Journal of Advances in Science and Technology
Vol. 16, Issue No. 1, March-2019, ISSN 2230-9659

A Study of Convolutional Neural Network for Natural Image Classification

Amit Kumar Pandey[1]* Dr. P. K. Bharti[2] Dr. Prashant Singh[3]

[1] Research Scholar, Department of Computer Science and Engineering, Shri Venkateshwara University, Gajraula, Amroha, Uttar Pradesh

[2] Vice Chancellor, Shri Venkateshwara University, Gajraula, Amroha, Uttar Pradesh

[3] Department of Information Technology, Dr. Akhilesh Das Gupta Institute of Technology and Management, Delhi

Abstract – The accuracy of deep convolutional neural networks, which can also fulfill the function of implicit model ensemble without incurring extra model training costs. Simultaneous data augmentation throughout training and testing phases helps assure network optimization and boost its generalization ability. Augmentation in two phases needs to be consistent to guarantee the appropriate transmission of particular domain knowledge. Picture classification finds its use in practically every sector including quality inspection, illness diagnosis, face identification, image, video recognition, etc. The development of convolutional neural network is a significant advance in the area of machine learning. On the other hand, a convolutional neural network receives the picture itself as the input and classifies the image based on the likelihood scores generated. A convolutional neural network may be constructed according to the demand. One of the remarkable advantages of a convolutional neural network system is its ability to handle massive volume of data. A convolutional neural network is a unique sort of neural network that is constructed by stacking many nonlinear layers one after the other. The input picture is translated into semantic features and supplied to subsequent layers and eventually gets converted to class score.

Keywords – Convolutional Neural Network, CNN, Natural Image, Classification, etc.

---------------------------------X---------------------------------

INTRODUCTION

The application of deep convolutional neural networks (CNNs) in computer vision has exploded, yielding answers for a wide range of computer vision problems. As more data and processing power become accessible, over parameterized deep learning models' capacity to perform better based on their extremely nonlinear fitting skills has developed. Deep CNNs trained on massive datasets nevertheless fall short when applied to new datasets that haven't been trained because to the "over fitting" problem [1]. Larger models tend to do better, but this comes at the cost of sacrificing accuracy in favor of reasoning speed. If a photograph's natural picture is to maintain its emotive traits, it must be cleansed of any noise. The process of image categorization is carried out by a computer/machine system. It entails object detection in photos based on the spatial relativity of pixels [2].

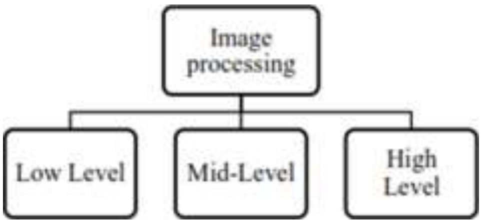

Figure 1: Levels of Image Processing

Image processing is classified into three levels based on the depth of processing and the result obtained: low-level, middle-level (mid-level), and high-level image processing. Figure depicts a schematic of the three layers of image processing. Picture preprocessing, image enhancement, image restoration, de-noising, contrast enhancement, and other low-level image processing techniques are included [3].

OBJECT DETECTION SYSTEM

The convolution neural network of the object detection system is trained in three distinct contexts. In the first configuration, data from the datasets is used without any data augmentation. This was be used as a comparison point for different data augmentation methodologies. In the second situation, data augmentation procedures are utilized to generate more data. This synthetically manufactured data is added to the training data. This new, bigger training data set is used to train the convolutional neural networks. In the third and final option, GAN-generated samples are added to training data. A convolutional neural network is trained using this additional training data. The network established in section 3.4 is used in training, testing, and all three scenarios. The identical set of training conditions was used across all trials in order to compare the results. The network is trained with a batch size of 32 for a total of 20k iterations. As a result, unpredictability must be considered when employing stochastic gradient descent to train the network. Training is repeated 10 times with a fresh random seed to eliminate unpredictability. An object detection system was now be used to refer to this network [3- 4].

- Object detection system with no augmentation
- Object detection system with manual data augmentation
- Objection detection system with neural data augmentation

UNSUPERVISED CLASSIFICATION

Unsupervised classification is used to arrange pixels in an image in a natural way. Unlike supervised classifiers, it does not require any training. An unsupervised classification technique based on similarities inside and across classes is used to classify pixels. The following are some of the most often used unsupervised classification methods [4]:

- Hierarchical clustering algorithm
- K-means clustering algorithm
- Fuzzy C-means clustering algorithm
- Iterative self-operating data analysis technique (ISODATA)

SUPERVISED CLASSIFICATION

In supervised classification, classifiers are trained. This type of classifier takes advantage of what it has learned throughout its training. In supervised classification, feature extraction from photos is utilised to establish a unique signature for each class. Once the training has been performed properly, the performance of the classification system is assessed using test photographs. A classifier employs a signature to determine the classification of an image while it is being tested. One or more of the following techniques can be used to do supervised classification [4-5]:

- Minimum Distance from Mean (MDM) classifier
- Parallelepiped classifier
- Maximum Likelihood (ML) classifier
- Support Vector Machines (SVM) classifier
- K-Nearest Neighbours (KNN) classifier
- Artificial Neural Networks (ANN)
- Convolutional Neural Networks (CNN)

CONVOLUTIONAL NEURAL NETWORKS

Convolutional neural network architecture takes use of the picture's spatial structure. Furthermore, the design is more logically constrained. The primary distinctions between CNNs and fully connected architectures are local receptive fields, weight sharing, and pooling. In a fully connected architecture, pixels arranged in a vertical line are straightforward to grasp. Understanding CNN is made easier by visualizing pictures as two-dimensional structures. In terms of structural closeness, it's a reasonable assumption that pixels next to each other are more associated than pixels far apart. CNN employs a similar method. In a completely connected architecture, a neuron's activity is determined by the activation of all neurons in the layer before it. CNNs employ a different method to determine the value of neurons: they use a small chunk of neurons from the layer preceding it to do so. The activity of a neuron may be calculated using a 5 x 5 sub-window and an image with a resolution of 28 × 28 pixels [5].

FEED-FORWARD NEURAL NETWORKS

To understand CNNs, you must first understand what neural networks are. A fundamental objective for neural networks is to create robots that imitate the brains of animals. With the use of neural networks, unsupervised and semi-supervised learning tasks are becoming increasingly popular. We'll look at supervised learning problems and an example neural network model to get a better understanding of what neural networks are. A labelled dataset may be used to identify features and labels for each training data set (xi, yi) (x, yi). Neural networks may be used to learn non-linear connections between yi and xi (xi). In neural network topologies, circles and edges represent neurons and connections between neurons, respectively. The way neurons are layered is referred to as neural network architecture. In any neural network architecture, there are three sorts of layers. The input layer is called "hidden," followed by the output layer, and any other levels in between are called "invisible." The sample design only shows one hidden layer, however there are multiple hidden layers in reality. The output of the preceding layer determines the value of each hidden and output neuron, and all neurons in the hidden and output layers are computing units. Except for the input layer, each neuron in the previous layer is coupled to every other neuron in the previous layer [6].

IMAGE CLASSIFICATION METHODS BASED ON APPROACHES

Pattern matching classifiers: Pattern matching classifiers generally employ either quantitative or structural patterns to determine the best match for the input pattern. The pattern classes are grouped closely and kept as far away from each as possible.

Optimum statistical classifiers: Optimum statistical classifiers use a probabilistic approach for classification. This kind of classifiers generally provides lowest probability of incurring errors in classification.

Neural network classifiers: Classifiers that use neural networks with nonlinear computing aspects are known as neural network classifiers. The pieces are linked in a similar way to how actual neurons are connected. Parallel distributed processing model, neuro computer, layered self-adaptive network, neuromorphic system, and connectionist model are all terms used to describe a neural network.

IMAGE CLASSIFICATION METHODS BASED ON INFORMATION EXTRACTED

Spectral pattern classification: The image in spatial domain is converted to spectral domain and the measurements taken from different frequency bands are used for classifying the images.

Spatial pattern classification: The information extracted from each pixel and the relation it holds with the neighboring pixels are used for image classification.

Temporal pattern classification: Variations observed in pixels over a period of time are collected. Data on temporal variations are converted into meaningful features. The features thus generated are used for classification [7].

LITERATURE REVIEW

Akcay et al. (2018) the use of deep learning convolutional neural networks for luggage identification and classification has been examined. Based on extensive training on generic photos, a pre-trained CNN was used to identify and classify items. Using a transfer learning strategy, the problem of a lack of access to the item of interest was eliminated. Additionally, a variety of CNN-driven detection patterns, including sliding window-based CNN (SW-CNN), region-based fully convolutional networks (R-FCNs), faster region-based networks (F-RCNNs), and YOLOv2 (you only look once - version 2), were employed. Six-class object recognition was handled by the YOLOv2 with an 88.5 percent accuracy rate.

Yang et al. (2018) An advanced neural network was used to recognise facial expressions (FER). The goal of this method was to better understand how people communicate their emotions. There is no doubt that people's emotional expressions differ from one another and from one another at different points in time. This problem was taken into account when designing the suggested system. A weighted mixed deep neural network was used to automatically extract features in this technique (WMDNN). Several pre-processing methods, such as face identification, data augmentation, and rotation correction, were used to minimise the regions that would be subject to FER. Two channels of data were sent into the WMDNN for processing. Grayscale facial images were given by one channel and the equivalent binary pattern images were provided by the other.

Khan et al. (2016) system of convolutional activation (DUCA) for scene identification in indoor environments has been proposed. It was first necessary to establish a code book of Scene Representative Patches (SRPs), from which the input image's extracted mid-level patches were encoded. It was possible to use supervised and unsupervised SRPs in the code book. In order to encode the characteristics in terms of broad object categories while still including the target's discriminative properties, a highly discriminative feature space was established. Class-belonging determinations were made using a linear classifier that aggregated relationships of the data. In order to recognise the interior environment, we had to contend with inter- and intra-class similarities. It was determined that the suggested approach could accurately classify three different types of interior scenes: those collected by New York University, Massachusetts Institute of Technology (MIT), and 15 other types of scene classifications.

Chen et al. (2016) SAR (Synthetic Aperture Radar) image categorization approach was advocated by the authors. Using CNN as a tool is part of the process (convolutional neural network). To avoid over-fitting, a particular sort of CNN, called 'all convolutional neural networks' (A-ConvNets), was developed for the categorization of SAR images using synthetic aperture radar (SAR). Rather of a completely linked layer, A-ConvNets have a sparsely connected layer. For the AConvNets framework, four convolutional layers and three max-pooling layers were used. As a result, the computing time was lowered due to the lack of a completely linked layer. On the Moving and Stationary Target Acquisition and Recognition (MSTAR) dataset, this element dramatically improved the classification accuracy to 99 percent.

Jaswal et al. (2014) Developed a deep learning technique to address picture categorization issues. SUN database scene photos and aerial images from UC Merced land use dataset were used to examine how well this algorithm performed. The Mean Squared Error (MSE) and classification accuracy were measured. Using the pyramid reduction approach, the pictures from the datasets were reduced to 32 by 32 pixels. The characteristics in the aerial photos were categorised into building region, dense residential area, agricultural region, forest region, built region, and green region by using the algorithm and the land use pattern of the images. A beach scene, garden scene, road scene and combat scenario were all categorised. Classification accuracy for training data was 97.50%, while for test data for road scenes, it was 91%. For created regions, 97 percent categorization accuracy was achieved using aerial data.

Priya et al. (2013) employing Probabilistic Neural Network (PNN), Bayes Classifier, and Support Vector Machine (SVM) to categorise DR images as NPDR or PDR (SVM). The raw pictures were analysed to identify blood vessels, haemorrhages, and exudates. A total of 130 photos from "DIARETDB0: Evaluation Database and Methodology for Diabetic Retinopathy" were used to test out the algorithm's performance, and the suggested system achieved classification accuracies of 87.69%, 90.76% and 95.38%, respectively.

Balint Antal & Andras Hajdu (2012) shown how to use a simulated annealing-based search technique to find micro aneurysms in pictures by picking the best preprocessing and candidate extractor pair combinations. These techniques include Walter-Klein contrast enhancement, vessel elimination and extrapolation in the pre-processing step of the image processing. Candidate extraction was contemplated using a variety of successful methods. Simulated annealing was used to find the best mix of pre-processing and candidate extraction pairings. The ideal pair was chosen because it had the fewest false positives. Energy function selection was critical to the achievement of scalability in simulated annealing To ensure the results were accurate, clinical specialists carefully calculated the centroids of micro aneurysms.

Sabeenian et al. (2012) novel techniques to identify and find faults in fabric have been proposed by combining the multi resolution combined statistical and spatial frequency (MRCSF) and Markov random field matrix (MRF) approaches. With the addition of second-order statistical characteristics like Markov random field matrix (MRFM) or gray-level cooccurrence matrix (GLCM), the researchers were able to mix first and second order statistical features. Weaving flaws such as holes, shuttle mesh, float, missing ends and temple markings were examined in the method's testing on a database of local flaws. Features were compared to library features and identified as defective or non-defective using the nearest neighbour approach. 96.66% accuracy was achieved in the testing of 25 samples using the suggested method.

Nanni et al. (2012) proposed the Linear Quinary Pattern as a five-value encoding method (LQP). For images with nearly consistent noise levels, we wanted a more robust description. The centre pixel may have any of the five values in LQP if two threshold values were used. It was then divided into four binary patterns, and the histograms generated from each of these binary patterns were concatenated. In the beginning, LQP was used to extract the rotation-invariant (Set A) and uniform bins (Set B). 125 characteristics were chosen from Set A, which had 250 bins with substantial variation. Using principle component analysis (PCA), we were able to minimize the number of features and then feed the reduced features into a support vector machine (SVM) for further categorization.

Vijaya Kumari et al. (2010) extracted an image of the optic disc from a retinal picture using this approach. Principal component analysis was used by the researchers. The detection probability of exudates around the optic disc was raised by 25-40% using upgraded MDD classifier. Exudates were discovered in 35 of the 39 photos evaluated utilizing the suggested technique, while healthy retinas were found in four of the images.

Zhang et al. (2010) proposed a method for detecting retinal micro aneurysms using a two-stage algorithm. First, coarse-level micro aneurysms were detected, and then fine-tuning was implemented. With a multi-scale Gaussian kernel sliding neighborhood filter, we were able to calculate the correlation coefficient for coarse candidate selection. It was found that items with a circular shape had the highest correlation with the Gaussian kernel. The correlation coefficients ranged from zero to one. More likely to have micro aneurysms were pixels with coefficients closer to one. Color intensity, gray-scale pixel intensity, shape, responses from Gaussian filter-banks, and correlation coefficient values were used to extract features in fine-level detection of color. Refinement tables for each of the 31 characteristics of genuine micro aneurysms were created.

Mak et al. (2009) method to identify weave faults has been created. In addition to the algorithm, morphological filters were also used. The textural properties of the textile fabric were retrieved using a Gabor wavelet network. It was determined which characteristics could be gleaned from the data. The morphological filters were used to identify and separate the fabric flaws from the background pixels. To limit the amount of computational complexity, just a few morphological filters were used. Few false alerts were generated by the system. Different types, forms, and sizes of fabric flaws were used to test the suggested architecture's performance.

OBJECTIVES OF THE STUDY

- To design a deep learning architecture for image classification using convolutional neural networks

- To apply the suggested architecture in the medical profession to grade images based on illness severity.

- To assess the proposed Convolutional Neural Network architecture's performance using hyper parameters such as filter size, stride, zero padding, activation functions, pooling layer, and epochs.

RESEARCH METHODOLOGY

Problem Formulation

Given a deep Convolutional Neural Network model $\mathrm{M}_\theta : f(\mathbf{x};\theta_0)$ trained on the training set, $D : \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N}, (\mathbf{x}, \mathbf{y})$ and $\theta_0 : \{\mathbf{W}_0^l, \mathbf{b}_0^l\}_{l=1}^{L}$ represent the real-world images and the associated ground truth labels, as well as the network parameters itself. Four-dimensional tensors and two-dimensional matrices are used to group parameters in the convolutional and fully connected layers. An SGD algorithm based on back propagation is used to optimize the network in a small batch.

This level is referred to as "forward propagation" since each layer's output is sent into the next; the deep CNN's hl output for l = 1 is... The expression for L 1 is as follows:

$$\mathbf{h}_l = \sigma\left(\mathbf{W}_0^l \mathbf{h}_{l-1} + \mathbf{b}_0^l\right),$$

Where $\mathbf{h}_0 = \mathbf{x}$ and $\sigma(\cdot)$ an element-wise non-linear activation function, such as Leaky-ReLU

$$\sigma(x) = \begin{cases} x, & \text{if } x > 0, \\ \dfrac{x}{a}, & \text{if } x \leq 0, \end{cases}$$

Where a is a fixed hyper parameter in (1, +∞). Finally, the deep Convolutional Neural Network model's final output may be achieved by:

$$f(\mathbf{x}) = \operatorname{softmax}\left(\mathbf{W}_0^L \mathbf{h}_{L-1} + \mathbf{b}_0^L\right),$$

Where softmax (·)The logarithmic normalizing function of a finite term discrete probability distribution may be used to calculate and define this parameter.

$$\operatorname{softmax}(f)_i = \frac{e^{f_i}}{\sum_{j=1}^{C} e^{f_j}}, \quad \text{for } i = 1, 2, \ldots, C,$$

The last layer of neurons, or neurons in that layer, has a total number of categorization categories equal to C. A deep Convolutional Neural Network's training loss may also be calculated using this method.

$$\mathscr{L}(\mathbf{x}_i, \mathbf{y}_i) = -\frac{1}{C}\sum_{j=1}^{C}\left[y_i^j \log f(\mathbf{x}_i)^j + \left(1 - y_i^j\right)\log\left(1 - f(\mathbf{x}_i)^j\right)\right] + \lambda \sum_{k=1}^{L} \|\mathbf{W}^k\|_F.$$

The first term is the L2 regularization of all weights, while the second is the negative log-likelihood loss. Λ the decay rate of the weight influences both the regularization intensity and the Fresenius norm. It is used for testing and training by constantly refining the loss function and adjusting the network parameters.

Reducing the number of parameters (weights W and biases b) in deep Convolutional Neural Networks is a primary goal of the back propagation stage in the algorithm. A mini-batch SGD implementation allows for parameter changes during the tth training run.

$$\mathbf{W}_t^l = \mathbf{W}_{t-1}^l - \alpha \cdot \frac{1}{M}\sum_{i=1}^{M}\frac{\partial \mathscr{L}(\mathbf{x}_i, \mathbf{y}_i)}{\partial \mathbf{W}_{t-1}^l},$$

$$\mathbf{b}_t^l = \mathbf{b}_{t-1}^l - \alpha \cdot \frac{1}{M}\sum_{i=1}^{M}\frac{\partial \mathscr{L}(\mathbf{x}_i, \mathbf{y}_i)}{\partial \mathbf{b}_{t-1}^l},$$

Where α and M represent the learning rate and batch size, respectively.

Data Augmentation during Training Process

Images are only one of a number of ways in which the potential anatomy of the subject may be seen through various spatial transformations and noise disturbances. Biased outcomes are possible when relying solely on pictures that have been gathered by means of direct inference. The "overfitting" problem in deep Convolutional Neural Networks can be reduced by using a full-stage data augmentation method proposed here.

Translation, horizontal flipping, and noise disruption techniques can be used to expand the training samples in a mini-batch set at the tth training iteration. In this augmentation, the translation step, rotation range, and noise level are all fixed. There are no augmentations made at the beginning of the training process; rather, they are made at the time of data input. In this way, the training data are expanded to $\widetilde{M}/M$ time and number of training iterations in the original data is nearly identical.

Data Augmentation during Testing Process

There are two levels of augmentation parameters that we apply at this step, which is known as the test stage. Enhancements are made to each test image $\widetilde{M}/M$ Similar to the training procedure; photos are augmented with similar data. To achieve proper domain information transfer, data augmentation in all stages must be consistent. The $\widetilde{M}/M$ The final forecast is based on a majority vote of the prediction results:

$$f(\mathbf{x}) = \frac{M}{\widetilde{M}} \sum_{i=1}^{\widetilde{M}/M} f(\tilde{\mathbf{x}}_i)\,.$$

The final prediction is the label that corresponds to the biggest value in the one-dimensional vector f (x). In the event of a tie vote, the final forecast outcome is determined by the category with the highest likelihood.

Interpretation as Model Ensemble

The use of deep Convolutional Neural Network's trained on a variety of noisy datasets is often beneficial. To train each neural network, however, takes an increasingly huge number of noisy data sets. This is extremely costly. Each test sample undergoes a data augmentation procedure at this point. $(\mathbf{x} \longrightarrow \tilde{\mathbf{x}})$ Can be viewed as $\tilde{\mathbf{x}} = g(\mathbf{x})$ where g (·) is an example of how the same data augmentation might be applied. Various symbols can be used to represent data augmentation techniques g (·).Accordingly, the final forecast based on majority vote may be expressed as follows:

$$f(\mathbf{x}) = \frac{M}{\widetilde{M}} \sum_{i=1}^{\widetilde{M}/M} f\left[g_i(\mathbf{x})\right] = \frac{M}{\widetilde{M}} \sum_{i=1}^{\widetilde{M}/M} \tilde{f}_i(\mathbf{x})\,,$$

Where $\tilde{f}_i$ learners with different interests and abilities might be considered. Data augmentation in the test step can provide an implicit model ensemble if all samples are independent and identically distributed. On the training set, over fitting may be avoided by using a converging network that minimizes bias and variance.

CONCLUSION

The proposed Convolutional neural networks classification system was performed very well with max pooling layer, learning rate of 0.01, and 2000 epochs. The overall classification accuracy produced by the proposed Convolutional neural networks architecture under specific network conditions falls in the range of approx 80% to 90% for images, and The proposed CNN architecture was tuned to function under different combinations of network parameters. Production of feature map from each layer was much helpful in the extraction of significant features from the images.

REFERENCES

1. Samet Akcay, Mikolaj E Kundegorski, Chris G Wascocks & Toby P Breckon 2018, *'Using Deep Convolutional Neural Network Architectures for Object Classification and Detection within X-ray baggage Security Imagery'*, IEEE Transactions on Information Forensics and Security, vol. 13, no. 9, pp. 2203-2215.

2. Biao Yang, Jinmeng Cao, Rongrong Ni & Yuyu Zhang 2018, *'Facial Expression Recognition using Weighted Mixture Deep Neural Network based on Double Channel Facial images'*, vol. 18, IEEE Access.

3. Khan, SH, Hayat, M, Bennamoun, M, Tognerim, R & Sohel, F 2016 , *'A Discriminative Representation of Convolutional Features for Indoor Scene Recognition',* IEEE Transactions on Image Processing, vol. 25, no.7, pp. 3372- 3383.

4. Sizhe Chen, Haipeng Wang, Feng Xu & Ya–Qiu 2016, 'Target Classification using the Deep Convolutional Networks for SAR Images', IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 8, pp. 4806-4817.

5. Jaswal, D, Vishvanathan, S & Soman, KP 2014, *'Image Classification using convolutional neural Networks',* International Journal of Scientific and Engineering Research, vol. 5, no.6, pp. 1661-1668.

6. Priya, R & Aruna, P 2013, 'Diagnosis of Diabetic Retinopathy Using Machine Learning Techniques', ISSN: 2229-6956(Online) ICTACT Journal on Soft Computing, vol. 3, no. 4.

7. Balint Antal, and Andras Hajdu, 2012, 'An Ensemble-based system for micro aneurysm Detection and Diabetic Retinopathy Grading', IEEE Transactions on Biomedical Engineering, vol. 59, no. 6, pp. 1720

8. Sabeenian, RS, Paramasivam, ME & Dinesh, PM 2012*, 'Computer vision based Defect detection and identification in Handloom silk industries',* International Journal of Computer Applications, vol. 42, no. 17, pp. 41-48

9. Lorris Nanni, Alessandra Luminiz & Sheryl Brahnam 2012, *'Survey on LBP based texture descriptors for image classification', Expert Systems with Applications,* vol.39, pp. 3634-3641.

10. Vijaya Kumari, V & Suriya Narayanan, N 2010, *'Diabetic Retinopathy-Early Detection Using Image Processing Techniques*', (IJCSE) International Journal on Computer Science and Engineering vol. 02, no. 02, pp. 357-361

11. Zhang, B 2010, 'Detection of Micro aneurysms using Multiscale Correlation Coefficients', Pattern Recognition, vol. 43, pp. 2237-2248

12. Mak, KL, Peng, P & Yiu, KFC 2009, 'Fabric defect detection using morphological filters', Image and Vision Computing, vol. 27, no. 10, pp. 1585-1592.

Corresponding Author

Amit Kumar Pandey*

Research Scholar, Department of Computer Science and Engineering, Shri Venkateshwara University, Gajraula, Amroha, Uttar Pradesh