International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Performance Comparison between VGG16 and Inception V3 for Organic Waste and Recyclable Waste Classification

Abba Suganda Girsang[1], Andrew Dharma Saputra[2], Vincentius Yanrie[3]

Submitted: 08/11/2022 **Accepted**: 12/02/2023

Abstract: Computer vision is used for learning image recognition, where a CNN algorithm is used to carry out the learning of the image itself. In this paper, a comparison is made between the two algorithms to determine which is better by comparing VGG16 and Inception V3 using a dataset that distinguishes the types of organic waste and recyclable waste. This study proposes algorithms using VGG-16 and Inception V3 that use a semi-supervised learning system to train algorithms from different images. By training the two algorithms, it can be seen that VGG16 and Inception V3 have quite good accuracy, but for the dataset used, it is better to use the VGG16 algorithm because Inception V3 has a fairly complicated algorithm model that makes the performance of the algorithm with the dataset not optimal. Therefore, further research is needed to optimize the two models to train the dataset. This case is taken because waste management is still a problem for our environment.

Keywords: *Computer Vision, VGG16, Inception V3, waste classification*

1. Introduction

Human population in the world is increasing each year, as well as the growth of technology that always continues to develop. The more human population in an area the more waste is produced. World bank reports show that nearly 4 billion tons of waste worldwide each year. People like to bury waste in the land, thrown in the street, thrown in the sea. Indonesia itself produces more than 64 million tons of waste[1]. Indonesia in 2025 based on the news report will be the second largest contributor of plastic waste in the world. According to data from the Ministry of Environment and Forestry[2], the total waste produced in Indonesia is 175,000 tonnes per day. Waste is a problem that has a direct or indirect impact on humans and the environment. People need to manage waste so the world can be saved.

In several studies, waste has been classified into two classes: organic and inorganic, such as research conducted by organizing waste, especially organic sections that can be recycled or not[3]. organizing waste using algorithms in Computer Vision such as VGG16 and Inception V3 that will compare which model has a good accuracy to classified waste between organic and recyclable waste. Waste classifiers are being developed in many countries that researchers have been attempting to make and develop waste classifiers. Wang used CNN with the VGG-16 model on a waste dataset consisting of 47,000 images and an accuracy of 86.1%[4]. The accuracy of the model in classifying organic and inorganic waste still can be improved with hyperparameter optimization. So this paper will compare the model VGG-16 and Inception V3 in classifying organic and recyclable waste that is a little bit different from Wang research that uses organic and inorganic dataset.

In recent years, many technologies have finally used computer vision techniques to deepen several sectors in the field of informatics engineering so that they can develop this technology better in the future. There are many ways to improve deep learning performance, one of which is using hyperparameter optimization, which sets parameters in the CNN algorithm that can improve performance and accuracy [5]. This optimization technique uses a metaheuristic technique which is carried out to find a fixed layer architecture which is not suitable for making CNN models because there are layers that are not known when making the model[6].

The comparison of the two CNN models, namely VGG16 and Inception V3, demonstrates that both models study the same dataset and produce the same results, implying that more research will be conducted to develop these two models to achieve better results in terms of accuracy and efficiency of use model.

[1] *Computer Science Department, BINUS Graduate Program –Master of Computer Science, Bina Nusantara University, Jakarta 11380, Indonesia, ORCID ID: 0000-0002-0529-2095*
[2] *Computer Science Department, BINUS Graduate Program –Master of Computer Science, Bina Nusantara University, Jakarta 11380, Indonesia,*
[3] *Computer Science Department, BINUS Graduate Program –Master of Computer Science, Bina Nusantara University, Jakarta 11380, Indonesia*
* *Corresponding Author Email: vincentius.yanrie@binus.ac.id*

2. Basic Theory

2.1. Computer Vision

Computer Vision is a computation that is combined with a camera, artificial intelligence that can see or identify objects. Computer Vision uses deep learning to form a neural network which guides the system in management and analysis. Computer Vision is also a computer that can see objects around it and the computer will analyze the image itself which we will direct to the computer to perform certain commands. Computers in technique use specific software algorithms to obtain visual information, process it, and analyze the data [7].

The function of computer vision is where the information obtained can be clearer by knowing the object to be used where we as humans have difficulty seeing these objects so with computer vision it can make it easier to see these objects clearly. Then there is an analysis of computer vision which can capture the image and process it further to carry out what commands we as developers use [8].

2.2. Deep Learning

Deep Learning is a subfield of machine learning concerned with algorithms inspired by the structure and function of the brain called artificial neural networks. Deep Learning differs from traditional machine learning techniques in that deep learning automatically performs representations of data such as images, video, or text without introducing code rules or human domain knowledge [9].

Deep Learning refers to a group of deep learning algorithms that are influenced by the structure of the human brain. Deep Learning algorithms make use of multi-layered neural networks with non linear input data transformations that steadily raise the level of abstraction [10].

Deep learning can achieve greater identification precision than ever before. This enables consumers to exceed customer needs, which is important for safety-critical technologies such as self-driving vehicles. Deep Learning has progressed to the point that it now surpasses humans in certain functions, such as classifying objects in pictures [11].

2.3. Transfer Learning

Deeper neural networks require a lot of processing power to train, and creating them takes a long time. Transfer learning, a well-liked deep learning technique, uses pre-trained models as a starting point in computer vision and natural language processing, which can minimize computational resources and shorten the time needed to create an ideal model. In transfer learning, the basic network is trained first on the data set and the basic task, and the learned features are then reused or transferred to the second target network to train on the target data set and task. This technique tends to operate if the traits are universal, that is, applicable to fundamental and objective tasks rather than being particular to those activities [12].

2.4. Convolutional Neural Network

Image classification is one of the computer vision features that studies how a computer can see and understand what it sees visually. CNN is the most popular deep learning algorithm, and it has been widely used in image classification due to its high ability to classify images and patterns correctly or precisely. Image classification algorithmic methods include K-means, SVM (Support Vector Machine), and CNN (Convolutional Neural Network) [13].

CNN uses the convolutional process to its advantage by applying a convolutional kernel of a specific size to an image [14]. The convolutional results are then used as input to generate a feature representation. CNN can recognize an object thanks to convolution.

This process will be repeated for all parts of the image. Furthermore, the max-pooling process reduces the number of parameters while retraining the essential information from each component and lowering the image's resolution [15]. Meanwhile, the fully connected layer is the vector that results from several convolutions and pooling operations on the multilayer perceptron [16].

3. Proposed Method

In this research we used a public dataset that consists of 25077 images of organic waste and recyclable waste. The dataset itself has been licensed by CC BY-SA 4.0. Before the data is processed, the data is divided into three parts, namely training, validation, and testing. We divided it to 63% training, 27% validation, and 10% testing.

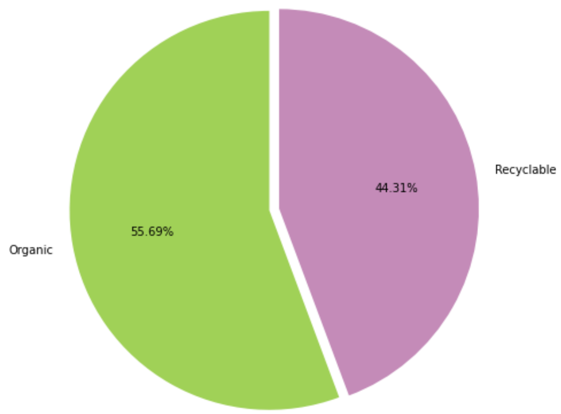

Fig. 1. Dataset Class

Figure 1 shows us that the dataset consists of 55.69% of organic waste image and 44.31% recyclable waste image. The data will be pre-processed so it can be used optimally.

With the help of OpenCV we can read the data but it will be read as BGR color. We can convert it back to RGB by using the OpenCV function. The data scale is also changed into 224 x 224 pixels so it can be used optimally.

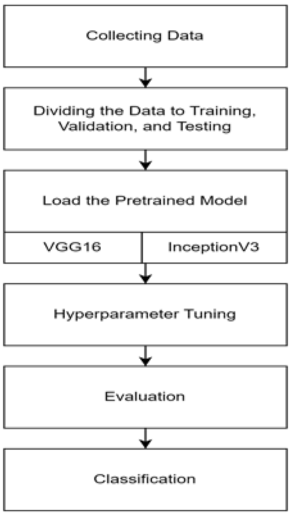

Fig. 2. Flow of the Method

In this research we will use two different deep learning models namely VGG16 and InceptionV3. Figure 2 shows the step of the method that we use. In the Hyperparameter Tuning step we tried to modify the learning rate, drop out rate, activation function, optimization algorithm, batch size, and epochs to find the optimal hyperparameter value. The evaluation step will show us the accuracy and loss in the training process and testing process.

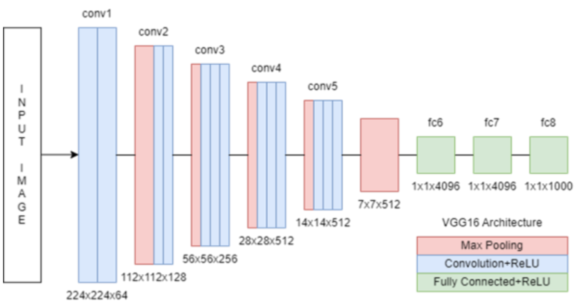

Fig. 3. VGG16 Architecture [17]

The deep learning model VGG16 is made of 13 convolution layers, 3 fully connected layers, and 5 max pooling layers as shown on Figure 3 [19]. The name itself came from the sum of 13 convolution layers and 3 fully connected layers, which is 16. The convolution kernels in each convolution layer in the VGG16 will increase from 64 to 512 [20]. In this research half of the VGG16 parameters will be freezed so half of the parameters can be trained but still maintain the other half trained information.

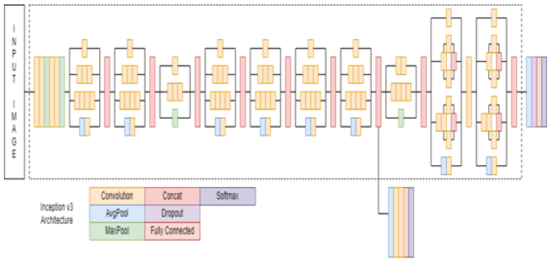

Fig. 4. Inception v3 Architecture [18]

The Inception V3 model in general is made of 6 convolution layers, 2 max pool layers, 3 inception with module A, 5 inception with module B, 2 inception with module C, a linear layer, and a softmax layer [18]. Below are the Figures for the modules architecture. In the earlier Inception version they used to have a larger spatial layer such as (5x5). This convolution will cost an expensive amount of computation. To reduce it in Inception V3 they replace the (5x5) convolution layer into two (3x3) convolution layers. This will reduce the computational cost because the parameters are decreased. For example, the (5x5) convolution layer will give 25 parameters and a single (3x3) convolution layer will give 9 parameters. The (5x5) convolution layer costs 2,778 times more than the (3x3) convolution layer [18].

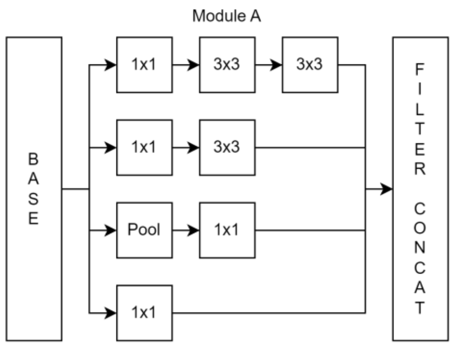

Fig. 5. Inception Module A

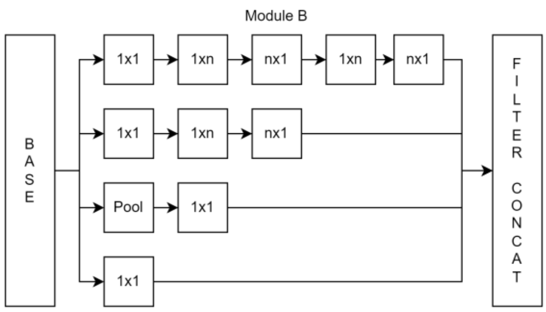

Fig. 6. Inception Module B

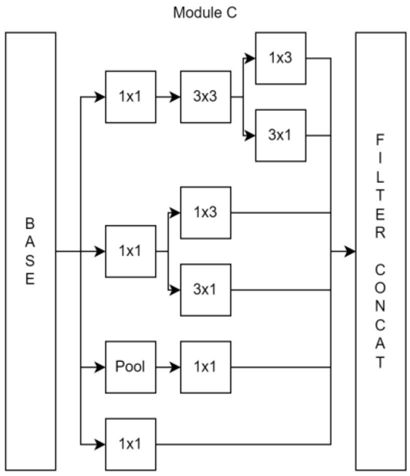

Fig. 7. Inception Module C

In this research half of the Inception V3 layers will be freezed so half of the parameters can be trained while maintaining the other half trained information.

At the start of both models an input layer will be added with input size of (224, 224, 3). At the end of both models a global average pooling 2D will be used to get the average pool, a dropout layer will be added by 0.5 value to reduce half of the neurons during training and to reduce overfitting, and lastly a dense layer for classify purposes.

4. Comparative Analysis

Efficiency of the algorithm that has been tested where there are two algorithms namely using VGG16 and Inception V3 where each algorithm has a different number of total parameters transferred. The total parameters of the first algorithm which is InceptionV3 are 21.806.884 with separators as trainable parameters of 10.545.090 and non-trainable parameters of 11.261.792. Meanwhile, the second algorithm which is VGG16 has a total of 14.715.714 parameters with 7.080.450 trainable parameters and 7.635.264 non-trainable parameters. From the differences in the parameters used for transfer learning where the number of parameters affects how well the algorithm works to produce high accuracy so there will be experiments to improve the accuracy of the two algorithms by comparing the loss and accuracy of the training and testing in the two algorithms.

Table 1. VGG16 Test Result

VGG16 Test Result		
Epoch	Accuracy	Loss
1	89,53%	0,29
2	90,37%	0,29
3	90,49%	0,29
4	91,33%	0,25
5	91,05%	0,27
6	**91,60%**	**0,24**
7	91,05%	0,29
8	88,30%	0,41
9	91,80%	0,27
10	92,32%	0,25
11	91,21%	0,34
12	91,13%	0,34
13	91,01%	0,37
14	90,93%	0,36
15	92,16%	0,40
16	91,68%	0,38
17	90,53%	0,43
18	90,17%	0,56
19	90,93%	0,55
20	91,88%	0,54
21	90,77%	0,65
22	90,05%	0,68
23	91,80%	0,57
24	91,56%	0,67
25	92,12%	0,46

Table 2. Inception V3 Test Result

Inception V3 Test Result		
Epoch	Accuracy	Loss
1	90,01%	0,27
2	89,65%	0,27
3	**91,25%**	**0,25**
4	90,97%	0,26
5	90,81%	0,29
6	91,44%	0,27
7	91,09%	0,33
8	90,77%	0,33

9	91,05%	0,34
10	91,72%	0,33
11	91,05%	0,37
12	91,21%	0,38
13	91,36%	0,38
14	90,85%	0,41
15	91,05%	0,42
16	91,56%	0,40
17	91,25%	0,46
18	89,45%	0,53
19	91,64%	0,41
20	91,21%	0,45
21	91,72%	0,44
22	92,08%	0,40
23	91,17%	0,48
24	91,68%	0,43
25	91,01%	0,48

Overall, a comparison of the two deep learning models can be seen from each epoch that we run with the same number of epochs, where for the VGG16 model in Table 1, the 6th epoch gets a fairly high accuracy of 91.60% with a fairly small loss of 0.24, but when compared with the Inception V3 model in Table 2, when it was in epoch 3, it got similar accuracy with a relatively small loss ratio of 91.25% and 0.25 loss.

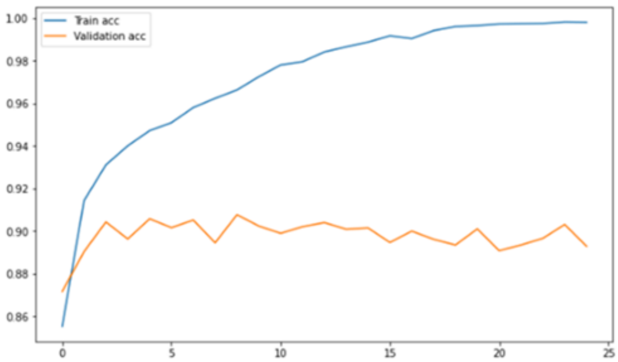

Fig. 8. VGG16 Train Accuracy Graph

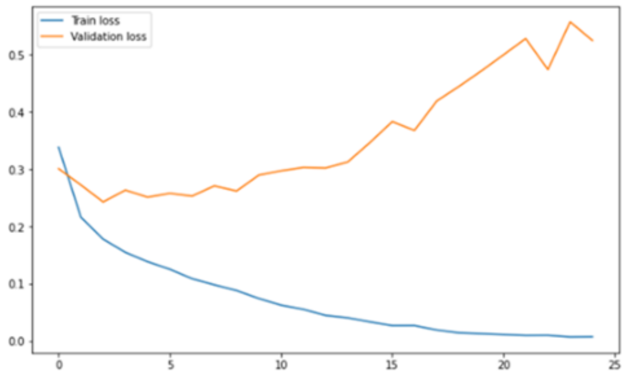

Fig. 9. VGG16 Train Loss Graph

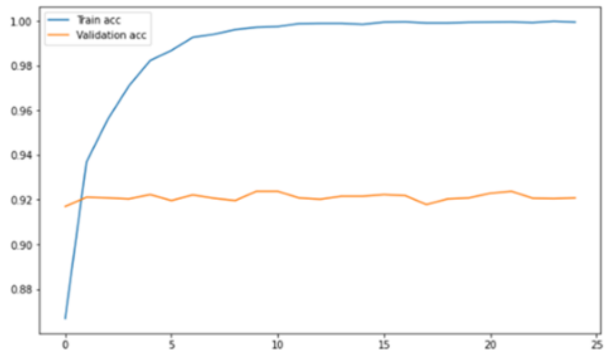

Fig. 10. Inception V3 Train Accuracy Graph

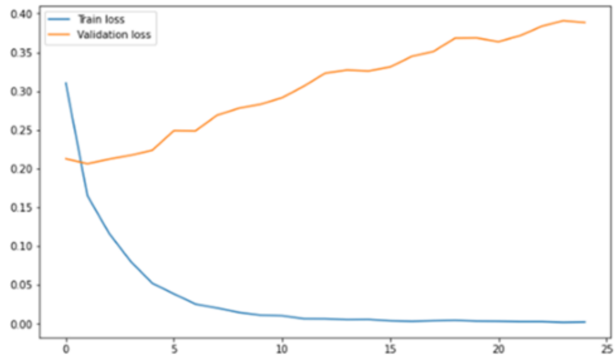

Fig. 11. Inception V3 Train Loss Graph

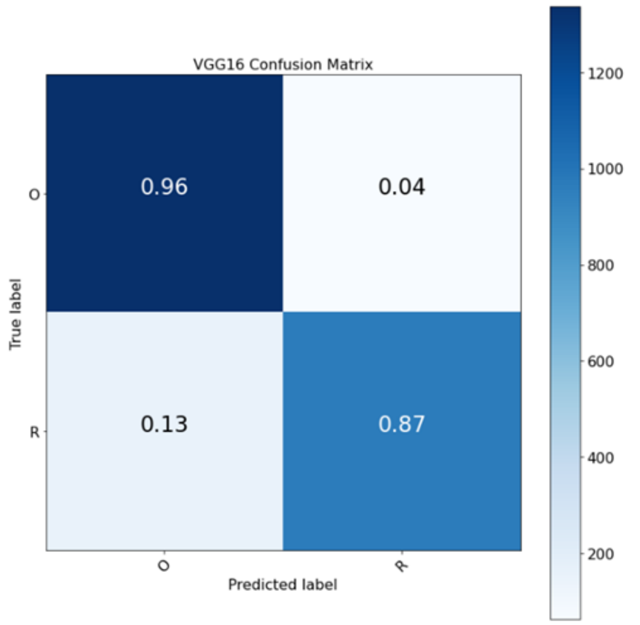

Fig. 12. Heatmap VGG16 Confusion Matrix

The confusion matrix for the VGG16 model can be seen in Figure 12. In the prediction of organic waste, there is a 0.96 correct prediction and 0.04 wrong prediction. Meanwhile, in the prediction of recyclable waste, there is a 0.87 correct prediction and 0.13 wrong prediction.

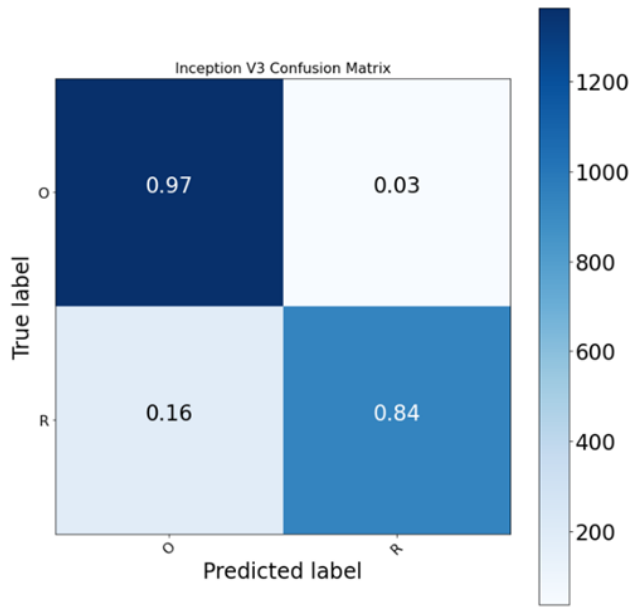

Fig. 13. Heatmap Inception V3 Confusion Matrix

The confusion matrix for the Inception V3 model can be seen in Figure 13. In the prediction of organic waste, there is a 0.97 correct prediction and 0.03 wrong prediction. Meanwhile, in the prediction of recyclable waste, there is a 0.84 correct prediction and 0.16 wrong prediction.

5. Conclusion

In this research, it appears that the classification of waste can be done using image classification with the help of deep learning algorithms. By optimizing the hyperparameter of the transferred learned models, we can achieve a fairly high accuracy, which is 91.60% for the VGG16 model and 91.25% for Inception V3 model.

From the comparison of the accuracy of the two, it can be seen that Inception V3 has a bit lower accuracy than VGG16, it seems that the dataset that we used for this problem is more compatible for VGG16. Both of the models can still be tuned to achieve better accuracy.

Author contributions

Abba Suganda Girsang: Conceptualization, Methodology, Field study, Writing-Original draft preparation

Andrew Dharma Saputra: Data curation, Software, Validation., Field study, Writing-Reviewing and Editing

Vincentius Yanrie: Writing-Original draft preparation, Visualization, Investigation, Writing-Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] H. N. Jong, “Indonesia in state of waste energy”, *The Jakarta Post* October 9, 2015.[Online], Available:https://www.thejakartapost.com/news/2015/10/09/indonesia-state-waste-emergency.html.

[2] C. S. Burke, E. Salas, K. Smith-Jentsch, and M. A. Rosen, “A Global Review of Solid Waste Management,” *A Glob. Rev. Solid Waste Manag.*, pp. 29–43, 2012, doi: 10.1201/9781315593173-4.

[3] O. Adedeji and Z. Wang, “Intelligent waste classification system using deep learning convolutional neural network,” *Procedia Manuf.*, vol. 35, pp. 607–612, 2019, doi: 10.1016/j.promfg.2019.05.086.

[4] H. Wang, Y. Li, L. M. Dang, J. Ko, D. Han, and H. Moon, “Smartphone-based bulky waste classification using convolutional neural networks,” *Multimed. Tools Appl.*, vol. 79, no. 39–40, pp. 29411–29431, 2020, doi: 10.1007/s11042-020-09571-5.

[5] Y. Wang, H. Zhang, and G. Zhang, “cPSO-CNN: An efficient PSO-based algorithm for fine-tuning hyper-parameters of convolutional neural networks,” *Swarm Evol. Comput.*, vol. 49, pp. 114–123, 2019, doi: 10.1016/j.swevo.2019.06.002.

[6] I. Y. Kim and O. L. De Weck, “Variable chromosome length genetic algorithm for progressive refinement in

topology optimization," *Struct. Multidiscip. Optim.*, vol. 29, no. 6, pp. 445–456, 2005, doi: 10.1007/s00158-004-0498-5.

[7] Szeliski, R. (2022). *Computer vision: algorithms and applications*. Springer Nature.

[8] Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience*, *2018*.

[9] Mathew, A., Amudha, P., & Sivakumari, S. (2021). Deep learning techniques: an overview. In *International conference on advanced machine learning technologies and applications* (pp. 599-608). Springer, Singapore.

[10] Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing*, *187*, 27-48.

[11] Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, *31*(3), 685-695.

[12] H. L. Jason Yosinski, Jeff Clune, Yoshua Bengio, "How transferable are features in deep neural networks?," *Proc. Int. Jt. Conf. Neural Networks*, vol. 2016-Octob, pp. 2560–2567, 2016, doi: 10.1109/IJCNN.2016.7727519. Available: http://dl.z-thz.com/eBook/zomega_ebook_pdf_1206_sr.pdf. Accessed on: May 19, 2014.

[13] Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*.

[14] Kattenborn, T., Leitloff, J., Schiefer, F., & Hinz, S. (2021). Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS journal of photogrammetry and remote sensing*, *173*, 24-49.

[15] Bhatt, D., Patel, C., Talsania, H., Patel, J., Vaghela, R., Pandya, S., ... & Ghayvat, H. (2021). CNN variants for computer vision: history, architecture, application, challenges and future scope. *Electronics*, *10*(20), 2470.

[16] Wang, S. Y., Wang, O., Zhang, R., Owens, A., & Efros, A. A. (2020). CNN-generated images are surprisingly easy to spot... for now. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 8695-8704).

[17] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[18] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).

[19] Qassim, H., Verma, A., & Feinzimer, D. (2018, January). Compressed residual-VGG16 CNN model for big data places image recognition. In 2018 IEEE 8th annual computing and communication workshop and conference (CCWC) (pp. 169-175). IEEE.

[20] Yang, H., Ni, J., Gao, J., Han, Z., & Luan, T. (2021). A novel method for peanut variety identification and classification by Improved VGG16. Scientific Reports, 11(1), 1-17..