# Comparing Different CNN Based Classification Models to Evaluate Performance On Natural Scene Images


1st Aditi Bharadwaj
*MSC Data Analytics*
*National College Of Ireland*
Dublin, Ireland
X19165552@student.ncirl.ie

2nd Mohit Jain
*MSC Data Analytics*
*National College Of Ireland*
Dublin, Ireland
X18200991@student.ncirl.ie

3rd Neelam Bist
*MSC Data Analytics*
*National College Of Ireland*
Dublin, Ireland
X19158769@student.ncirl.ie



***Abstract*—This study aims to classify the images of natural scenes using a convolutional neural network. An automizing technique of identifying and classifying the natural scenes will be useful in saving time to analyze scenes and accurately predict it. The Classification of Natural scenes will have many important applications in the defense field, physically challenged people, and many more. This paper proposed a CNN based technique through which we can identify and classify natural scenes such as buildings, streets, forests, glaciers, sea, and mountains. The dataset was initially given for the competition organized by Intel on the Kaggle platform and the total images are 24,335. The model has been trained on 14,301 images and tested on 3000 images whereas, a validation function has been introduced for 7301 images. The project is done on google collaboratory. For training, the model two convolutional neural network models are applied. CNN with 3 layers sequential model architecture and CNN with 4 layers sequential model architecture is applied. The models have been evaluated and compared based on two parameters accuracy and loss function. Data augmentation is done to overcome the overfitting problem. The accuracy of CNN with 3 layers is 82.13 percent whereas the accuracy of the CNN model with 4 layers is 77.8 percent. The final result shows that CNN with 3 layers outperformed CNN with 4 layers.**



***Index Terms*—CNN, Kaggle, Data Augmentation, Overfitting, Google Colaboratory.**


## I. Introduction

Scene classification is an open problem in computer vision and has applications in image and video database indexing. Natural scenes are considered to be a pleasant thing to the eye. It is also known to improve brain health and help in reducing depression and also it can increase the focus of the brain. Seeing the natural scene can help in many ways to improve our mental health. Computer vision has many applications and it can be used in various fields. The ability to see and understand the image or video by a computer is known as computer vision. Computer vision is an eye-catching field for every researcher. In this paper, we tried to identify the category of image based on the scene present in that for example classifying the images in a category like a forest, sea, street, glacier, etc present in the image. Computer vision has various applications in every filed. In today's era, everything is a machine-based human being is dependent on the machine for their day to day work. The scope of image classification is increasing day by day. Identifying the images and finding the pattern is a complex task and thus we choose to identify the pattern of images and classify them as forest, building, street, etc. Neural networks are very advance and give the best result in identifying or classifying any form of image, sound, text, video, etc.

The project is done on google collaboratory. It is a cloud service that provides free access to GPU and TPU and is widely used in machine learning. The data is accessed through google drive and then loaded through python command

Neural networks are the network consist of neurons and designed after the human brain. The way the human brain works the same way the neural network is built to identify or classify any form of data. There are various kinds of neural networks, different neural networks perform differently on every form of data. The task performed in this paper is image classification based on various images consist in the dataset. The study is performed using a convolutional neural network. The reason behind choosing CNN was that they give the best result in classification.

Convolutional neural networks are a class of deep neural networks and are the regularized form of multilayer perceptrons. CNN is often used for analyzing visual imagery. The model is trained using 14000 images. CNN with three-layers and CNN with four layers is used to perform the study, The result of both the model is compared.

## II. Litreature Review

[1]The proposed model begins with the Deep Convolutional Neural Network. This model consists of components for the identification of components, which are organized in layers. Distinct features are detected at the lowest layers and learned parameters are used to feed inputs to the next layer in the ranking order. The major characteristics of the Deep Convolutional Neural Network are inherently learning hierarchical component representations.[1] It states that if the primary layer components are generic and the final layer components are specified, there will be a transition from ordinary to unique in the network. As a result, generic components are obtained

using the CNN method. To apply the Transfer Learning technique, a CNN model based on VGG16 is pre-trained on ImageNet.

[2]In this study the author demonstrated the classification of images of Rocks. The author used SVM(support Vector Machine) algorithm for classification. The study shows classification of 9 different classes of Rocks extracted autonomously and the accuracy achieved is 96.71 percent. 3 layer CNN has been implemented and it shows significant improvement over results, 10 trails on the testing dataset was implemented. The study can outfit many practical situations in classifying natural landscape of rocks and the task has simplified in binary and 5 layers CNN which results in 89.43 percent accuracy.

[3]This study analyses a design of Deep Convolutional Neural Networks build on transfer learning concept for recognizing image. Deep CNN system was applied on large ImageNet dataset which consist of 1000 classes and 14 million images for learning the feature selection. This pre-trained model are forwarded to another situation of classification of images of UC Merced Land dataset which consist of 21 classes. The result shows accuracy of 87 percent.Various advances in the field of image recognition is dependent on Deep learning technology. Deep Learning is a component of Machine learning which employees Artificial Neural Networks with manifold layers, where top level components gets derived from Raw inputs. In this study the edges and curves of the images are determined by lower layers of neural networks and features such as trees, cars etc. are determined by higher layers of the network. The study use models such as VGG, Inception an Dense Net due to the high computational power and cost. To apply Transfer Learning the experiment have employed VGG16-based architecture of CNN which is pre-trained on ImageNet, carried our using torchvision. The final classifier has been reprocessed using Multilayer Perceptron using 3 fully connected layers and implementing ReLU activation function.The study the image of size 224*224 is used as input to the convolutional layer. The image surpasses through the filter of 3*3 size.

[4]The research is based on association of two models convolutional neural networkand shifted Legendre-Fourier moment's invariants . This combination obtains a model for image classification resulting in high accuracy considering small image datasets. The experiment consist of testing the invariance of characteristic for rotation and scaling function. 3 types of transformation was carried out like rotation, scaling and another rotation which resulted in output of 4 different images. The future work consist of increasing model accuracy and to apply SLFM algorithm in order to work on most significance to classify coloured images.

[5]The research is based on using transfer learning where pre-trained models are used for determining the solution . The experiment have used deep learning models such as Googlenet, Alexnet, VGG16, ResNet and DenseNet for image classification and VGG16 model shows highest performance.

[6]This paper address challenging problems of constructing huge image database. The paper propose wavelet analysis to extract the features of the image for classifying images. Seven classes of images are considered in the experiment with the total 2800 images. Wavelet analysis used for feature extraction proved to be very efficient the result shows 91 percent accuracy. As a result the experiment depict wavelet analysis performs better when organizing of huge images in the database.

[7]In this Study the author has proposed Radial basis function-based (RBF) methods to handle with image classification problems. The experiment shows there have been several drawbacks in RBF such as it was time consuming to train the model thus a self-organizing map (SOM) neural network is proposed. The RBF is proposed to expedite the training time. The study shows MRBF the modular RBF shows better performance than the traditional RBF neural network.

[8]In this study the author have discussed Convolutional Neural Network (CNN) . The Deep learning refers to Artificial Neural Network. CNN is convolution of matrixes. CNN has multiple layers such as convolutional layer, pooling layer ,fully connected layer and non-linearity layer. CNN shows effective performance while dealing with image data . CNN is widely used for Computer vision, Natural Language Processing and on largest ImageNet dataset. This research defines the important elements of CNN and how the efficiency of CNN is affected.

[9]In this research author has demonstrated and compared various solutions to the data augmentation issues in image classification. The experiment is carried out on ImageNet dataset and comparison on data augmentation methods are carried out. Neural augmentation was finally proposed in the experiment to improve the classification. Generative Adversarial Nets (GANs) Has been influential technique to generate unsupervised new images for training .Furthermore a style transfer can be applied to augment data for unbalanced dataset.

[10]In this study two major and well known classifier K-Nearest -Neighbor(KNN) and Support-Vector-Machine (SVM) classification are used to compare the results and outcome. The experiment predicts SVM classifier performs better than KNN classifier. In the future work such comparison must be performed using multiple objects and apply more sophisticated classifier.The experiment is been conducted using bags of words(BoW) model. The KNN classifier compares the histogram generated from training images to this histogram. In comparison the SVM uses learned model from training set and uses the histogram from test set for prediction. The objective of BoW is representation, which deals with image representation and feature detection.

## III. Methodology

This study is conducted using CRISP-DM methodology as the dataset which is choosen for the study has bussiness prospective.

Crisp-dm stands for cross industry process for data mining. It is a methodology which provides a structure for planning the data mining project. This methodology is used more often for planning the data mining project. This methodology has 6 steps:

- Business Understanding
- Data Understanding
- Data preparation
- Modelling
- Evaluation

### *A. Business Understanding:*

This study is conducted on identifying and classifying natural scenes. Image classification is a very complex task and thus identifying natural scenes is difficult. Recognizing and classifying natural scenes is unique and helpful for understanding the pattern of particular objects. This model can be useful for blind people by just adding the audio to it. It can be used in various existing items for blind people. An app can be made for blind to sense and feel the natural scenes. In autonomous cars, this model can be used and cars can identify the building, forest, sea, etc and accidents can be avoided. It will help increase the business in different aspects in different fields.

### *B. Data Understanding:*

The data has been downloaded from Kaggle. The dataset was first published on the analytics Vidhya website. This dataset is used by Intel for the competition which was organized by Kaggle. The dataset consists of 24,335 images. It has three folder train, test, and validation in which there are 14,301, 3000, 7301 images respectively. In each folder, there are 6 separate folders i.e building, street, forest, glacier, mountain, sea. All folders have the images respectively and the validation folder contains the images from every category to validate it. The original size of the images is 150*150.

### *C. Data Preparation:*

To access the data through the drive, a zip file has been made of data. The data is then uploaded in the drive by giving access to my drive through python command. The zip file is then unzipped in google collaboratory by python command. The data is then loaded to preprocess and apply the model. Data cleaning is an important task for a data mining project, However, in image classification cleaning is a bit different from other models. In image classification preprocessing of the image is done converting the image into compatible size by reducing the resolution of pixels. For this study image, pre-processing is done in multiple steps

First, the images were labeled into all different categories respectively. after labeling the images the images were resized and converted to 150*150*3 resolution. The training images are normalized in a range(0,1). After training the model the problem of overfitting is visualized and to avoid overfitting data augmentation is done in which multiple copies of images are made with the rotation view. The accuracy is improved after applying data augmentation.

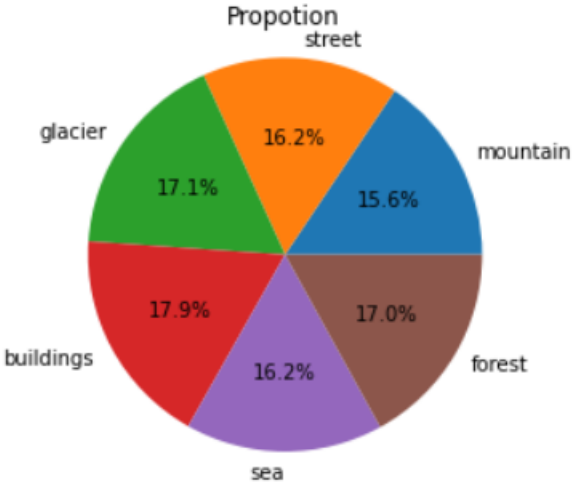


Fig. 1. Pie Chart of Data Distribution.

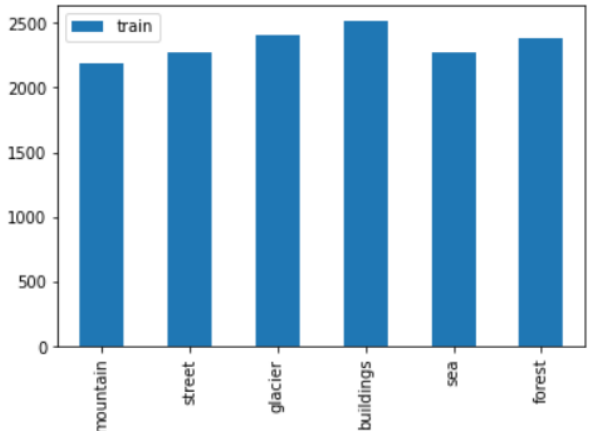


Fig. 2. Category wise Bar Graph Of Images .

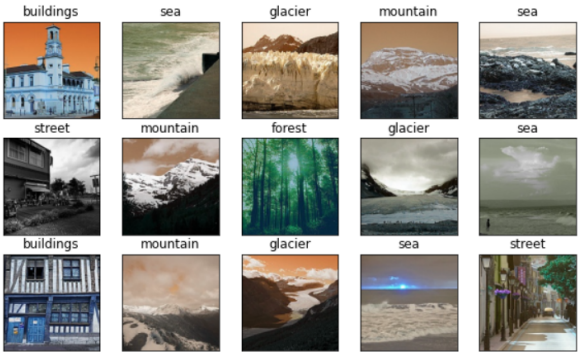


Fig. 3. Loaded Data for Training.

### D. *Modelling:*

This paper proposes image classification based on different natural scene images. The deep neural network has applications in various fields and is known for good and accurate predictions so in this paper we applied 2 convolutional neural networks one with 3 layers and one with 4 layers. The convolutional neural network gives good results on image classification and is mostly used for image classification that is the reason we choose CNN to perform this study. A convolutional neural network has different parameters including kernel size, filters, activation function, and input size. Filters are introduced for a better understanding of images it slices the images and learns the small portion of the image by sliding left to right across the image from top to bottom. A kernel is a matrix that moves over the input data and connect the dots and gives the product of dots as output data. The most commonly used kernel size is 3. The activation function is introduced to decide to weather the neuron should pass or not. The most common activation function for image classification is the rectified linear unit (relu). For compiling the model optimizer is used in order to reduce the losses and get a more accurate result. The loss function is a prediction error of the neural net and is introduced in the model to check the error in the CNN model. 3 layer CNN Sequential Model Architecture and 4-layer Sequential Model Architecture are developed and their performance is compared.

In 3 layer model architecture, the Conv2D layer with filter layers of 64, 32, and 64 is used with a kernel size of 3X3. During training, the model 30 percent of the data is used for validation. A max-pooling layer is applied to the model to reduce the dimensions of the data by combining the clusters of neurons into one neuron at the next step. One flatten layer is added to the model to convert the data into 1 dimensional in order to send it to the next output layer. The feature extraction is done using the conv2d layers and at last dense layers are applied to extract all the features and combine them as the output and classify the image according to their category. While compiling the model ADAM optimizer is used and cross-entropy loss function is used as it gives the best result for the classification of images.

After applying the model problem of overfitting is identified to overcome the overfitting problem data augmentation is done and after that, the result is improved. The data is augmented based on zooming the image and flipping the image horizontally and rotating the image. After augmenting the images the model is again run and the accuracy is improved.

In 4 layer model architecture Conv2d layer with filter layers of 32, 64, 128, and 32 is used in 4-layer CNN architecture with kernel sizes of 3X3. A max-pooling layer is applied to reduce the dimension and a dropout layer is applied for regularization in order to avoid overfitting. The activation function relu is used and to compile ADAM optimizer and cross-entropy loss function is used.

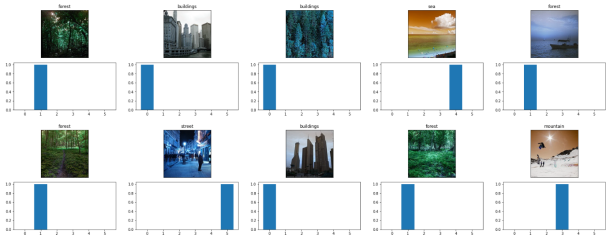

Fig. 4. Images and Graph of the Model Prediction.

## IV. Evaluation:

The models are compared and evaluated on the basis of two parameters accuracy and logarithmic loss function. The data is almost equal for every class that is the reason for choosing accuracy as evaluation metrics. The data consist of multi-class images and logarithmic loss works well for multi-class classification. The lesser the loss the greater the model performs.

When it comes to image classification we generally talks about the human face or animal face and talks about their feature extraction, but in this paper we are going to determine natural scene like mountains, forest, street, glacier, sea, and buildings. In this paper, we have used a residual network-based image extraction feature that uses scene recognition to provide low-level and high-level features, where low-level features are original image pixel-based attributes, and focus on subtle image texture information. High-level image features are extracted by low-level modelling and provide valuable semantic knowledge.

CNN with a 3 layers model gives accuracy 76 percent at the beginning. The problem of overfitting is visualized and because of overfitting the model performed poorly. The data is augmented to overcome the overfitting problem and then again the model is applied and the model gives 82.13 percent accuracy and the loss is 0.4634, Whereas CNN with 4 layers gives 77.8 percent accuracy and the loss is 0.606. CNN with 3 layers outperforms CNN with 4 layers. The figures shown below shows the graphs for both the models the accuracy and the loss function. The graphs clearly shows that CNN with 3 layer after applying the augmentation to the data performs better than CNN with four layer.

## V. Conclusion and Future Work

This research analyses two different deep learning models. Data set consists of natural scenes images is used to build a neural network to identify and classify the images. Two models are applied to the data set for classifying the images according to their categories CNN with 3 layers and CNN with 4 layers. The data is trained on 14k images and tested on 3k images. For overcoming the over-fitting problem data augmentation is done on the training images. The models are compared and evaluated on the basis of two parameters accuracy and logarithmic loss. The accuracy for CNN with 3 layers is 82.13 and the loss is 0.46, Whereas the accuracy and loss for CNN with 4 layers are 77.8 and 0.60 respectively. The above figures

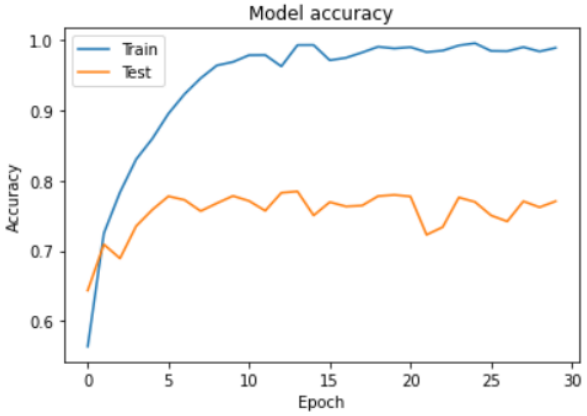


Fig. 5. Accuracy graph before applying augmentation.

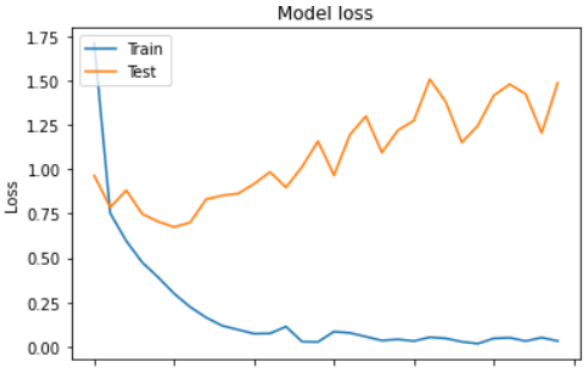


Fig. 6. Loss graph before applying augmentation.

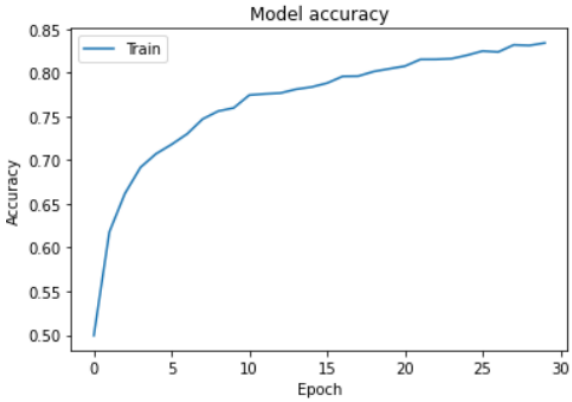


Fig. 7. Accuracy Graph of CNN-3 With Augmentation.

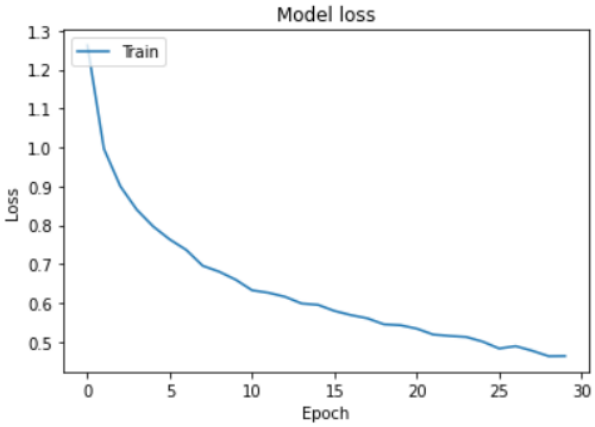


Fig. 8. Loss Graph of CNN-3 With Augmentation.

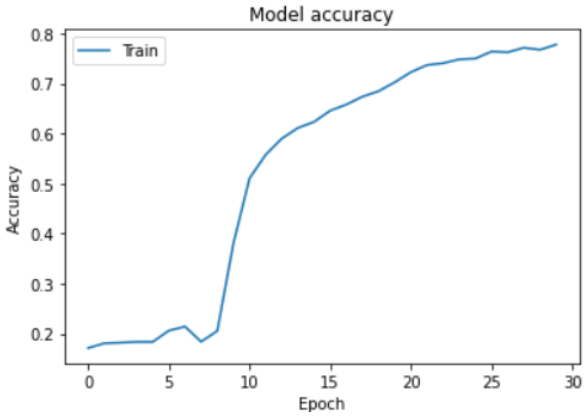


Fig. 9. Accuracy Graph of CNN With 4 Layers.

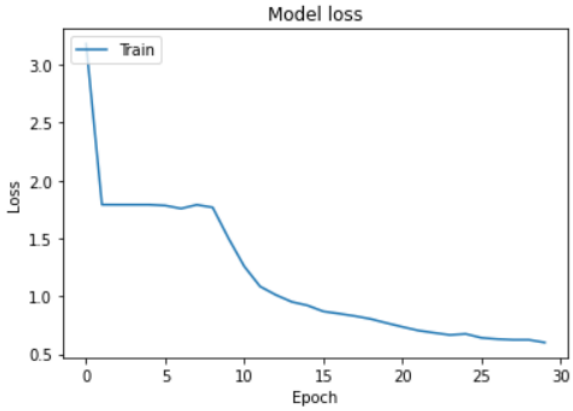


Fig. 10. Graph of Loss of CNN With 4 Layers.

shows the accuracy and loss graphs for both the models. CNN with 3 layers gives good results as compared to CNN with 4 layers.

The findings from this research unriddle the opportunity for wide future work. In the future, the accuracy of the model can be improved by applying different models and by using conv3D layers for CNN. Other data normalization techniques and different data pre-processing can be used to improve model performance. More accurate results can be achieved by increasing the epochs and other models.

An application can be made for the visually challenged people in the future using this classification model.

## REFERENCES


[1] Papers.nips.cc, 2020. [Online]. Available: https://papers.nips.cc/paper/5347-how-transferable-are-features-in-deep-neural-networks.pdf. [Accessed: 05- Aug- 2020].

[2] "Towards Natural Scene Rock Image Classification with Convolutional Neural Networks - IEEE Conference Publication", Ieeexplore.ieee.org, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/8861885. [Accessed: 05- Aug- 2020].

[3] "A Deep CNN Approach with Transfer Learning for Image Recognition - IEEE Conference Publication", Ieeexplore.ieee.org, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9042173/references. [Accessed: 05- Aug- 2020].

[4] "Image Classification using Shifted Legendre-Fourier Moments and Deep Learning - IEEE Conference Publication", Ieeexplore.ieee.org, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/8931326. [Accessed: 06- Aug- 2020].

[5] "Multiple Classification of Flower Images Using Transfer Learning - IEEE Conference Publication", Ieeexplore.ieee.org, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/8875953. [Accessed: 06- Aug- 2020].

[6] W. Zou, Z. Chi, and K. C. Lo, 'IMPROVEMENT OF IMAGE CLASSIFICATION USING WAVELET COEFFICIENTS WITH STRUCTURED-BASED NEURAL NETWORK', Int. J. Neur. Syst., vol. 18, no. 03, pp. 195–205, Jun. 2008, doi: 10.1142/S012906570800152X

[7] Chuan-Yu Chang and Shih-Yu Fu, 'Image Classification using a Module RBF Neural Network', in First International Conference on Innovative Computing, Information and Control - Volume I (ICICIC'06), Aug. 2006, vol. 2, pp. 270–273, doi: 10.1109/ICICIC.2006.295.

[8] S. Albawi, T. A. Mohammed, and S. Al-Zawi, 'Understanding of a convolutional neural network', in 2017 International Conference on Engineering and Technology (ICET), Aug. 2017, pp. 1–6, doi: 10.1109/ICEngTechnol.2017.8308186.

[9] L. Perez and J. Wang, 'The Effectiveness of Data Augmentation in Image Classification using Deep Learning', arXiv:1712.04621 [cs], Dec. 2017, Accessed: Aug. 09, 2020. [Online]. Available: http://arxiv.org/abs/1712.04621.

[10] M. Sarrafzadeh, 'Department of electrical engineering and computer science', SIGDA Newsl., vol. 20, no. 1, p. 91, Jun. 1990, doi: 10.1145/378886.380416

[11] W. Byeon, T. M. Breuel, F. Raue, and M. Liwicki, 'Scene labeling with LSTM recurrent neural networks', in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, Jun. 2015, pp. 3547–3555, doi: 10.1109/CVPR.2015.7298977

[12] B. Pan, Z. Shi, and X. Xu, 'R-VCANet: A New Deep-Learning-Based Hyperspectral Image Classification Method', IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing, vol. 10, no. 5, pp. 1975–1986, May 2017, doi: 10.1109/JSTARS.2017.2655516.

[13] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, 'PCANet: A Simple Deep Learning Baseline for Image Classification?', IEEE Trans. on Image Process., vol. 24, no. 12, pp. 5017–5032, Dec. 2015, doi: 10.1109/TIP.2015.2475625.

[14] S. B. Park, J. W. Lee, and S. K. Kim, 'Content-based image classification using a neural network', Pattern Recognition Letters, vol. 25, no. 3, pp. 287–300, Feb. 2004, doi: 10.1016/j.patrec.2003.10.015.

[15] K. Tan, J. Hu, J. Li, and P. Du, 'A novel semi-supervised hyperspectral image classification approach based on spatial neighborhood information and classifier combination', ISPRS Journal of Photogrammetry and Remote Sensing, vol. 105, pp. 19–29, Jul. 2015, doi: 10.1016/j.isprsjprs.2015.03.006

[16] S. Lazebnik, C. Schmid and J. Ponce, "Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories," 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), New York, NY, USA, 2006, pp. 2169-2178, doi: 10.1109/CVPR.2006.68.

[17] H. R. Roth et al., "Anatomy-specific classification of medical images using deep convolutional nets," 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), New York, NY, 2015, pp. 101-104, doi: 10.1109/ISBI.2015.7163826.