Journal of Artificial Intelligence Research & Advances
ISSN: 2395-6720
Volume 7, Issue1
www.stmjournals.com

# Multi-class Content-based Visual Information Retrieval using Multi-Layer Perceptron and Convolution Neural Network: A Comparative Study

***Shamik Tiwari****
Student, Department of Virtualization, School of Computer Science, University of Petroleum and Energy Studies, Dehradun, Uttarakhand, India

***Abstract***
*Content-based visual information extraction is an area, which utilizes the computer vision methods for searching and indexing the images in image library more competently than manual annotations. Due to the exponentially increase in image database volume, there is a demand of efficient retrieval system. Deep learning-based machine vision systems have the ability to handle large amount of data efficiently than ordinary machine vision systems. This paper proposed a content-based image search system, which utilizes convolution neural network based deep learning model for multi-class content-based image search. The results show the superiority of convolution neural network model on multi-layer perceptron model.*

***Keywords:*** *Image informatics, deep learning, multi-layer perceptron, convolution neural network*

****Author for Correspondence*** E-mail: shamiktiwari@hotmail.com

## INTRODUCTION

Content-Based Visual Information Extraction (CBVIE) also referred as Content-based Image Search (CBIS) and Query by Visual Information (QBVI) is an application of image informatics, that is used to be automated labelling of image in image gallery. The word "Content-based" signifies that the retrieval process examines the visual information of the image instead of the metadata like keywords, identifiers, or details related with the image [1]. Content based image search systems have wide range of applications including medical image processing [2], biometrics [3], biodiversity informatics [4], digital libraries [5], e-commerce [6], electronic retailing [7], product recognition [8] and many more. These days due to high availability of computational resources, deep learning based computer vision approaches have become an effective tool for CBIS. Deep learning is a part of machine learning domain. Deep learning is built upon both unsupervised and supervised learning approaches, which are based on Artificial Neural Network (ANN). Its architecture consists of multiple processing layers. Each layer produces nonlinear response, which is generated using the response of previous layer. Deep learning functions by imitating the human brain working for data processing and pattern creation. It uses a network, which is capable of learning from unsupervised data. Due to resemblance of this network with network of neurons in human brain, deep learning is also referred as deep neural learning or deep neural network [9]. This work utilizes the Convolution Neural Network Model (CNNM) for multi-class content based image search. The proposed approach is compared with Multi-Layer Perceptron Model (MLPM). The performance metrics shows that CNN model is superior to MLP model. The paper is distributed into seven sections including introduction section. In the next two sections, literature review and machine learning models are presented. Subsequent sections present the information about used multi-class image dataset and discuss performance metrics. A separate section presents simulation results and lastly conclusion is discussed.

## LITERATURE REVIEW

Researchers have offered wide range of schemes for content-based image search systems based on feature extraction methods. These features include colours, shapes, textures, moments, wavelets, statistical measures and many more. Colour based retrieval schemes are the widely used methods since it is invariant to image orientation or size. Colour histogram is used to calculate the distance measure such as Euclidian distance based on colour likeness in these schemes [10]. Many researchers for CBIS also use the shape features. Most of the times, shape is represented by the boundaries or edges in image which identifies the abrupt intensity variation. Canny, Kirch, Sobel, Prewitt and Robert are the most usually applied edge detection techniques [11]. Texture plays a central role as features in image retrieval systems. The texture features are categorized into two categories namely structural and statistical [12]. In statistical texture features, moments of histogram such as mean (average), standard deviation, kurtosis, skewness and other higher order moments are calculated for image matching [13]. Other methods in texture based image search are gray level co-occurrence matrix [14], energy [15], Gabor features [16], and so forth.

Multi-resolution transform based feature extraction schemes provides better performance for computer vision system that is proved by many researchers in last decade. Therefore, these approaches are extensively used for the texture classification in CBIS systems. Most commonly employed multi-resolution analysis techniques include the wavelet [17], curvelet transform [18], ridgelet transform [19] and counterlet transform [20].

After successful feature extraction distance measures [21] or machine learning based classification model are used to identify the similarities. In machine learning models, the most common models are based on neural network [22], k-nearest neighbour [23], support vector machine [24] and clustering [25].

These days deep leaning based classification models are used due to higher efficiency. It consists of input layer, which is followed by several hidden layers and at the end, there is one output layer. Innermost layer aggregate features from all the previous layers. So inner-most layer identifies the most complex feature. Each layer includes multiple neurons. On receiving input, neuron performs a weighted addition over its input and activation function is applied over the input to produce the output. Neuron of each subsequent layer repeat the same process of assigning weighted addition and applying of activation function. Output of the last layer/output layer symbolizes the prediction of the model. To identify the correctness of the model, a loss function is used. Loss function calculates the difference between true value and predicted value. This difference is called error rate and it is back propagated to reduce the error at every layer. This process of learning the deep neural network is repeated until error rate goes below a desired value [26].

## MACHINE LEARNING MODELS

This work utilizes two machine-learning models namely Multi-Layer Perceptron (MLP) and deep learning based Convolution Neural Network (CNN) to carry out the content based image search. The models are discussed briefly in following subsections.

### Multi-Layer Perceptron

A Multi-Layer Perceptron (MLP) is an artificial neural network has multiple neurons, arranged in multi-layers. This neural network model works on the principle of back-propagation learning procedure. A MLP has the capacity to learn nonlinear functions better than a perceptron that is able to learn only linear relationships in data [27]. This is first architecture that has with more than one layer. The layers are structured into three layers, which are referred as input, hidden and output layer. The input layer collects inputs and output layer is used to classify the inputs. The count of neurons in this layer in proportion of inputs receives by this layer. The middle layer i.e., hidden layer has varying number of neurons. This layer performs the mapping between input and output. It is proved in many applications that a single hidden layer with multiple neurons can establish impressible relationships [28]. The last layer produces the final output in terms of labels. The number of neurons in this layer are same as number of distinct classes in the data. This work uses MLP architecture with three traditional layers only. The number of neurons in input layer are $2352$ and output are $6$, as the input is our $28\ x\ 28\ X3$ image and the output is a 6x1 vector indicating the 6 classes.

 *© STM Journals 2020. All Rights Reserved* 

The 500 neurons in the hidden layer are used. ‘Adam’ optimizer is used that is an efficient variant of Gradient Descent Algorithm (GDA). Figure 1 displays the architecture of this network.

**Convolution Neural Network (CNN) Model**

Deep learning models are based on architecture of neural network. The term ‘Deep’ signifies the large number of hidden layers in the network. A standard neural network have $2-3$ layers while in deep network, it can go up to more than $three$ layers. It usages enormous set of labelled data to learn deep networks. Handcraft feature extraction is not required to learn these models [29].

A Convolutional Neural Network is one of the most noteworthy revolutions in the image processing domain. These classification models considerably better than customary machine learning methods and delivers the contemporary results. CNN is also termed as ConvNet model. It works on the principle of neural network that is based on feed-forward back propagation error. In recent times, outstanding accessibility of high computational power and enormous size of dataset, it is easy to use CNN classification models. It provides efficient performance in many image analysis applications. CNN learns valuable features straight from the learning images by the loss function optimization in contrast to the general machine learning schemes that need handcrafted features for training [30].

A CNN model principally consists of combination on one or more than convolutional layer, subsampling layer and dense layer [31], as displayed in Figure 2.

***Convolution Layers***

As the name proposes, the convolutional layer involves convolution as a basic building block. A convolution operation is performed to a region of an input image, transforming it into a single pixel after weighted sum. This process is repeated with every nonoverlapped region of image, to yield a new convolved image. The principle is that new values in the convolved image integrate information about the neighbouring pixels, hence reproducing how fine a feature is characterized in that particular area.

Convolution operation has two key elements, size and stride. The size represents the filter size i.e., number of pixels to be considered in a region. The size of filter $3x3$ operates on $3x3$ grids of pixels. The stride gives the gap between two regions. The stride $2$, signifies the gap between two regions, where the convolution is performed.

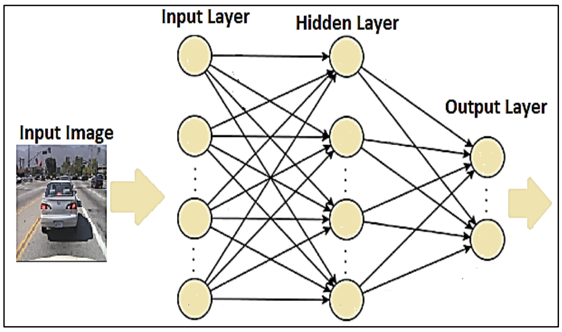


***Fig. 1:*** *Multi-Layer Perceptron Model for Multi-class Content-based Image Search System.*

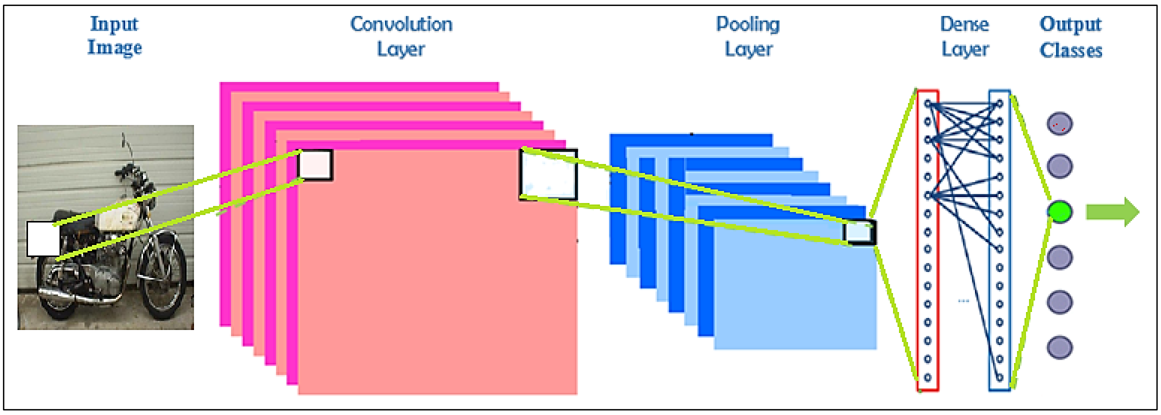


***Fig. 2:*** *Architecture of Convolution Neural Network for Multi-class CBIS.*

The size and stride of the convolution operation also regulates the size of the output image. For example, $3x3$ convolution with stride 1, converts a $5x5$ image to a $3x3$ image. It gives the grid of pixels over a weighted sum of the input pixels of region.

***Pooling Layers***

The pooling layers in CNN are used to decrease the spatial dimensions, without disturbing the depth. The main uses of pooling layer are

- To increase the computation performance by reducing the spatial dimension.
- To reduce the chances of over-fitting since reduced spatial information needs less learning parameters.
- To achieve translation invariance.

The widely used type of pooling is max pooling. A max pool layer with a size of $2x2$ reduces the height and width of image by half by selecting the maximum value of $four$ input values. Other examples of pooling operations are minimum pooling and average pooling.'

***Dense Layers***

The last layer of any CNN model is always a dense layer. The dense layer of the CNN does the classification part. It is also called as fully connected layers. The dense layer is generally used with a $sigmoid$ or $softmax$ activation function, for producing the result. Therefore, this layer predicts what a network has learned.

Additionally, activation functions are placed in between or at the end of layers in CNN. The activation function is the nonlinear mapping that is applied on the input signal. This transformed output is then passed to the subsequent layer of model as input. The widely acceptable activation function is Rectified Linear Unit (ReLU). Statistically, it is expressed as $y = max(0, x)$. ReLU is linear for all positive inputs, and zero for all negative inputs. It has fast convergence than other activations.

CNN model used in this content-based image search problem has $three$ convolutional layers. The size of first convolution layer is $64 - 3\ x\ 3$, the 2nd convolution layer has size $128\ - 3\ x\ 3$ and the final convolution layer has size $512 - 3\ x\ 3$. A subsampling layer using max pooling follows all these three convolution layers. The size of these max pooling layers is $2\ x\ 2$. The final layer of CNN model is a dense layer that has a $softmax$ nonlinear activation function with $six$ units. These six units in last layer are essential for this six-class multi-classification problem.

The model is compiled after design. The optimizer is the gradient descent algorithm based 'Adam' optimizer and cross-entropy loss is used to calculate the prediction error rate.

This optimizer uses backpropagation to update the weights of the neurons. It computes the derivative of the loss function regarding each weight and deducts it from the weight. A cross entropy loss function is utilized due to multi-class nature of the problem.

## TRAINING AND TESTING DATASET

A multi-class image dataset consisting of six classes is used in this work. This image dataset has total 3513 images with different number of images in the following six classes. Figure 3 presents the example images for all the six categories [32].

***Class I:*** Aeroplane (1074 samples);
***Class II:*** Car (451 samples);
***Class III:*** Background (526 samples);
***Class IV:*** Motorbikes (450 samples);
***Class V:*** leaves (186 samples);
***Class VI:*** Faces (826 samples).

All the colour images are reshaped to size $28x28$. After resizing all the images are normalized in range 0 to 1 after normalization. The motive for this transformation is simply that it is easier for the CNN to learn if the features values are not too large.

## PERFORMANCE EVALUATION AND RESULTS

Following performance metric are measured to evaluate the efficiency of the CBIS classification model [33].

**Precision (*P*):** Precision is calculated as:

$$P = T^+/(T^+ + F^+)$$

**Recall (*R*)**: Recall is defined as:

$$R = T^+/(T^+ + F^-)$$

© STM Journals 2020. All Rights Reserved

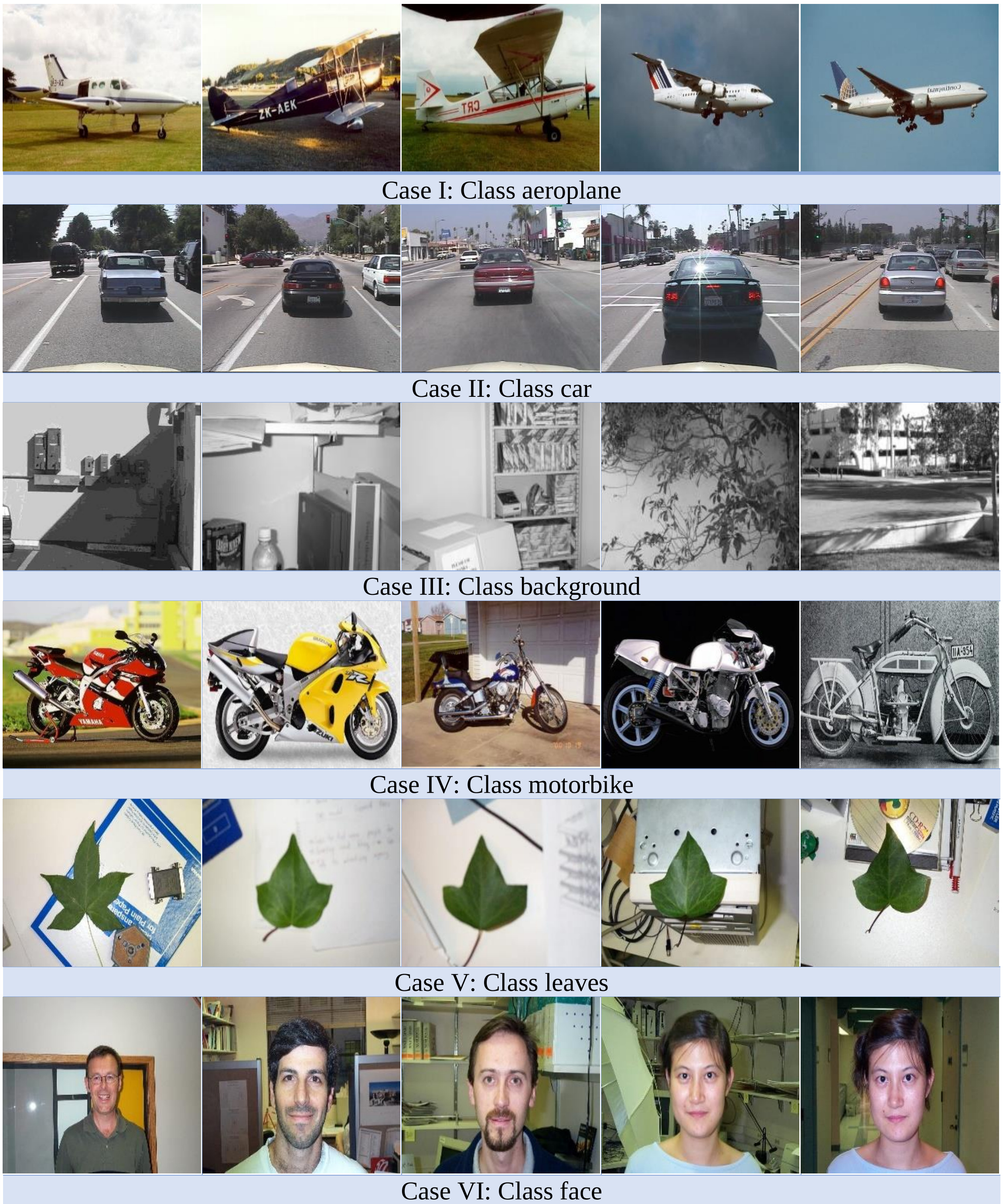


***Fig. 3:*** *Sample Images from each Class of Dataset.*

**F-Score ($f1$)**: this measurement is the weighted mean of $P$ and $R$. It is defined as:

$$\mathrm{F} = (2 * \mathrm{P} * \mathrm{R})/(\mathrm{P} + \mathrm{R})$$

**Accuracy:** Accuracy is the overall perfection of the CBIS model. It is defined as:

$$\text{Accuracy} = (T^{+} + T^{-})/(T^{+} + T^{-} + F^{-} + F^{+})$$

Here the terms $T^{+}$, $T^{-}$, $F^{+}$ and $F^{-}$ are True Positive, True Negative, False Positive and False Negative, respectively.

In a Receiver Operating Characteristic ($ROC$) plot, the True Positive Rate ($TPR$) is plotted as function of the False Positive Rate ($FPR$) for multiple thresholds. Individual point on the plotted curve indicates a $TPR/FPR$ pair associated to a specific decision threshold [34]. The area covered by ROC curve (referred as AUC) measures ability of a parameter to separate two classes.

## SIMULATION AND RESULTS ANALYSIS

Two separate experiments are carried out to perform multi-class CBIS as discussed in following subsections.

### *Experiment 1: Multi-Layer Perceptron for multi-class CBIS*

In this simulation, MLP model as discussed in the above section is used for multi-class CBIR. The extent of epochs represents the number of times the model is repeated through the data. The training of MLP model is iterated for 100 epochs (steps). The image dataset divided intotraining and validation subsets. 70% and 30% images are part of validation and training datasets, respectively. Therefore, 2459 images are in training dataset and rest images are in validation dataset. The complete image dataset is used as test set. Figures $4(a)$ and $4(b)$ plots the loss and accuracy for the training and validation data sets. It is evident from the curves that they are though nonlinear; however, it clarifies that the model is generalized well after learning. All the performance measures are provided in Table 1 individually for each class. The overall test loss and the test accuracy of the model are 0.048 and 0.94, respectively [35].

### *Experiment 2: Convolution Neural Network for Multi-class CBIS*

In this experiment, CNN model as discussed above is applied for content-based image search. 50 epochs are used for learning. The image dataset is parted into validation and training subsets. 70% and 30% images are the part of these two partitions. Therefore, 2459 images are in training dataset and rest images are in validation dataset. Complete image dataset is used as test set. A drop out layer is also used after each convolution layer with parameter 0.2. Dropout enables a network to learn more strong features that are valuable in aggregation with many altered random subsets of the neurons. Figures $5(a)$ and $5(b)$ plot the loss and accuracy for both the subsets. It is evident from the curves that they are though nonlinear; however, it demonstrates that the trained model is generalized well. All the performance measures are provided in Table 2 individually for each class. The overall test loss and the test accuracy of the model are 0.003 and 0.99, respectively.

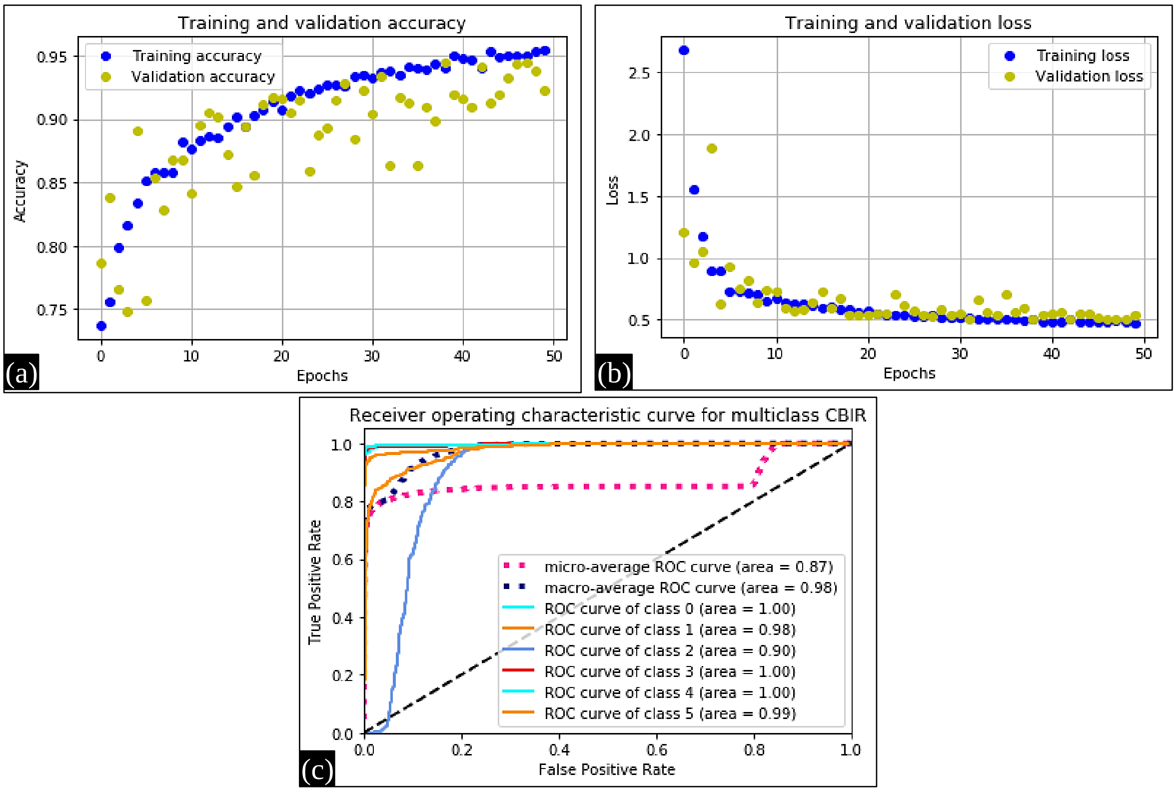


***Fig. 4:*** *(a) Plots of Training and Validation Accuracies for Simulation with MLP Model (b) Plots of Training and Validation Losses for Simulation with MLP Model (c) ROC Plot.*

 *© STM Journals 2020. All Rights Reserved* 

**Table 1:** *Classification Results for Experiment with MLP.*

| Classes/Metrics | $p$ | $r$ | $f1$ |
|---|---|---|---|
| Class Aeroplane | 0.76 | 0.97 | 0.85 |
| Class Car | 0.94 | 0.93 | 0.93 |
| Class Background | 0.40 | 0.34 | 0.45 |
| Class Motorbikes | 0.77 | 0.99 | 0.87 |
| Class Leaves | 0.99 | 0.94 | 0.96 |
| Class Faces | 0.87 | 0.97 | 0.92 |
| Average | 0.76 | 0.86 | 0.79 |

**Table 2:** *CBIS Classification Results for Experiment with CNN.*

| Classes/Metrics | $p$ | $R$ | $f1$ |
|---|---|---|---|
| Class Aeroplane | 0.99 | 0.99 | 0.99 |
| Class Car | 0.98 | 0.98 | 0.98 |
| Class Background | 1.00 | 1.00 | 1.00 |
| Class Motorbikes | 0.99 | 1.00 | 1.00 |
| Class Leaves | 0.99 | 0.98 | 0.99 |
| Class Faces | 1.00 | 0.99 | 0.99 |
| Average | 0.99 | 0.99 | 0.99 |

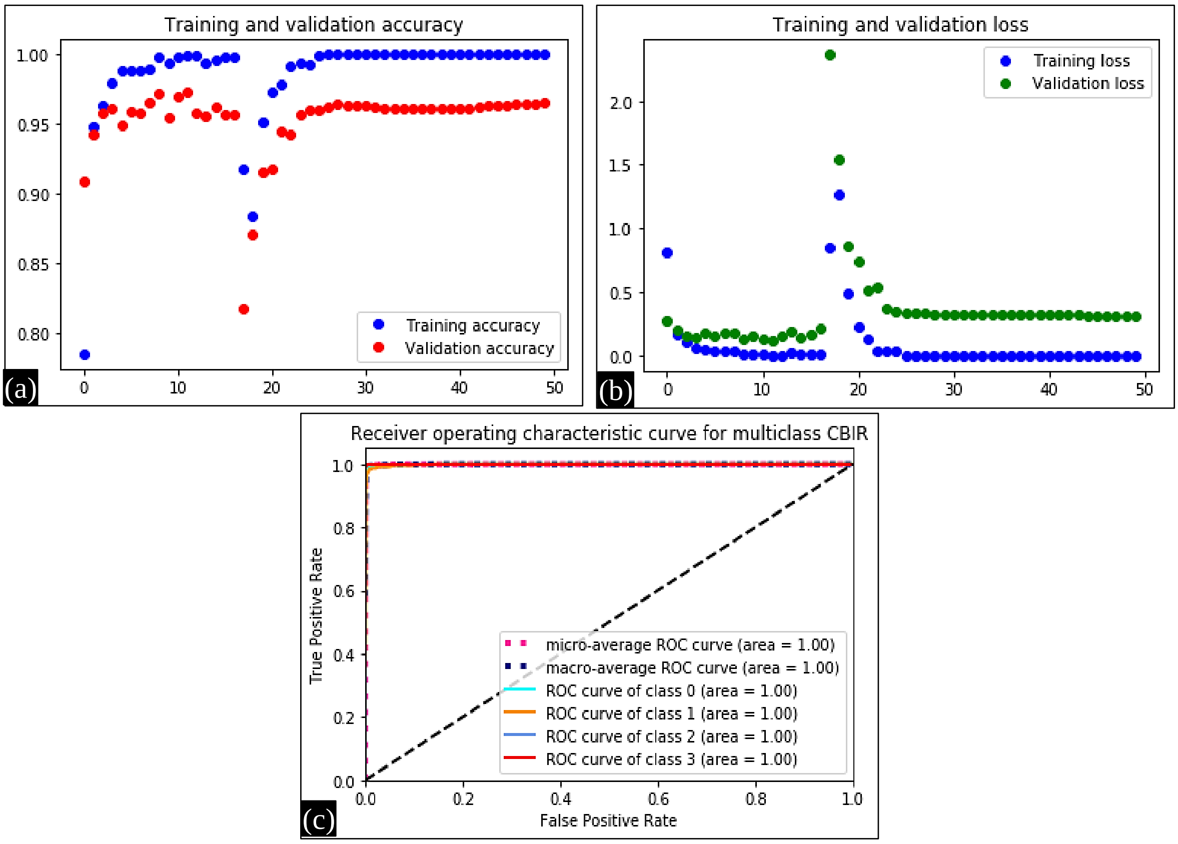


***Fig. 5:*** *(a) Plots of Validation and Training Accuracies for Simulation with CNN Model, (b) Plots of Training and Validation Losses for Simulation with CNN Model, (c) ROC Plot.*

It is evident from these results that the classification accuracy of CNN model is higher than the accuracy of multi-layer perceptron mode. ROC plots confirm the same.

## CONCLUSION

In this work, two different machine-learning models are analysed for content-based image search system. A multi-class image database with six classes is used to evaluate the models. To compare the results performance metrics precision, recall and f-score are calculated and to further confirm these results ROC curves are plotted. It is obvious from these metrics and plots that performance of convolution neural network-based model is better than the performance of multi-layer perceptron model. This work can be stretched by introducing more classes in future.

## REFERENCES

1. Smeulders AW, Worring M, Santini S, Gupta A, Jain R. Content-based image retrieval at the end of the early years. *IEEE T Patt Anal Mach Intell*. 2000; (12), 1349–1380p.

2. Müller H, Michoux N, Bandon D, Geissbuhler A. A review of content-based image retrieval systems in medical applications—clinical benefits and future directions. *Int J Med Info*. 2004; 73(1), 1–23p.
3. Iqbal K, Odetayo MO, James A. Content-based image retrieval approach for biometric security using colour, texture and shape features controlled by fuzzy heuristics. *J Comp Syst Sci*. 2012; 78(4), 1258–1277p.
4. da Silva Torres R, Falcao AX. Content-based image retrieval: theory and applications. *RITA*. 2006; 13(2), 161–185p.
5. Breiteneder C, Eidenberger H. Content-based image retrieval in digital libraries. In *Digital Libraries: Research and Practice, 2000 Kyoto, International Conference on*. IEEE, 2000, 288–295p.
6. Datta R, Li J, Wang JZ. Content-based image retrieval: approaches and trends of the new age. In *Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval*, ACM, 2005, November, 253–262p.
7. Gangopadhyay A. An image-based system for electronic retailing. *Decis Support Syst*. 2001, 32(2), 107–116p.
8. Deb S. Multimedia Systems and Content-based Retrieval. In *Multimedia Systems and Content-based Image Retrieval*, IGI Global, 2004, 1–13p.
9. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015; 521(7553), 436p.
10. Liu GH, Yang JY. Content-based image retrieval using color difference histogram. *Patt Recog*. 2013; 46(1), 188–198p.
11. Kavitha H, Sudhamani MV. Content-based image retrieval using edge and gradient orientation features of an object in an image from database. *J Intell Syst*. 2016; 25(3), 441–454p.
12. Hiremath PS, Pujari J. Content based image retrieval using color, texture and shape features. In *Advanced Computing and Communications, 2007. ADCOM 2007. International Conference on*, IEEE, 2007 December, 780–784p.
13. Tiwari S, Shukla VP, Biradar S, Singh A. Texture features based blur classification in barcode images. *Int J Info Eng Electron Bus*. 2013; 5(5), 34p.
14. Liu GH, Yang JY. Image retrieval based on the texton co-occurrence matrix. *Patt Recog*. 2008; 41(12), 3521–3527p.
15. Singhai N, Shandilya SK. A survey on: content based image retrieval systems. *Int J Comp Appl*. 2010; 4(2): 22–26p.
16. Zhang D, Wong A, Indrawan M, Lu G. Content-based image retrieval using Gabor texture features. *IEEE T PAMI*. 2000; 13–15p.
17. Tiwari S, Shukla VP, Biradar SR, Singh AK. Blur Classification Using Wavelet Transform and Feed Forward Neural Network. *Int J Modern Educ Comp Sci*. 2014; 6(4), 16p.
18. Tiwari S. A Pattern Classification Based approach for Blur Classification. *Indonesian J Elect Eng Info (IJEEI)*. 2017; 5(2), 162–173p.
19. Tiwari S, Shukla VP, Biradar SR, Singh AK. Blur classification using ridgelet transform and feed forward neural network. *Int J Image Graph Signal Process*. 2014; 6(9), 47p.
20. Arun KS, Menon HP. Content based medical image retrieval by combining rotation invariant contourlet features and fourier descriptors. *Int J Recent Trends Eng*. 2009, *2*(2), 35-39p.
21. Rubner Y, Tomasi C, Guibas LJ. The earth mover's distance as a metric for image retrieval. *Int J Comput Vis*. 2000, 40(2), 99–121p.
22. Park SB, Lee JW, Kim SK. Content-based image classification using a neural network. *Patt Recog Lett*. 2004; 25(3), 287–300p.
23. Xia Z, Wang X, Zhang L, Qin Z, Sun X, Ren K. A privacy-preserving and copy-deterrence content-based image retrieval scheme in cloud computing. *IEEE T Inf Foren Sec*. 2016, 11(11), 2594–2608p.
24. Chen Y, Zhou XS, Huang TS. One-class SVM for learning in image retrieval. In Image Processing, 2001. *Proceedings. 2001 International Conference on*, IEEE, 2001, 1, 34–37p.
25. Patil RV, Jondhale KC. Edge based technique to estimate number of clusters in k-means color image segmentation. In *Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE*

*International Conference on,* IEEE, 2010, July, 2, 117–121p.
26. Izadinia H, Russell BC, Farhadi A, Hoffman MD, Hertzmann A. Deep classifiers from image tags in the wild. In *Proceedings of the 2015 Workshop on Community-Organized Multimodal Mining: Opportunities for Novel Solutions,* ACM, 2015, October, 13–18p.
27. Kruse R, Borgelt C, Klawonn F, Moewes C, Steinbrecher M, Held P. Multi-layer perceptrons. In *Computational Intelligence,* Springer, London, 2013, 47–81p.
28. Zhang Z, Lyons M, Schuster M, Akamatsu S. Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron. In *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on,* IEEE, 1998, April, 454–459p.
29. Sahiner B, Chan HP, Petrick N, Wei D, Helvie MA, Adler DD, Goodsitt MM. Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images. *IEEE T Med Imag*. 1996; 15(5): 598–610p.
30. Simard PY, Steinkraus D, Platt JC. Best practices for convolutional neural networks applied to visual document analysis. In *null*, IEEE, 2003, August, 958p.
31. Hatanaka Y, Ogohara K, Sunayama W, Miyashita M, Muramatsu C, Fujita H. Automatic microaneurysms detection on retinal images using deep convolution neural network. In *Advanced Image Technology (IWAIT), 2018 International Workshop on,* IEEE, 2018, January, 1–2p.
32. Griffin G, Holub A, Perona P. Caltech-256 object category dataset, 2007 Available: Caltechedu. (2019). Caltechedu. Retrieved 1 October, 2019, from https://authors.library.caltech.edu/7694/1/CNS-TR-2007-001.pdf.
33. Li G, Yu Y. Visual saliency based on multiscale deep features. In *Proceedings of the IEEE conference on computer vision and pattern recognition,* 2015, 5455–5463p.
34. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Patt Recog.* 1997; 30(7), 1145–1159p.
35. Qiu Z, Zhuang Y, Yan F, Hu H, Wang W. RGB-DI Images and Full Convolution Neural Network-Based Outdoor Scene Understanding for Mobile Robots. *IEEE T Instru Measure,* 2018, 68(1), 18076-18088p.

**Cite this Article**
Shamik Tiwari. Multi-class Content-based Visual Information Retrieval using Multi-Layer Perceptron and Convolution Neural Network: A Comparative Study. *Journal of Artificial Intelligence Research & Advances.* 2020; 7(1): 1–9p.