/fhTq2S. [Accessed: 20- Oct- 2016].

[2] Chatterji, S. K., "The origin and development of the Bengali language", 2002.

[3] Pal, U., and Chaudhuri, B. B., "OCR in Bangla: an Indo-Bangladeshi language." In Pattern Recognition, 1994. Vol. 2-Conference B: Computer Vision and Image Processing., Proceedings of the 12th IAPR International. Conference on, vol. 2, pp. 269-273. IEEE, 1994.

[4] Bhattacharya, U., and Chaudhuri, B. B., Handwritten numeral databases of indian scripts and multistage recognition of mixed numerals, In IEEE transactions on pattern analysis and machine intelligence, 31(3) on pp.444-457. IEEE,2009.

[5] LeCun, Y., Cortes, C., and Burges, C. J., "The MNIST database of handwritten digits.", 1998. APA

[6] Liu, C. L., and Suen, C. Y., A new benchmark on the recognition of handwritten Bangla and Farshi numeral characters, In Proc. 11 th ICFHR, 2008.

[7] Wan, L., Zeiler, M., Zhang, S., Cun, Y. L., and Fergus, R. , Regularization of neural networks using dropconnect, In Proceedings of the 30th International Conference on Machine Learning (ICML-13), pp. 1058-1066. 2013.

[8] Hossain M.Z., Amin M.A. and Yan H., "Rapid feature extraction for BangIa handwritten digit recognition", In Machine Learning and Cybernetics (ICMLC), 2011 International Conference on, vol. 4, pp. 1832-1837. IEEE, 2011.

[9] Sazal, M. M. R., Biswas, S. K., Amin, M. F., and Murase, K., Bangla Handwritten Character Recognition using Deep Belief Network, In Electrical Information and Communication Technology (EICT), 2013 International Conference on, pp. 1-5. IEEE, 2014. December 2013.

[10] Xu, J. W., Xu, J., and Lu, Y, Handwritten Bangla digit recognition using hierarchical Bayesian network, In Intelligent System and Knowledge Engineering, 2008. ISKE 2008. 3rd International Conference on, vol. 1, pp. 1096-1099. IEEE, 2008.

[11] Khan, H. A., Al Helal, A., and Ahmed, K. I., "Handwritten Bangla digit recognition using Sparse Representation Classifier," In Informatics, Electronics and Vision (ICIEV), 2014 International Conference on, pp. 1-6. IEEE, 2014.

[12] Roy, A., Mazumder, N., Das, N., Sarkar, R., Basu, S., and Nasipuri, M.,"A new quad tree based feature set for recognition of handwritten bangla numerals." In Engineering Education: Innovative Practices and Future Trends (AICERA), 2012 IEEE International Conference on, pp. 1-6. IEEE, 2012.

[13] Hassan, T., and Khan, H. A. "Handwritten Bangla numeral recognition using Local Binary Pattern," In Electrical Engineering and Information Communication Technology (ICEEICT), 2015 International Conference on, pp. 1-4. IEEE, 2015.

[14] George, D., and Hawkins, J. A Hierarchical Bayesian model of Invariant Pattern Recognition in the Visual Cortex, In Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005., vol. 3, pp. 1812-1817. IEEE, 2005.

[15] Akhand, M.I.H.R.I. M.A.H., Ahmed, M., Convolutional neural network training with artificial pattern for bangla handwritten numeral recognition, ICIEV, vol. 1, no. 1, pp. 16, 2016.

[16] Kamavisdar, P., Saluja, S., and Agrawal, S. "A Survey on Image Classification Approaches and Techniques", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 1, pp. 1005 1009, January 2013

[17] Bottou, L., Cortes, C., Denker, J. S., Drucker, H., Guyon, I., Jackel, L. D., and Vapnik, V., "Comparison of classifier methods: a case study in handwritten digit recognition." In International conference on pattern recognition, pp. 77-77. IEEE Computer Society Press, 1994.

[18] LeCun, Y., "LeNet-5, convolutional neural networks.", 2015. [Online] Available: http://yann.lecun.com/exdb/lenet. [Accessed: 20- Oct- 2016]

[19] Vincent, P., Larochelle, H., Bengio, Y., and Manzagol, P. A., "Extracting and composing robust features with denoising autoencoders." In Proceedings of the 25th international conference on Machine learning, pp. 1096-1103. ACM, 2008.

[20] Krizhevsky, A., and Hinton, G. E., "Using very deep autoencoders for content-based image retrieval." In ESANN. 2011.

[21] Erhan, D., Bengio, Y., Courville, A., Manzagol, P. A., Vincent, P., and Bengio, S., "Why does unsupervised pre-training help deep learning?," Journal of Machine Learning Research,pp. 625-660, February 2010.

[22] Huang, F. J., Boureau, Y. L., and LeCun, Y., "Unsupervised Learning of Invariant Feature Hierarchies with Applications to Object Recognition," In 2007 IEEE conference on computer vision and pattern recognition, pp. 1-8. IEEE, 2007.

[23] LeCun, Y., and Bengio, Y. "Convolutional networks for images, speech, and time series." The handbook of brain theory and neural networks 3361.10 (1995): 1995.

[24] Nair, V., and Hinton, G. E. "Rectified linear units improve restricted boltzmann machines." Proceedings of the 27th International Conference on Machine Learning (ICML-10). 2010.

[25] Basu, S., Sarkar, R., Das, N., Kundu, M., Nasipuri, M., and Basu, D. K., Handwritten bangla digit recognition using classifier combination through ds technique, In International Conference on Pattern Recognition and Machine Intelligence, pp. 236-241. Springer Berlin Heidelberg, 2005.

[26] Basu, S., Das, N., Sarkar, R., Kundu, M., Nasipuri, M., and Basu, D. K. , A novel framework for automatic sorting of postal documents with multi-script address blocks, Pattern Recognition, vol. 43, pp. 3507- 3521, 2010.

[27] Sarkhel, R., Das, N., Saha, A. K., and Nasipuri, M., A multi-objective approach towards cost effective isolated handwritten bangla character and digit recognition, Pattern Recognition, vol. 58, pp. 172189, 2016

[28] Das, N., Reddy, J. M., Sarkar, R., Basu, S., Kundu, M., Nasipuri, M., and Basu, D. K., A statisticaltopological feature combination for recognition of handwritten numerals, Applied Soft Computing, vol. 12, no. 8, pp. 24862495, 2012.

[29] Nasir, M. K., and Uddin, M. S., Hand written bangla numerals recognition for automated postal system, IOSR Journal of Computer Engineering, vol. 8, no. 6, pp. 4348, 2013.

[30] Wen, Y., and He, L. , A classifier for bangla handwritten numeral recognition, Expert Systems with Applications, vol. 39, no. 1, pp. 948 953, 2012.

[31] Das, N., Sarkar, R., Basu, S., Kundu, M., Nasipuri, M., and Basu, D. K., A genetic algorithm based region sampling for selection of local features in handwritten digit recognition application, Applied Soft Computing, vol. 12, no. 5, pp. 15921606, 2012.

[32] Rahman, M. M., Akhand, M. A. H., Islam, S., Shill, P. C., and Rahman, M. H., Bangla handwritten character recognition using convolutional neural network, I.J.Image, Graphics and Signal Processing(IJIGSP, vol. 7, no. 3, pp. 4249, 2015.