2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) | 979-8-3503-9926-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICACITE57410.2023.10183255

Image Classification using Deep Learning: A Comparative Study of VGG-16, InceptionV3 and EfficientNet B7 Models

Shivam Aggarwal
Chitkara University Institute of Engineering & Technology
Chitkara University
Punjab, India
shivam0017.cse19@chitkara.edu.in

Ashok Kumar Sahoo
Graphic Era Hill University
Dehradun,Uttarakhand,India
ashoksahoo2000@yahoo.com

Chetan Bansal
Chitkara University Institute of Engineering & Technology
Chitkara University
Punjab, India
chetan0065.cse19@chitkara.edu.in

Pradeepta Kumar Sarangi
Chitkara University Institute of Engineering & Technology
Chitkara University
Punjab, India
pradeeptasarangi@gmail.com

Abstract— Image classification is the process of identification and classification of an input image or visual from a predetermined set of labeled images. This work comes under computer vision and machine learning. This work is dedicated towards image classification using deep learning by comparing various classification models based on their architecture performance and accuracy. Convolutional Neural Network (CNN) is a type of Artificial Neural Network (ANN) that learns the structural representation of an image and make predictions about its contents. The main layers of the model, known as the Convolutional layers. It detects and analyzes specific patterns in the image using filters. In this research development, image classification is done on several images of butterflies and spiders. The data set has been acquired from the ImageNet dataset. The technological progress and innovation include the observations of different CNN models to find out which one gives the most accurate outputs, i.e., identifying the type of insect in the input image. The focus has been on image recognition, data set size, classification techniques, and recognition accuracy. Three machine learning models such as VGG-16, InceptionV3 and EfficientNetB7 have been implemented in this work and the accuracies observed from the models are 97.67 %, 97.2 % and 99 % respectively. From the results it can be seen that EfficientNetB7 was found to be the best among the three models.

Keywords –classification technology, Convolutional Neural Networks, accuracy, deep learning,

I. INTRODUCTION

We sometimes need to process a large amount of data in images format and also need to classify them according to specific classes. This is not feasible manually, so we need the assistance of a computer program that can study and learn from the data provided, and give us some accurate and expected results. The task of image classification uses an algorithm that is trained to recognize and categorize objects or visuals in digital images. Machine learning algorithms, such as CNNs, are commonly used for image classification because they can automatically learn features from raw image data, form feature maps and make predictions based on those features [1]. There are various CNN models that can be used for our task. We wanted to study which CNN model is the best for image recognition tasks at present.

The basic structure of a Convolutional Neural Network consists of two main blocks. The first block extracts features (or patterns) from the input image and is the most unique characteristic of a CNN [2]. It generally includes one or more convolutional layers, and a pooling layer. Convolutional layers perform certain operations on the input image to extract various features from the input image using small filters, which finally results in a feature map. The filter size and the number of filters utilized in each convolutional layer can be adjusted to control the level of abstraction and complexity of the features learned by the CNN. Finally, all the aspects captured from the feature maps by the filters are put together in the form of a vector that defines the input for the second block. The second block is common to most of the neural networks used for classification. It aims to extract complex features from the input image. The output from the former block is passed as input to the initial convolutional layer of the second block as the feature map is passed through the network. This layer of the second block puts a set of filters to the input image, which is trained to learn and recognize several features or patterns in the image. The output of the convolutional layer is then passed through ReLU (explained in section V), which is an activation function that introduces non-linearity into the network. This output is then passed through further layers having their own set of filters, to extract more complex features from the image. There are additional filters in these layers as compared to the first block, and these filters are larger in size to recognize more global features in the image. In the second block, a pooling layer is generally used after the final convolutional layer, to reduce the dimensions of the output. This layer reduces the structural resolution of the output of the convolutional layers and helps to make the network more technically efficient.

VGG-16, InceptionV3, and EfficientNet B7 are all deep-learning models commonly used in image classification tasks. While they share similarities, they differ in their architecture, working, performance, and accuracy. VGG-16 is a CNN architecture that consists of 16 convolutional layers and three fully connected layers. The architecture is characterized by the use of small convolution filters (3x3). This helps the network to learn features at different scales and makes it more robust to image variations. VGG-16 (fig. 1) is a relatively simple architecture that is easy to understand and implement.

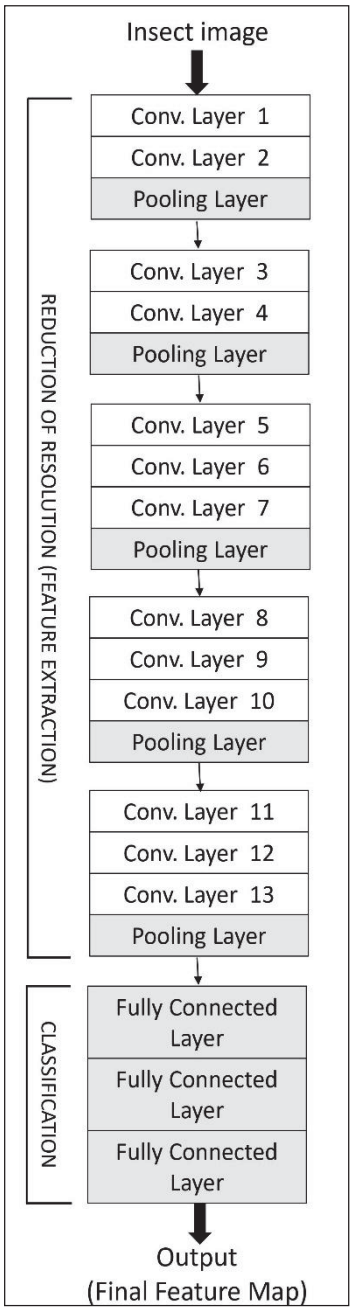

Fig. 1. Architecture of VGG-16 [3]

InceptionV3 (fig. 2) is a convolutional neural network architecture that uses a combination of convolutional filters of various sizes to recognize features.

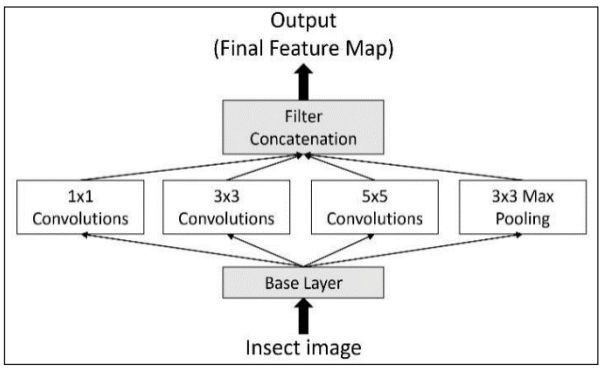

Fig. 2. Layer structure of InceptionV3 [4]

EfficientNet B7 is a type of CNN model. The number of parameters used in EfficientNet B7 is higher than InceptionV3 and lower than VGG16. This makes it faster than VGG16 and higher accuracy than InceptionV3. The architecture of an EfficientNet B7 model is given in fig.3.

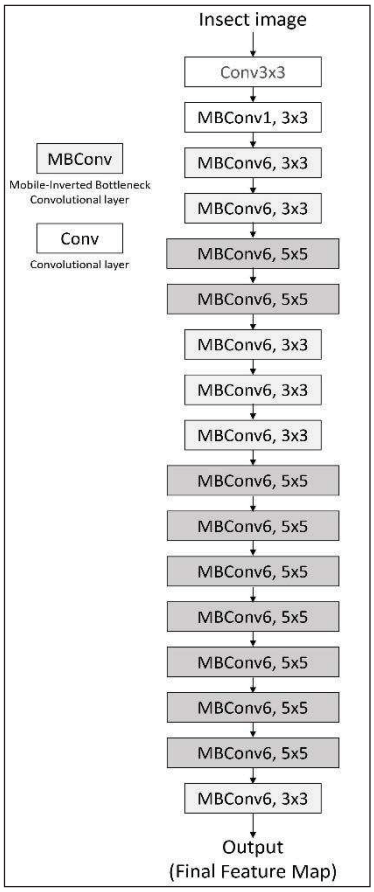

Fig. 3. Architecture of EfficientNet B7 [5]

All the three models (VGG-16, InceptionV3, and EfficientNet B7) are very effective in image classification. The only difference is in their architecture, working, and performance.

II. Literature Review

The authors [1] in their work have implemented a CNN model in classifying citrus fruit images. Using VGGNet 16 model the authors have achieved high accuracy in detecting the severity level of citrus fruit diseases.

In another work [2], the authors proposes a transfer learning approach for the classification of powdery mildew wheat disease from leaf images. It used a pre-trained convolutional neural network (CNN) with 450 wheat images, where each image is labeled with the corresponding class of powdery mildew disease. This pre trained model is then applied to another CIAGR dataset which demonstrated the effectiveness and improved results of transfer learning.

In [6], the paper presents a target tracking and recognition system that uses deep learning techniques to recognize and track different types of targets, such as vehicles, pedestrians, and bicycles. This study used dataset of images and videos captured by UAVs to train and evaluate their deep-learning models. In this paper, the authors used CNNs to classify and track different types of targets in real time. It also

demonstrates several layers of processing like the feature layers and the model layers which has the applications for feature extraction and image recognition for this study.

In [7], the paper is not directly related to deep learning models. However, it does compare the performance of 3D local descriptors with state-of-the-art methods in target recognition, including deep learning approaches. The paper provides insights into the potential of using 3D local descriptors in combination with deep learning models for target recognition in LIDAR (Light Detection and Ranging) missiles. Further research could explore the use of these local descriptors as input to deep learning models to improve the accuracy of target recognition. They generally use image hues, texture, and other information to compute various features from images.

Use of Gaussian mixture models (GMMs) is reported by the authors in their work [8]. The dataset included 800 samples for all target types. The study represents the use of GMM and feature extraction to reduce the impact of weather conditions on the target image. This study includes three stages, first is data collection through a UWB transceiver equipment, second stage involves feature extraction and then finally a variety of classifiers to classify the targets. Further it shows the potential use of image classification algorithms after feature extraction to achieve better results.

Besides these, there are a number of works published by various researchers such as implementation of residual learning framework [9], introduction of AlexNet architecture demonstrating the potential of CNN in computer vision tasks [10], overview of various CNN architectures (LeNets, VGG, AlexNet, ResNet) and their performance on various datasets [11], application of transfer learning in improvement of CNN architecture for image recognition tasks [12], transferability of features learned by CNN for image recognition [13], and performance analysis of GoogLeNet architecture on various datasets [14], and rethinking model scaling in convolutional neural network [15]. Summary of some of the notable works is given in table 1.

TABLE I. SUMMARY OF SOME NOTABLE WORKS

Ref No.	Dataset	Technique	Conclusion
[1]	Images of citrus fruits with disease	CNN	Highlights the importance of using deep learning techniques for image classification tasks, especially in the domain of agriculture and food security.
[2]	Images of wheat leaf	CNN	Highlights the importance of using transfer learning techniques for image classification tasks, especially in the domain of agriculture and plant disease diagnosis.
[5]	Images of Vehicles, pedestrians, and animals	CNN	The proposed system demonstrates the potential of using deep learning and machine learning techniques for target tracking and recognition in UAV-based applications.
[6]	3D point cloud data of vehicles and buildings	CNN	It has applications in extracting features from images which can be input to deep learning models for object classification, tracking, and recognition.
[8]	Flower dataset	CNN	Shows relevance in the broader field of computer and vision and relevance in target recognition in images.

III. OBJECTIVE AND METHODOLOGY

The main objective of this work is to recognize the input images using various image recognition classification techniques, by using different pre-trained models and finding out their test accuracies, and to find which one is the most efficient image classification model among them. The methodology diagram is given in the fig. 4.

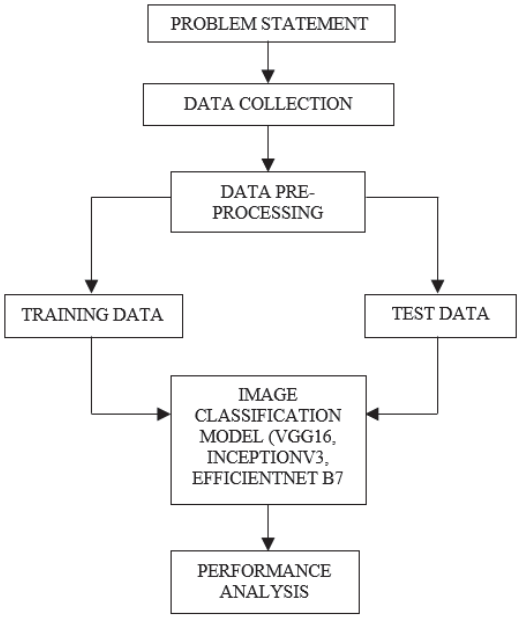

Fig. 4. Implementation Flowchart

IV. DATASET

This work uses a common dataset for all the pre-trained models to determine their performance.

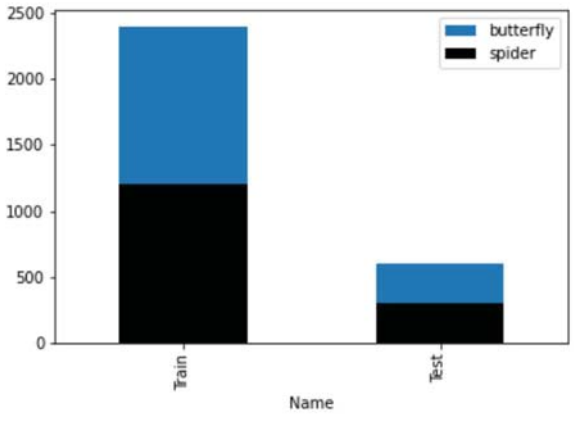

Fig. 5. Dataset distribution bar chart

The dataset contains about 7K medium-quality animal images belonging to 2 classes of insects, that is, Spiders and Butterflies. All the images have been collected from Kaggle [16] and have been checked by a human. The main directory is divided into folders one for each category, each having a

specific type of image. Fig. 5 is a graph depicting the distribution of data for training and test directories which is divided in an 80:20 ratio. Some of the sample inputs are given in the fig. 6.

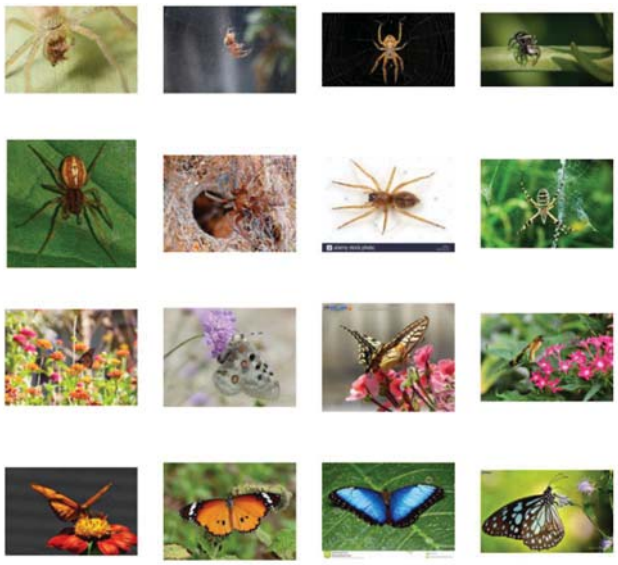

Fig. 6. Sample Dataset

V. Implementation And Result Analysis

The overall implementation is carried out in mainly 4 steps. Firstly, we will define the training and validation sets for the models. After that is complete, we load the base model. These models are designed to handle up to 1000 classes, but since this is just a binary classification, we will be using only the basic models. After making changes to the final layer of the model we will compile it and perform operations on it that is the flattening of the output to just one dimension and adding a fully connected layer with hidden units and ReLU (rectified linear unit) activation, while also specifying the dropout rate (is used to avoid overfitting and to improve the performance of the models) and RMSProp.

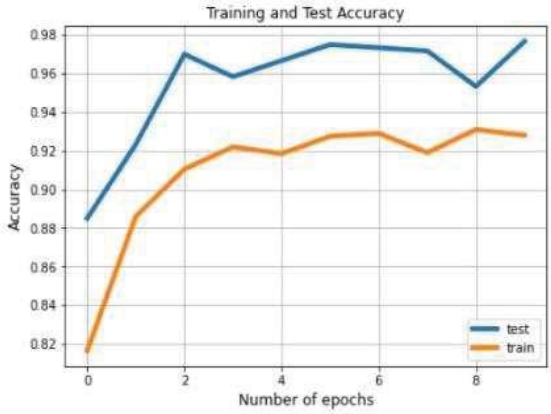

Fig. 7. Training v/s Test Accuracy graph for VGG-16 model

RMSProp (Root Mean Square Propagation), is an optimization algorithm used in deep learning to update the parameters of a neural network during the training process. It is commonly used in deep learning for its ability to converge faster and being more reliable than other optimization algorithms. It uses an adaptive learning rate, which means that it acts as a moving average filter that considers previous the gradients while updating the learning rate. If a parameter has large and consistent gradients, the algorithm will decrease its learning rate to prevent overshooting the minimum of the cost function.

The results in the fig.7 are of VGG-16 which shows a test accuracy of 97.67 % and a graph of training and test accuracy v/s no. of epochs.

The results in the fig. 8 are of InceptionV3 which shows a test accuracy of 97.20 % and a graph of training and test accuracy v/s no. of epochs.

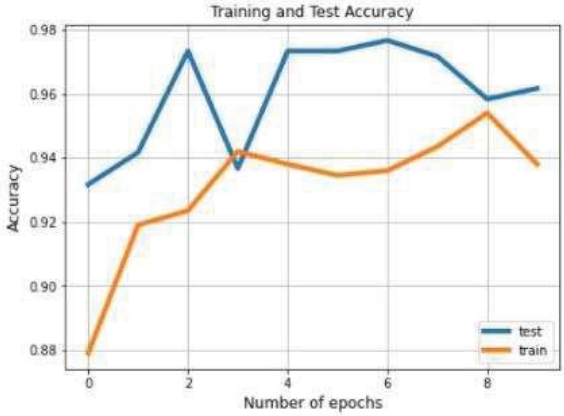

Fig. 8. Training v/s Test Accuracy graph for InceptionV3 model

Fig. 9. Training v/s Test Accuracy graph for EfficientNet B7 model

The results in the fig. 9 are of EfficientNet B7 which shows a test accuracy of 99.00 % and a graph of training and test accuracy v/s no. of epochs.

A reliable image classification model must exhibit high accuracy in both training and testing. Overfitting may be indicated if there is a significant difference between the two accuracies, with test accuracy falling behind. The graph depicting the training and test accuracy over epochs is a crucial tool for monitoring model performance during training and for detecting potential issues, such as overfitting. The graph plots the accuracy of the model on the training and test set against the number of epochs. At the start of training, both the training and test accuracy will typically be low. As the number of epochs increases, the training accuracy will generally increase, while the test accuracy may plateau or even decrease if the model is overfitting.

Overfitting happens when a model fits too closely or exactly to a limited set of data points and fails to perform well on new or unseen data. As the model is trained for more epochs, it is expected to learn, to make more accurate predictions on the training set, which should also translate into enhanced performance on the test set.

VI. Conclusion

Image classification is a challenging task hence, all of it depends upon the quality of the input data and the classification algorithm. In this study, we compared our existing models on a common dataset of butterflies and spiders. From the results shown in the previous section, we can see that the accuracy of EfficientNet B7, i.e., 99%, is the highest among all the three CNN models, with just 10 epochs, achieving both higher accuracy and better efficiency over other existing models. The accuracy comparison of all models is given in the table 2.

TABLE II. Accuracy Comparison Of Models

Model	Accuracy
VGG-16	97.67 %
InceptionV3	97.2 %
EfficientNet B7	99 %

The graphical representation of accuracies of all models is given in the fig. 10.

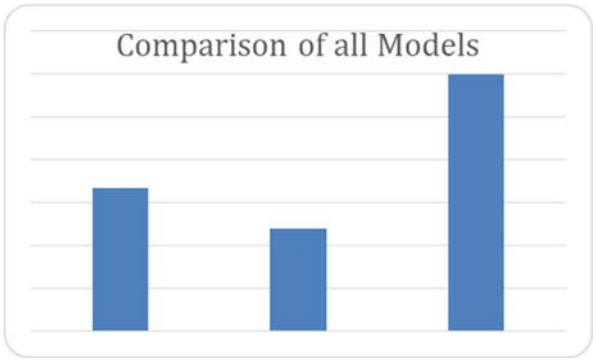

Fig. 10. Accuracy comparison of all models

The choice of the model depends on the specific requirements of the task and the resources available to us. However, image classification is a growing spectrum and there is always a new innovation to look forward to and push the boundaries further. By applying significant adjustments to model efficiency, we can expect these image classification models can serve as a foundation for future computer vision tasks.

References

[1] P. Dhiman, V. Kukreja, P. Manoharan, A. Kaur, M.M. Kamruzzaman, I. B. Dhaou, and C. Iwendi, “A novel deep learning model for detection of the severity level of the disease in citrus fruits”, Electronics, 11(3), 495, 2022.

[2] D. Kumar, and V. Kukreja, “N-CNN-based transfer learning method for classification of powdery mildew wheat disease”, In 2021 International Conference on Emerging Smart Computing and Informatics (ESCI) (pp. 707-710), 2021, IEEE.

[3] Available at: https://neurohive.io/ and retrieved on 13.03.2023.

[4] Available at: https://iq.opengenus.org/ and retrieved on 13.03.2023.

[5] Available at: https://ai.googleblog.com/2019/05/efficientnet-improving-accuracy-and.html and retrieved on 13.03.2023.

[6] S. J. Wang, F. Jiang, B. Zhang, R. Ma, and Q. Hao, “Development of UAV-based target tracking and recognition systems,” IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 8, pp. 3409–3422, 2020.

[7] O. Kechagias-Stamatis and N. Aouf, “Evaluating 3D local descriptors for future LIDAR missiles with automatic target recognition capabilities,” e Imaging Science Journal, vol. 65, no. 7, pp. 428–437, 2017.

[8] W. L. Xue and T. Jiang, “An adaptive algorithm for target recognition using Gaussian mixture models,” Measurement, vol. 124, pp. 233–240, 2018.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", *Proc. IEEE Conf. Comput. Vis. Pattern Recognit*, pp. 770-778, June, 2016.

[10] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks[J]", *Advances in neural information processing systems*, vol. 25, pp. 1097-1105, 2012.

[11] Y.-n. Dong and G. -s. Liang, "Research and Discussion on Image Recognition and Classification Algorithm Based on Deep Learning", *2019 International Conference on Machine Learning Big Data and Business Intelligence (MLBDBI)*, pp. 274-278, 2019.

[12] D. Xing, W. Dai, G.-R. Xue, Y. Yu, "Bridged Refinement for Transfer Learning", Proc. 11th European Conf. Principles and Practice of Knowledge Discovery in Databases, pp. 324-335, 2007-Sept.

[13] Jason Yosinski, Jeff Clune, Yoshua Bengio, “Advances in Neural Information Processing Systems” 27, pages 3320-3328. Dec. 2014; [http://arxiv.org/abs/1411.1792 arXiv:1411.1792].

[14] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions”, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1–9, 2015.

[15] Mingxing Tan, Quoc V. Le, “ EfficientNet : Rethinking Model Scaling for Convolutional Neural Networks”, pages 1-11, 2019.

[16] Available at:https://www.kaggle.com/ and retrieved on 13.03.2023.