Optical Odia Character Classification using CNN and Transfer Learning: A Deep Learning Approach

Rakesh Kumar Sethi[1], Kalyan Kumar Mohanty[2]

[1]*M.Tech, I & E department, College of Engineering and Technology, Bhubaneswar, India*
[2]*Center of Excellence on Artificial Intelligence,CET, Bhubaneswar, India*

---***---

Abstract - *OCR (optical character Recognition) is widely used in the electronic document analysis system. It belongs to the newline intersection of image processing, pattern recognition, and Natural Language Processing and it helps to extract texts from a document or an image. Though tremendous research efforts have been made, the state of the art in the newline OCR has in recent years only reached the point of partial use. Nowadays, the off-the-shelf OCR program can reliably recognize cleanly written text in documents with clear layouts. There is only modest progress in handwriting character recognition (HCR), and that too for a small vocabulary for new line isolated and neatly hand-typed characters and words. The literature on the identification of Odia's character is limited, in the context of the literature that is based in the present work. In this work transfer learning technique ResNet50, VGG16 is used and compared with the traditional CNN-based method for classification of Odia characters.*

Key Words*: *Optical Character recognition, Convolutional Neural Network, Image processing, Transfer learning.

1. INTRODUCTION

Handwritten identification of characters belongs to the field of artificial intelligence, computer vision, and pattern recognition. It is said that a computer that performs handwriting recognition will understand and recognize the characters in documents, images, mobile devices, and many other sources and can convert it into an encoded form that can be processed in machines encoded languages. The application is used in smarter, more sophisticated character recognition systems for optical character recognition among them the neural network is one of the algorithms in machine learning which is widely used for most of the systems. Handwriting is one of the most widely used natural means of paper document production [1]. Majority of paper records are handwritten since it is a normal means of ease and time. The widespread acceptance of digital computers, as the keyboard becomes the most common method of text generation, seems to threaten the future of handwriting. Yet the simplicity of pen and paper is the cause why the handwriting is still standing in the generation of digital technology and is still present as a more realistic way of entering the data into the computer so that technology can decrypt the handwritten text. From another perception, we still find difficulties for computer generation of the local language text data for conventional keyboards because the conventional keyboards generally designed for the English language and yet need to be taught to map the various combinations to generate the local language scripts that typically have a large number of characters. It is yet one more reason why machine handwriting can be a more understandable input tool [2]. India has multiple languages in different regions and very rich literary scripts. Nearly all of the Indian scripts originated from the Ancient Brahmi through different transformations. A script may be shared in multiple languages, and multiple scripts may contain one language. The script systems for alphabets include simple characters, composite characters, and complex characters. The combination of one or more consonant and vowels results in both vowel and consonant set shape changes with complex characters. However, Indian scripts contain a great many character shapes.

2. RELATED WORK

This section provides a comprehensive historical analysis of OCR research and development. Tusheck first conceptualized OCR in 1927 and Handel in 1933 to create a system capable of reading characters and numerals. Using optical and mechanical OCR components Tusheck proposed the idea of matching models. OCRs of second-generation were able to recognize both handwritten and machine-printed characters in the early '70s and mid-'60s. Commercial OCR aimed for invoices, pre-printed order forms, and typed documents [3]. Many multinational companies marked for it. But automatic shorting machine for postal code numbers was first built by Toshiba. In 1996 IEEE patterns recognition dataset was established. Weightage was provided for the OCR pre- and post-processing stages [4]. In the third-generation OCRs, the exponential development in software and hardware technologies has helped create third-generation OCRs with high performance and low price. Such OCRs sought to tackle the problems of low-quality papers and large printed and hand-written character sets with complex characters. The researcher gave tremendous efforts in past years but the general usability of OCR seen in recent years. But still, there is less development in the field of handwritten characters due to limited vocabulary.

RC Sahoo in his work described the Odia classification using a Hopfield network that has similar architecture like LeNet-5 to avoid backpropagation. Hopfield network architecture is a fully connected network that works in associative memory in which neurons can be updated synchronously or asynchronously. The weighted sum is calculated at the same time for synchronous and randomly for asynchronous respectively. Tested on both vowels and consonants of Odia character images which gives the accuracy of 95.63% and 92.87% respectively [5]. B. Majhi described the digit recognition of Odia handwritten data using Discrete Fourier Transform, S-transform, wavelet transform, Short-time Fourier Transform, curvelet transform for the feature extraction, and then to classify multilayer perception, radial bias function network and probabilistic neural network. RBF-CT gives greater accuracy compared to FLANNP-DFT that is 98.70 % and 84.54% respectively [6]. Gradient-based and clonal selection based feature extraction method is used by P. Pujhari for handwritten Odia digits. For dimensionality reduction, he used principal component analysis. With 90.75 % accuracy, Genetic based Multilayer Artificial Neural Network model performed better compared to Clonal Selection Algorithm Multilayer Artificial Neural Network to recognize the handwritten digits. For individual digits, recognition accuracy lies between 90 % - 95 % [7].

3. OPTICAL CHARACTER RECOGNITION

Optically acquired images in a computer-processable form like UNICODE or ASCII are called optical character recognition. The method of interpreting the machine-printed or handwritten text (numbers, symbols, and letters). OCR usually refers to a technology that reads typed or handwritten, characters from ordinary documents and converts the symbols into a form that the machine can understand [8].

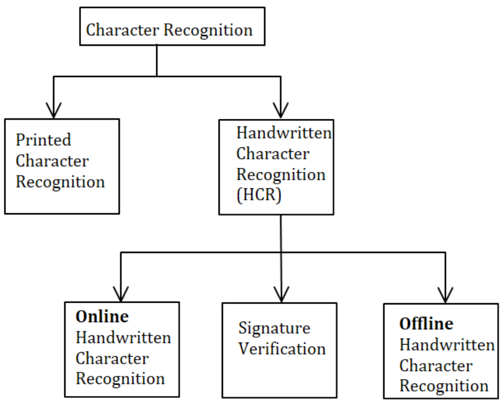

Fig-1: Classification of an optical character

4. MATERIALS AND METHOD

A total of 35 characters were selected in Odia-language. 10 different individuals have hand-written each of the characters and they is all stored together to form a database. The database is split into the training of 60% and testing of 40 percent. The input RGB images of handwritten data are initially gone through the pre-processing where it is converted to grayscale image and the further threshold is applied. The pre-processed image is now given as input to segmentation and reshaping block where the ROI of the image is extracted for further analysis. Finally, the image data from this block is provided to the CNN block for the classification of the characters are shown in fig2.

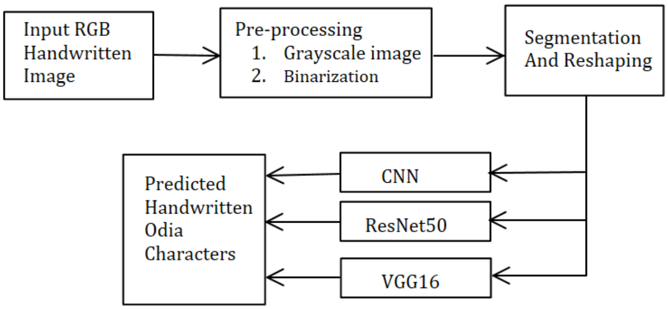

Fig-2: The proposed method for classification of Odia optical characters

5. PRE-PROCESSING

The characters are in RGB shape, so there isn't much detail on the other color channels. And you can pick any color channel to which the dimension while the information is intact for processing. Thus every sample image in RGB is converted to Grayscale in the present approach and further processing has been carried out. After the grayscale image is obtained, thresholding of intensity is done to minimize the complexity of the image. That is because the characters typically don't have any background information, and can be overlooked. So, the thresholding approach can simply eliminate the background. Such an implementation is laid out below [9].

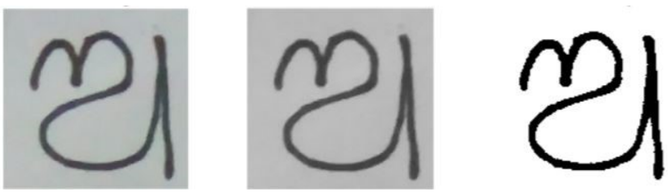

Fig-3: (a) RGB image (b) Grayscale image (c) Threshold binary image

6. SEGMENTATION AND RESHAPING

Obtained binary images undergo the morphological operation to get true character information. Major two steps are involved in the morphological operation are; 1) small noise is removed 2) larger connected area are identified. Noise reduction is achieved using opening operations and closing operations, and additionally, field analysis is performed. It selects the region that covers most of the linked pixels and chooses the largest region as a character.

Collected, and chosen as a character the largest area. To locate the rectangular pixel boundary, usually referred to as the Region of Interest (ROI), the pixel positions collected will be further analyzed. Figure 3 presents Implementation. After the boundary is again collected, the picture is reshaped to its original dimension 50 x 50 [10].

Fig-4: Segmentation to detect the ROI of a binary image.

7. CONVOLUTIONAL NEURAL NETWORK

The Convolutional Neural Network (CNN) is a traditional algorithm used by extracting features from images to solve computer vision problems. CNN has initially used for object detection but with time the improvements made it to grow in other fields such as pattern recognition, text detection, and recognition, object tracking, etc. In 1990 CNN was able to detect handwritten digits but not to a greater extent because of less computational power and low training data.

CNN uses a variant of the multilayer perceptron (MLP) that is designed to allow minimal preprocessing. The visual cortex of animals is the inspiration of CNN architecture. The basic unit of a convolutional network is the layers that are responsible for mathematical computations that are generally a set of feature maps where artificial neurons are arranged inside it. The complexity of the network is reduced by sharing weights within the same feature map. Convolutional networks are also trained with the backpropagation algorithm. The convolution layer and pooling layers are stacked together in a simple CNN layer, and their output is connected to a completely connected network. The final layer of a fully connected layer calculates the loss which shows the difference between actual and predicted output [11].

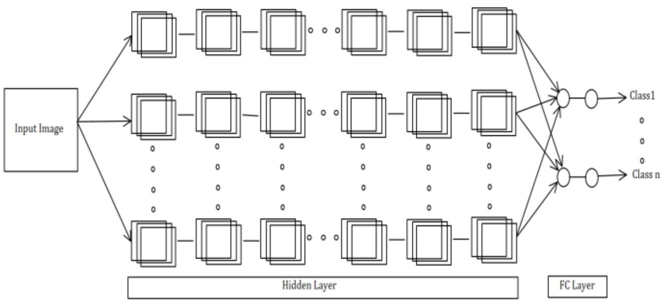

Fig-5: Basic CNN architecture

CNN uses different activation functions which helps to do some mathematical operations inside a network. The sigmoid function is a nonlinear activation function which gives the output in the range from 0 to 1 by taking as input a real-evaluated function. Unlike sigmoid tanh the output ranges from -1 to 1. Both sigmoid and tanh have the drawback of vanishing gradient. Tanh is a linear activation function having a threshold at 0. Leaky relu is a slight modification to relu which is mathematically represented below.

$$Sigmoid = f(x) = \frac{1}{1+e^{-x}} \tag{1}$$

$$tanh(x) = 2f(2x) - 1 \tag{2}$$

$$relu(x) = max(0, x) \tag{3}$$

$$leaky\ relu = f(x) = \begin{cases} x, & x < 0 \\ 0.01x, & otherwise \end{cases} \tag{4}$$

CNN shows better compared to other algorithms in the area of image processing and also the pre-processing part in CNN is very less. This ensures that the network knows the filters that were hand-engineered in conventional techniques. There's a huge advantage to this independence from previous experience and human effort in feature design. These include tools for image and video recognition, suggest structure, and face detection, biomedical image analysis, and NLP. Although there are many pre-trained models like GoogleNet, VGGNet, AlexNet, etc. In this work, we have used the same architecture as CNN which was used to classify the Malayalam language. The architecture of the CNN consists of six layers, to be further discussed. CNN's suggested approach is set out below.

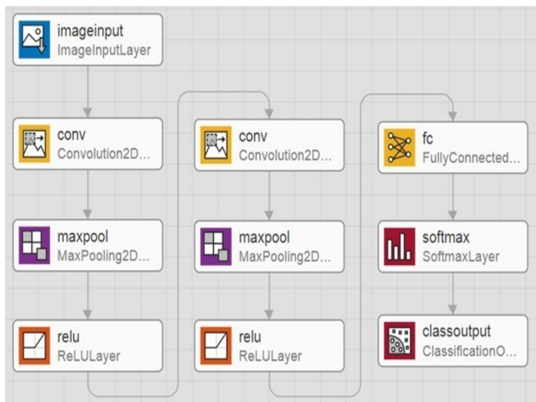

Fig-6: CNN architecture of the proposed method

8. ResNet50

The key concept of ResNet is to skip the connection between convolutional layers which is called the Residua Building

Blocks (RBB). That helps to overcome the problem of exploding/ vanishing gradients. Residual building blocks are consist of many batch normalizations (BN), Conv layers, the rectified linear unit as activation, and finally 1 shortcut. RBB-1 and RBB-2 have different structures are used in this work. Batch normalization and Convolutional layers are present in both RBB-1 and RBB-2. But the shortcut is provided with another Conv and BN in case of RBB-2 whereas RBB-1 is the identity x of a nonlinear function x. The output of both RBB-1 and RBB-2 can be obtained by using eq(5) and eq(6).

$$y = f(x) + x \qquad (5)$$

$$y = f(x) + H(x) \qquad (6)$$

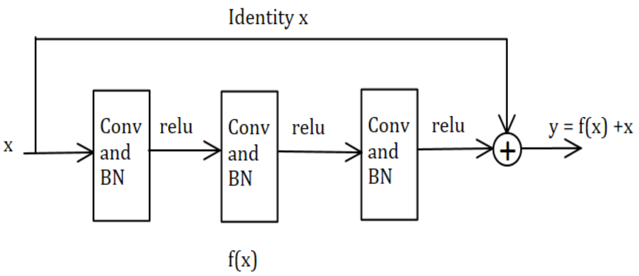

Fig-7: Residual Building Block -1 architecture

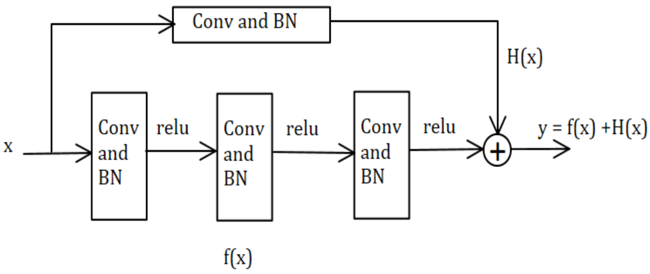

Fig-8: Residual Building Block -2 architecture

The architecture of ResNet50 is presented in fig 8 One Conv layer and 16 residual blocks complete the fifty layers of the model. We have added a fully connected layer and a softmax classifier having 35 classes of Odia characters [12].

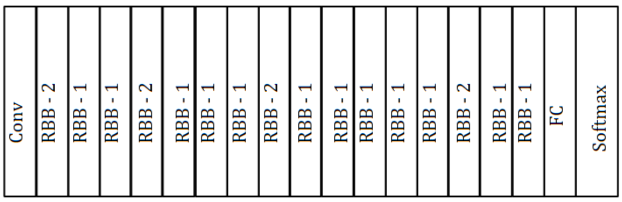

Fig-9: ResNet50 architecture having RBB-1, RBB-2, FC and softmax unit.

9. VGG16

Simonyan and Zisserman proposed the VGG network in 2014. VGG16 architecture contains five Conv layers with 3 fully connected layers (FC) at the last pooling layer of the VGG16 network. 3 x 3 kernels are used for the Conv layer of stride and padding of 1. Activation function relu is used right after each convolution that minimizes the spatial dimension and 2 x 2 kernel is used max pool having no padding and stride of 2. Which halves the spatial dimension from the previous layer of the activation function [13].

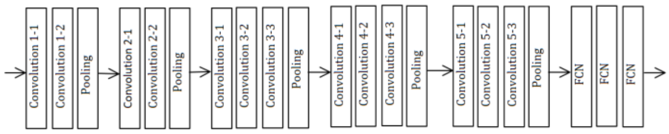

Fig-10: VGG16 architecture

10. RESULTS

We have taken 25 random input samples from the database out of 35 Odia characters. Then the sample is binarized. After binarization, the images are ready for segmentation which are shown in fig 11, 12, and 13 respectively. Further, the input to the convolutional neural network is provided with segmented images to classify.

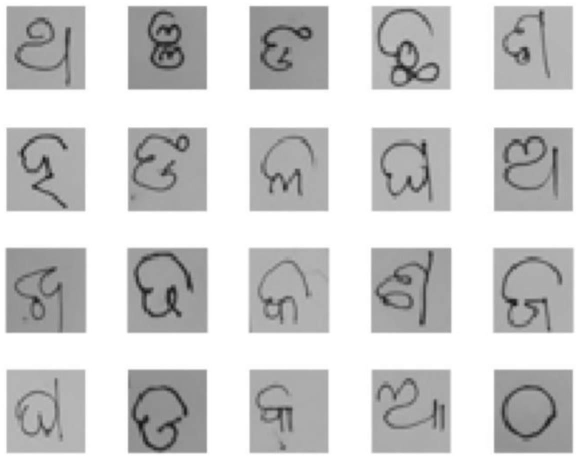

Fig-11: Random character samples

Fig-12: Thresholding on the selected sample

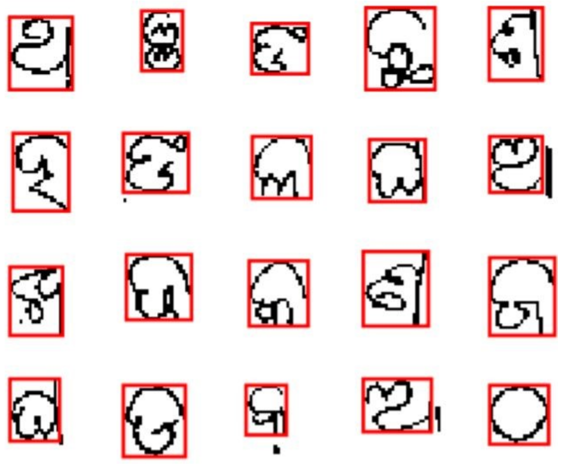

Fig-13: Segmentation result

The training and test process is repeated 5 times, and the training and testing accuracy are documented in fig. 14, 15 and 16 respectively. Training and testing accuracy increased linearly in the case of ResNet50 and VGG net whereas CNN performance was average as giving 70% accuracy.

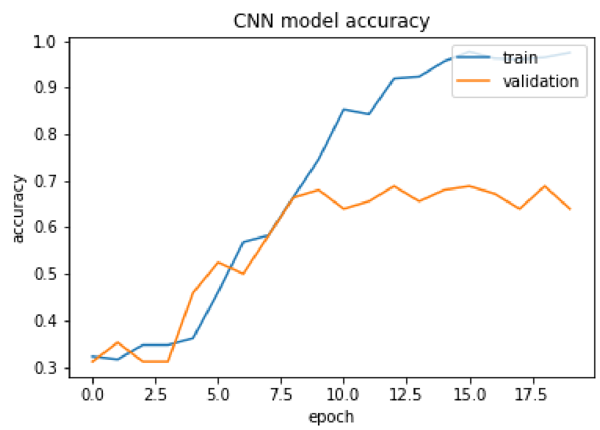

Fig-14: CNN model accuracy graph

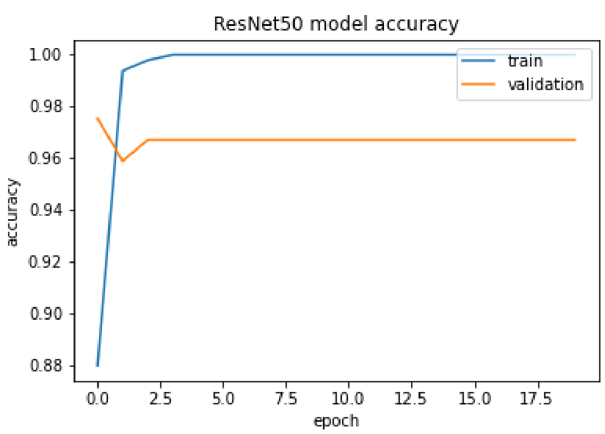

Fig-15: ResNet50 model accuracy graph

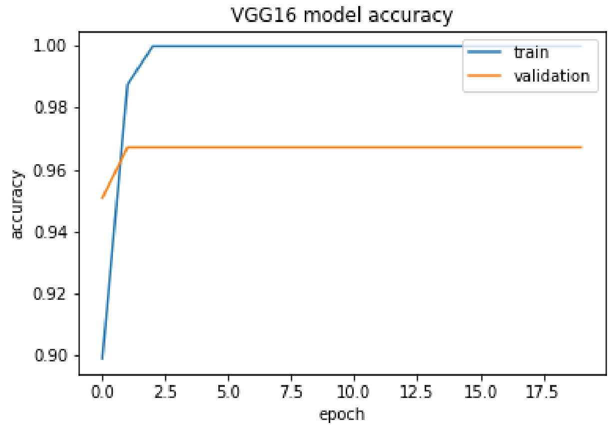

Fig-16: VGG model accuracy graph

11. CONCLUSIONS

Optical character recognition for Odia handwritten characters is one of the active research fields which still needs improvement inaccuracies. This work focuses on using convolutional neural networks (CNN) as well as transfer learning model ResNeand VGG16 to classify Odia characters, which are handwritten. The accuracy obtained promised to be around 96 percent an effective technique for character classification using transfer learning and 70% using CNN. However other feature extraction techniques such as Local binary pattern, Gabor features, etc. can be used to extract features efficiently. Similarly, the classification algorithms like Naïve Baise, Support Vector Machine (SVM), and decision tree can be implemented. Moreover, deep learning algorithms have always been on the top of all while handling the image data.

REFERENCES

[1] M. J. a. R. K. S. Kumar Munish, "Review on OCR for handwritten Indian scripts character recognition," in International Conference on Digital Image Processing and Information Technology, Springer, Berlin, Heidelberg, 2011.

[2] M. J. a. R. K. S. Kumar Munish, "Review on OCR for handwritten Indian scripts character recognition," in International Conference on Digital Image Processing and Information Technology, Springer, Berlin, Heidelberg, 2011.

[3] A. S. A. C. NV Rao, "OPTICAL CHARACTER RECOGNITION TECHNIQUE ALGORITHMS," Journal of Theoretical & Applied Information Technology, vol. 83.2, 2016.

[4] A. P. M. E. M. M. K. Anil R, "Malayalam Character Recognition using Singular Value Decomposition," International Journal of Computer Applications , vol. 92, 2014

[5] Ramesh Chandra Sahoo , Sateesh Kumar Pradhan ,Poonam Tanwar, "HopNet based Associative Memory as FC layer in CNN for Odia Character Classification," in 2020 10th International Conference on Cloud

Computing, Data Science & Engineering (Confluence), Noida, India, India, 2020.

[6] B. M. Puspalata Pujari, "Recognition of Odia Handwritten Digits using Gradient based Feature Extraction Method and Clonal Selection Algorithm," International Journal of Rough Sets and Data Analysis , vol. 6, pp. 19-33, 2019.

[7] Babita Majhi, Puspalata Pujari, "On Development and Performance Evaluation of Novel Odia Handwritten Digit Recognition Methods," COMPUTER ENGINEERING AND COMPUTER SCIENCE, 2017.

[8] U.PalB.B.Chaudhuri, "Indian script character recognition: a survey," Author links open overlay panelU.PalB.B.Chaudhuri, vol. 37, pp. 1887-1899, 2004.

[9] AmitChoudhary, RahulRishi, Savita Ahlawat, "Off-line handwritten character recognition using features extracted from binarization technique," AASRI Procedia, vol. 6, pp. 306-312, 2013.

[10] Musbah J. Aqel, Ziad ALQadi, Ammar Ahmed Abdullah, "RGB Color Image Encryption-Decryption Using Image Segmentation and Matrix Multiplication," International Journal of Engineering & Technology, vol. 3.13, pp. 104-107, 2018.

[11] M. G. Neena Aloysius, "A review on deep convolutional neural networks," in 2017 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2018.

[12] Long Wen, Xinyu Li &, Liang Gao, "A transfer convolutional neural network for fault diagnosis based on ResNet-50," Neural Computing and Applications volumeNeural Computing and Applications volume, 2019.

[13] Edmar Rezende, Guilherme Ruppert, Tiago Carvalho, Antonio Theophilo, Fabio Ramos, Paulo de Geus, "Malicious Software Classification Using VGG16 Deep Neural Network's Bottleneck Features," in Information Technology - New Generations, Springer, 2018, pp. 51-59.

BIOGRAPHIES

Rakesh Kumar Sethi completed his B.Tech in Electronics and Telecommunication Engineering and received M.Tech in Signal processing Engineering from the College of Engineering and Technology, Bhubaneswar. His active research area belongs to computer vision.

Kalyan Kumar Mohanty is working as a research intern at the *Center of Excellence on Artificial Intelligence(COE-AI)*, CET, Bhubaneswar completed his B.Tech in Electronics and Telecommunication Engineering and received M.Tech in Signal processing Engineering from the College of Engineering and Technology, Bhubaneswar. His active research area belongs to Computer vision and Natural Language Processing.