Performance Analysis on Bangla Handwritten Digit Recognition using CNN and Transfer Learning

Afsana Hossain
Department of Computer Science and Engineering, East West University, Dhaka Bangladesh
Email: afsanahossain65@gmail.com
Md. Sabbir Hasan
Department of Computer Science and Engineering, East West University, Dhaka Bangladesh
Email : sabbirhasananik@gmail.com
Md. Mujtaba Asif
Department of Computer Science and Engineering, East West University, Dhaka Bangladesh
Email :mjtbasif@gmail.com
Amit Kumar Das
Department of Computer Science and Engineering, East West University, Dhaka Bangladesh
Email :amit.csedu@gmail.com

ABSTRACT

A computer can read and interpret intelligible handwritten input from sources like paper, photos, and other devices, known as Handwriting recognition (HWR). Besides, handwritten recognition is an interesting challenge in machine learning and deep learning. Because several strategies and approaches have been followed already to solve this challenge, machine learning and deep learning provided the best results.Handwritten digit recognition is a part of HWR. It is getting popular day by day because many applications could be made using this system like OCR, postal code recognition, license plate recognition, bank checks recognition, etc. Besides, the importance of recognizing the Bangla digit from the document is increasing. But the works available in Bangla handwritten digit recognition are very few.

Similarly, none of them are robust, and some of them are overfitted. Therefore, we need to make some improvements to this system considering its importance. This paper explores the presentation of transfer learning with the help of some best-in-class profound CNN strategies for the acknowledgment of manually written Bangla digits. It considers two deep CNN architectures, such as Mobile Net and Residual Network (Reset) based on performance and accuracy. This model was trained and tested with the CMATERdb dataset. The study suggests that transfer learning provides 97% accurate results, where traditional CNN provides 86-92%.

Keywords - **BangleHandwritten Digit, CNN, MobileNet, ResNet50, Transfer learning.**

Date of Submission: June 29, 2021 Date of Acceptance: July 26, 2021

I. INTRODUCTION

Our study has tried different strategies to recognize Bangla's handwritten digit recognition accurately and efficiently. Taking handwritten digits as input and recognize it is the main objective of a digit recognition system. This system's importance is increasing in our country as many applications could be made using it, such as automatic license plate recognition for vehicles, national id card recognition, postal code recognition on envelopes, OCR, amount recognition on bank cheques, and many more. OCR or Optical Character Recognition is one of the essential applications, and it is used to convert various documents, like scanned paper, pdf files, also images. The camera captures us into editable data. Besides, bankers, teachers, bankers will be benefited from this digit recognition system because the number of bank checks can be recognized automatically. This system can be used in our online education system too. The teacher will check and edit the online exam scripts if needed by the character recognition system, which can be implemented with the strategies, we have followed to build the digit recognition. This system can be used for making advanced map applications, like capture a picture, then extract the street no using OCR. And much more applications can be implemented with this recognition system. So, we can see this digit recognition system can make a significant contribution by helping the people of different fields of our society. Therefore, we have tried to develop a better Bangla handwritten digit recognition.Bangla is one of the most important languages around the globe. More than 250 million people speak in Bangla, and it's the 7th most spoken language internationally. Bangla is the primary language in Bangladesh, and also it is used as a primary or secondary language in the Indian subcontinent. There are ten unique characters in Bangla digits. But recognizing handwritten digits is much more complicated than recognizing typed digits. Because writing style varies from person to person, and also there are some similarities between different digit's shapes.

Volume: 13 Issue: 01 Pages:4809-4815(2021) ISSN: 0975-0290

Table 1: Example of Bangla Digits

০	১	২	৩	৪	৫	৬	৭	৮	৯
0	1	2	3	4	5	6	7	8	9

To recognize Bangla, handwritten digits have used traditional CNN and two deep CNN architectures (MobileNet and ResNet50) along with transfer learning. We have used CNN because it is preferable for image recognition and transfer learning with CNN architecture to a better result with less computational power. Though there are many other ways, two build this digit recognition system, water reservoir scheme,deep CNN with HOG features, autoencoder, some other CNN architecture (AlexNet, CapsuleNet), etc.

II. BACKGROUND ANALYSIS AND RELATED WORK

We have used TensorFlow and two deep CNN architectures (MobileNet and ResNet50) for our study. TensorFlow is a python library, and it is used for fast numerical computation [13] [14] [15][16]. On the other hand, MobileNet is used for image classification, and its specialty is that it takes less computational power when applying transfer learning compared to other models. ResNet is a 50-layer deep convolutional neural network, and we can load pertained versions of the network trained on more than a million images from the ImageNet database [17][18][19].

Several approaches and strategies have been made to solve the digit recognition problem. In 2000 Pal and Chaudhury made some attempts to solve digit recognition problems for Bangla numerals using the water reservoir concept, dependent on the extracted features. T. Hassan and A.H. Khan used the Local Binary Pattern (LBP) approach in three different schemas. LBP has been generally used for face recognition. The writers of the paper used the K-nearest neighbor classifier to classify characters. This paper proposed an OCR system on a dataset from the CMATERdb3.1.1 database, which accurately recognized 96.7 % of characters [1]. Researcher U. Bhattacharya of this paper worked on Devanagari handwritten database. This database has 22,556 data points collected from 1049 individuals. The researcher also used Bangla handwritten numerical database with 23,392 samples collected from 1106 individuals. He had chosen the nearest neighbor classifier to do his research and got results for k = 1, 3, 5, 7, 9, and 15. In the case of resizing, he used Daubechies wavelet filters to classify the images, and he used the Multistage Recognition System. In the Devanagari dataset, he trained 1,67,940 images and got 99.27% accuracy, and for validation, 20,000 data points were used, and he got 99.02% accuracy. In the Bangla dataset, he used 173920 images for training and 20,000 images for validation.

The model gave his 99.14% and 98.20% accuracy, respectively. [2]. The researchers introduced deep CNN for handwritten digit recognition to improve accuracy and for better-supervised learning. They have used the NumtaDB database to do their research and to preprocess images. They did resize and gray-scaling, interpolation, removing blur from the image, sharpening images. They build a deep neural network model, which gives a testing accuracy of 92.72% and training accuracy of 99.59% [3]. In this paper, writers used a deep CNN model to recognize Hindi's handwritten digits. They constructed their model using a convolutional layer, Maxpooling2D layer, flatten layer. They used 20,000 images from Kaggle to conduct their research. Their CNN model used root mean square propagation optimizer and 99.85% accuracy they achieved using the model [4]. P V Bhagyasree, Ajay James, Chandran Saravanan proposed a new technique to recognize English handwritten characters. To conduct their research, they used a convolutional network as well as a deep neural network. Their unique approach was Directed Acyclic Graph-Convolutional Neural Network. They did not implement their proposed model yet. In CNN, the output value of a layer passes to the next layer, and because of the vanishing gradient, the accuracy goes down as the network goes deeper. In DAG-CNN, the whole CNN model will get some subdivided ways using some directed edges [5]. Here is another research on this aspect, but their digit was Arabic, and researchers used 46,000 Arabic digits collected from 840 people. In preprocessing, they revised their image into 64 x 64 pixels. In the final layer of CNN, they used the Softmax activation function to minimize the error. As they used GPUs to fasten in the building model, they used 10,000 epochs to get a superior result of 95.7% [6].

This paper's researchers used the CMATERdb dataset and tried some combinational approach of deep learning and traditional algorithms to build a Bangla handwritten digit recognizer. Their applied techniques are SVM, deep belief network, CNN + Gaussian, CNN + Gabor, CNN + Gaussian + Dropout, CNN + Gabor + Dropout. The highest accuracy they got using the CNN + Gabor + Dropout approach was 98.78% [12].

III. DATASET

We have used the CMATERdb 3.1.1 database for our study. It is a popular dataset and preferred by many researchers for the handwritten digit recognition challenge. It is handwritten and a balanced dataset of a total of 6000 Bangla numerals (32x32 RGB colored, 6000 images), each having 600 images per class (per digit) [7]. In the CMATERdb, the Bangla handwritten dataset ids are level from 0 to 9. We have divided the data set for training and testing to avoid the same subject in both segments. This database has created in the Jadavpur University research lab in Kolkata.

IV. METHODOLOGY

Modern technology provides a vast number of multimedia devices, simulation software, network technologies. People use them rapidly and are increasing amounts of information in the forms of images. This massive number of images needs to organize effectively. These images

Int. J. Advanced Networking and Applications 4811
Volume: 13 Issue: 01 Pages:4809-4815(2021) ISSN: 0975-0290

contain several complex and logical information, so they needed to be labeled accurately. In general, the images are labeled in multiple labels to classify, predict, and group them. Some traditional supervised algorithm helps to retrieve and process the images, but now, in the recent year, researchers are more interested in deep learning to solve the problems

1.1 CONVOLUTION NEURAL NETWORK (CNN)

In the 1960s, Hubel and Wiesel [8] first proposed the Convolution neural network during a study of neurons in monkey cortexes. CNN's work by sharing weights and extracting some essential features from the images. CNN's take images as input, process them, and classify them under specific categories. For image resolution computer considers an image as an array of pixels, and it sees h× w × d [height × weight × dimension]. In the RGB value, dimension assigns to 3, whereas in grayscale, it assigns to 1. Each image passes through a series of layers during training and testing. The combination of a convolutional layer, pooled layer, and fully connected layer builds the CNNs. In the output layer, the Softmax activation function is applied to classify objects with probabilistic values 0 to 1 [9][20][21][22].

The convolution layer works for extracting features from an input image, and in this layer "relu" activation function works well. It takes two inputs, a matrix of pixels of the input image and a filter or kernel. The mathematics behind this is –

- An image dimension ($h \times w \times d$)
- A filter ($f_h \times f_w \times d$)
- Output dimension ($h - f_h + 1) \times (w - f_w + 1) \times 1$

Rectified Linear Unit (ReLU) uses for non-linearity in our ConvoNet. ConvoNet should learn the non-negative linear value, and that's what ReLU does [2][23][24][25].

The pooling layer reduces the number of parameters in the case of a large image. Spatial pooling reduces the dimensionality of each map, which is also known as subsampling or downsampling. Different types of pooling can be:

- Max pooling
- Average pooling
- Sum pooling

After pooling the feature, the matrix is converted into vectors [x1, x2, x3, ….] and fed it to the fully connected layer. Finally, as output, it becomes a complete model able to classify the images [2].

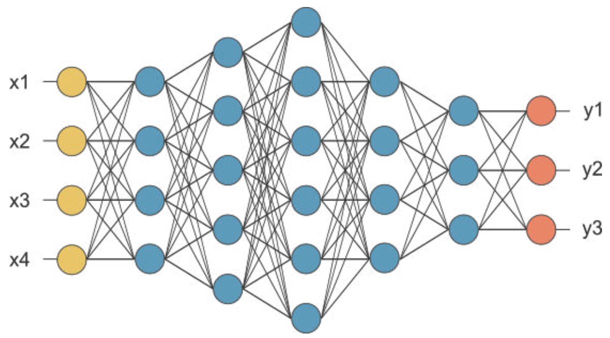

Fig 1:Flattened as the fully connected layer

1.2 RESIDUAL NETWORK (RESNET)

CNN has different types of architecture that participate in the ImageNet challenge, and one of them is ResNet, which stands for Residual Network. A classic neural network works behind it. This model won first place in the ILSVRC 2015 competition, competing with a top-5 error rate of 3.57%. The model that helps to train with 150+ layers deep neural network is developed byKaiming [3].

This model has a concept called skip connections. With the ReLU activation function in classic CNN, the input matrix calculates the linear transformation one after another. Still, in RestNet it skips the first transformation and directly passes the input matrix to the second transformation output, and finally, all the outputs sum up in the final ReLU function [10].

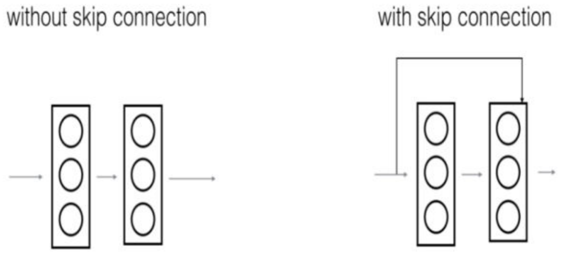

Fig 2: Skip connection concept of ResNet architecture comparing with plain.

1.3 MOBILE NET

A CNN class is MobileNet that Google publicly released, and hence, this gives us a fantastic beginning stage for preparing our classifiers that are madly little and madly quick. This model is designed for a mobile application that uses depth-wise separable convolution. The model is used to do transfer learning. This ImageNet classification model is defined to meet the resource limitations of different cases [11].

In our research, we have used an end-to-end machine learning open-source platform, TensorFlow. It has several libraries and resources that help developers build a model in a short time.

Volume: 13 Issue: 01 Pages:4809-4815(2021) ISSN: 0975-0290

1)Preprocessing*:* This step holds the most crucial part during the research as a result, it mostly depends on the preprocessing of the dataset. We did resize the dataset and then grayscaled them, and finally sharpen them.

a) resizing*:* Though the sizes of images in the database were mentioned as 32x32 pixels, we took a little initiative to confirm the image size.

b) grascaling*:* The images were in RGB scaling, so we needed to convert the images into Gray scaling. We have changed the color channel also to 1 so that our model can be built properly. Initially, the color channel was 3, which was a restriction.

b) normalizing*:* To make it computationally efficient, we need to reduced the grayscaled values by 0 to 1. It uses a pixel intensity value.

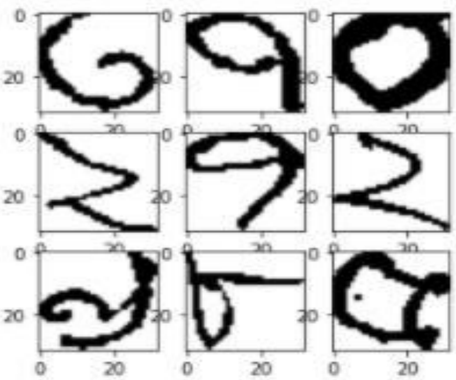

Fig 3: Random image before preprocessing

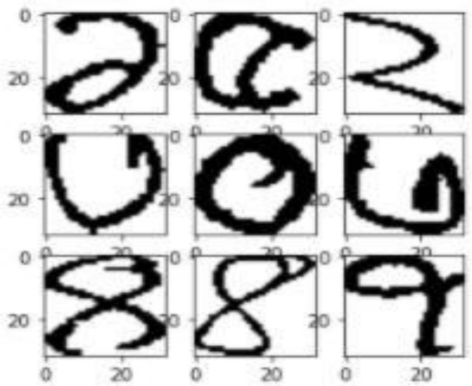

Fig 4: Random image after Gray scaling

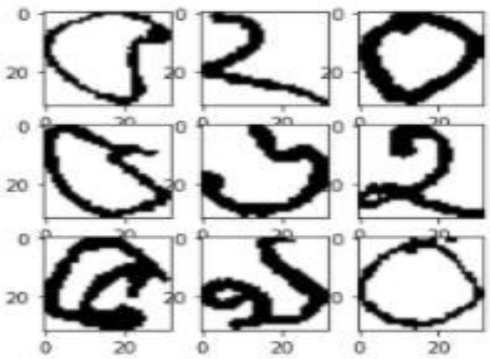

Fig 5: Random image after a normalization

The whole process of our work-

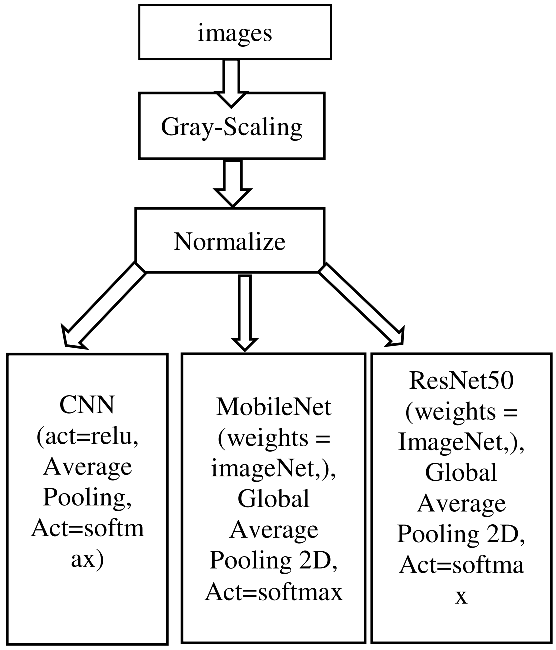

Table 2: A short summary of three models

Model Name	Total parameters	Trainable parameters	Non-trainable parameters
CNN	61,706	61,706	0
MobileNet	3,228,864	3206976	21888
ResNet50	23,587,712	23,534,592	53,120

V. RESULTS

We have used 83% of the training dataset and 17% of the dataset for testing, 1000 data points, and ten classes. As our PC configuration was not much good and the experiments typically take so much time to train a model, we set epochs 5. We first trained the model with traditional CNN and found that the model predicts Bengali handwritten digits with 91% accuracy. Figure 6 shows that for digits 1,3, 6, and 9, recall and f1-score are less than 90%.

It means the model correctly identifies the digits less than 90% of the actual digits. Figure 6 clearly shows that our CNN model accurately trained up to 92% within four epochs and validation accuracy up to 91%. Figure 8 establishes evidence of training and validation loss. The validation loss is less than 25%, whereas training loss is an exact 25% within four

	precision	recall	f1-score	support
0	0.94	0.98	0.96	100
1	0.89	0.88	0.88	100
2	0.98	0.95	0.96	100
3	0.85	0.89	0.87	100
4	0.92	0.94	0.93	100
5	0.97	0.87	0.92	100
6	0.88	0.80	0.84	100
7	0.93	0.97	0.95	100
8	0.96	1.00	0.98	100
9	0.83	0.86	0.84	100
accuracy			0.91	1000
macro avg	0.91	0.91	0.91	1000
weighted avg	0.91	0.91	0.91	1000

Fig 6: Classification report of CNN model

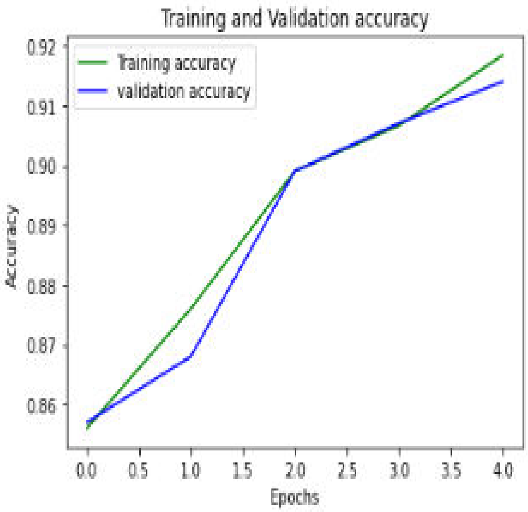

Fig 7: Training and validation accuracy of CNN model

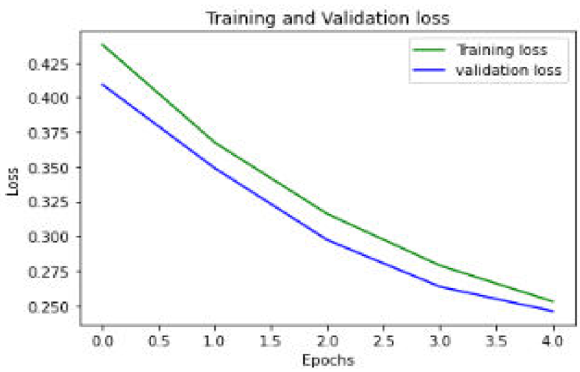

Fig 8: Training and validation loss of CNN model

	precision	recall	f1-score	support
0	1.00	0.28	0.44	100
1	0.60	0.90	0.72	100
2	0.94	0.99	0.97	100
3	0.94	0.97	0.96	100
4	0.84	1.00	0.91	100
5	0.92	0.97	0.94	100
6	0.99	0.95	0.97	100
7	1.00	0.82	0.90	100
8	1.00	0.91	0.95	100
9	0.82	0.99	0.90	100
accuracy			0.88	1000
macro avg	0.91	0.88	0.87	1000
weighted avg	0.91	0.88	0.87	1000

Fig 9: Classification report of MobileNet model

After CNN, we have built a MobileNet model to see the performance of our dataset. The model gives an accuracy of 88%, which is less than the CNN model. The precision of digit 1 is 60%, meaning the model can accurately identify the digit 60%. The recall of the digit 0 only 28% means the model can identify the digit 28% of all the 0 digits. As the precision and recall affect the f1-score, f1-score for these two digits is so low. All of this information can be seen in figure 9.

	precision	recall	f1-score	support
0	0.99	0.97	0.98	100
1	0.99	0.87	0.93	100
2	0.99	0.99	0.99	100
3	0.98	0.97	0.97	100
4	0.92	1.00	0.96	100
5	0.95	1.00	0.98	100
6	0.97	0.98	0.98	100
7	1.00	0.96	0.98	100
8	1.00	0.97	0.98	100
9	0.91	0.97	0.94	100
accuracy			0.97	1000
macro avg	0.97	0.97	0.97	1000
weighted avg	0.97	0.97	0.97	1000

Fig 10: Classification report of ResNet50 model

Finally, we have tried ResNet50 to see the performance on our dataset, and yet, we have got satisfied with the result, and it has given 97% accuracy, which is more convenient to predict our digit. Precision, recall, f1-score for all digits is above 90%, and this model predicts our digits above 90% accurately.

VI. CONCLUSION

Recognizing Bengali handwritten characters are becoming more and more important day by day. As mentioned before, systems like OCR can be highly benefited from it. This study proposed two different methods for recognizing Bengali handwritten digits using Convolutional Neural Network (CNN) and Transfer Learning with ResNet50 and MobileNet. As it suggests, the transfer learning approach produces a better result. The traditional CNN provides an accuracy of 91%, MobileNet 88%, and ResNet50 offers the best accuracy, 97%. We can conclude that conventional CNN is not enough to recognize handwritten digits, and we can use deep CNN architecture to build a robust digit recognition system.

There were a few limitations that we faced when we proceeded with this study. We have shortlisted the crucial elements. The dataset we used to train and implement our model was relatively small. A larger dataset would increase efficiency and accuracy for the test results. We used to run out test models with limited computation power; therefore, we had to resort to fewer models for optimum results. Our work has a vast field to contribute if worked on for an extended time, leading to more efficient outputs. The test can be implemented through different data models to regulate and find more efficient solutions among the models and define a better approach. A different dataset with a larger size will increase the accuracy of the results, which can be obtained from an existing source or be collected by personal research means. The dataset we used only contained numerals. This study can be extended further if we incorporate letters from the Bengali alphabet to change the dynamic and bring the research work into a broader perspective.

REFERENCES

[1] T. Hassan, and A.H. Khan, “Handwritten Bangla numeral recognitionusing Local Binary Pattern,” *In Electrical Engineering and InformationCommunication Technology (ICEEICT), 2015 InternationalConference on, pp. 1-4. IEEE, 2015,doi: 10.1109/ICEEICT.2015.7307371*

[2] U. Bhattacharya, and B. B Chaudhuri, “Handwritten numeral databasesof indian scripts and multistage recognition of mixed numerals”, *InIEEE transactions on pattern analysis and machine intelligence, 31(3)on pp.444-457. IEEE,2009.*

[3] A. Shawon, M. Jamil-Ur Rahman, F. Mahmud, and M. M. Arefin Zaman, "Bangla Handwritten Digit Recognition Using Deep CNN for Large and Unbiased Dataset," *2018 International Conference on Bangla Speech and Language Processing (ICBSLP), Sylhet, 2018, pp. 1-6, doi: 10.1109/ICBSLP.2018.8554900.*

[4] R. V. Kumar Reddy, B. Srinivasa Rao and K. P. Raju, "Handwritten Hindi Digits Recognition Using Convolutional Neural Network with RMSprop Optimization," *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2018, pp. 45-51, doi: 10.1109/ICCONS.2018.8662969.*

[5] P. V. Bhagyasree, A. James and C. Saravanan, "A Proposed Framework for Recognition of Handwritten Cursive English Characters using DAG-CNN," *2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT), CHENNAI, India, 2019, pp. 1-4, doi: 10.1109/ICIICT1.2019.8741412.*

[6] Al-Wzwazy, Haider. (2016). Handwritten Digit Recognition Using Convolutional Neural Networks. *International Journal of Innovative Research in Computer and Communication Engineering. 4.*

[7] Das, Nibaran & Sarkar, Ram &Basu, Subhadip & Kundu, Mahantapas&Nasipuri, Mita&Basu, Dipak. (2012). A genetic algorithm-based region sampling for selection of local features in handwritten digit recognition application. Applied Soft Computing. 12. 10.1016/j.asoc.2011.11.030.

[8] D. H. Hubel and T. N. Wiesel, “Receptive fields and functional architecture of monkey striate cortex,” *Journal of Physiology, vol. 195, no. 1, pp. 215–243, 1968.*

[9] "Understanding of Convolutional Neural Network (CNN) — Deep Learning", *Medium*, 2020. [Online]. Available: *https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148. [Accessed: 30- Sep- 2020*].

[10] "CNN Architectures — LeNet, AlexNet, VGG, GoogLeNet and ResNet", *Medium*, 2020. [*Online]. Available: https://medium.com/@RaghavPrabhu/cnn-architectures-lenet-alexnet-vgg-googlenet-and-resnet-7c81c017b848. [Accessed: 30- Sep- 2020].*

[11] A. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications", *arXiv.org, 2020. [Online]. Available: https://arxiv.org/abs/1704.04861. [Accessed: 30- Sep- 2020].*

[12] M. Alom, P. Sidike, T. Taha and V. Asari, "Handwritten Bangla Digit Recognition Using Deep Learning", *arXiv.org, 2020. [Online]. Available: https://arxiv.org/abs/1705.02680. [Accessed: 16- Oct- 2020].*

[13] O. F. Rakib, S. Akter, M. A. Khan, A. K. Das and K. M. Habibullah, "Bangla Word Prediction and Sentence Completion Using GRU: An Extended Version of RNN on Ngram Language Model," 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI), 24-25 December, Dhaka.

[14] E. A. Emon, S. Rahman, J. Banarjee, A. K. Das and T. Mittra, "A Deep Learning Approach to Detect Abusive Bengali Text," 2019 7th International Conference on Smart Computing & Communications (ICSCC), Sarawak, Malaysia, Malaysia, 2019.

[15] M. M. Hossain, M. F. Labib, A. S. Rifat, A. K. Das and M. Mukta, "Auto-correction of English to Bengali Transliteration System using Levenshtein Distance," 2019 7th International Conference on Smart Computing & Communications (ICSCC), Sarawak, Malaysia, Malaysia, 2019.

[16] M. D. Drovo, M. Chowdhury, S. I. Uday and A. K. Das, "Named Entity Recognition in Bengali Text Using Merged Hidden Markov Model and Rule Base Approach," 2019 7th International Conference on Smart Computing &Communications (ICSCC), Sarawak, Malaysia, Malaysia, 2019.

[17] E. Biswas and A. K. Das, "Symptom-Based Disease Detection System In Bengali Using Convolution Neural Network," 2019 7th International Conference on Smart Computing & Communications (ICSCC), Sarawak, Malaysia, Malaysia, 2019.

[18] A. K. Das, A. Ashrafi and M. Ahmmad, "Joint Cognition of Both Human and Machine for Predicting Criminal Punishment in Judicial System," 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), Singapore, 2019, pp. 36-40.

[19] R. A. Tuhin, B. K. Paul, F. Nawrine, M. Akter and A. K. Das, "An Automated System of Sentiment Analysis from Bangla Text using Supervised Learning Techniques," 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), Singapore, 2019, pp. 360-364.

[20] J. Islam, M. Mubassira, M. R. Islam and A. K. Das, "A Speech Recognition System for Bengali Language using Recurrent Neural Network," 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), Singapore, 2019, pp. 73-76.

[21] T. F. Mumu, I. J. Munni, and A. K. Das, “Depressed People Detection from Bangla Social Media Status using LSTM and CNN Approach”, J. eng. adv., vol. 2, no. 01, pp. 41-47, Mar. 2021.

[22] M. T. Hossain, M. W. Hasan and A. K. Das, "Bangla Handwritten Word Recognition System Using Convolutional Neural Network," 2021 15th International Conference on Ubiquitous Information Management and Communication (IMCOM), 2021, pp. 1-8.

[23] A. K. Das, A. A. Asif, A. Paul and M. N. Hossain, "Bangla hate speech detection on social media using attention-based recurrent neural network" Journal of Intelligent Systems 30, no. 1 (2021): 578-591.

[24] M. S. S.Khan, S. R. Rafa, and A. K. Das, "Sentiment Analysis on Bengali Facebook Comments To Predict Fan's Emotions Towards a Celebrity" Journal of Engineering Advancements,pp. 118-124, July 2021.

[25] M. Mubassira, and A. K. Das, “Implementation Of Recurrent Neural Network With Language Model For Automatic Articulation Identification System In Bangla”, International Journalof Advanced Networking and Applications - IJANA., vol. 12, no. 06, pp. 4800-4808, May-June, 2021.