 https://doi.org/10.24191/jeesr.v20i1.002

Comparison Study on Convolution Neural Network (CNN) Techniques for Image Classification

Siti Maisarah Zainorzuli, Syahrul Afzal Che Abdullah, Husna Zainol Abidin and Fazlina Ahmat Ruslan*

***Abstract*— Deep Learning is an Artificial Intelligence (AI) function which can imitate the human brain to process data and deciding. It has networks that able to learn the unsupervised data that unlabeled or unstructured. It also identified as Deep Neural Network or Deep Neural Learning. Convolutional Neural Network (CNN) is a subset of Deep Neural Network which frequently used to analyse images. CNN also called as ConvNet which can be trained using an existing model that has been fine-tuned or trained from zero by using a large data set. CNN was often used in image classification due to its effectiveness and accuracy. However, there are several CNN architectures such as AlexNet, GoogleNet and ResNet-50. To select the appropriate architecture for our research in agriculture, a preliminary study to evaluate the architecture were conducted by using five different types of flower datasets that obtained from Matlab and Kaggle database. The three types of CNN architecture were compared in terms of accuracy in classifying the flowers. Result of this study indicated that the optimal configuration is by setting the number of epochs at 30, with the learning rate at 0.0005, to obtain the highest accuracy at 99.82%.**

***Index Terms*— Convolutional Neural Network (CNN), Image classification, AlexNet, GoogleNet, ResNet-50.**

I. INTRODUCTION

COMPUTER Science and Artificial Intelligence (AI) has made an astonishing development for the past years. The AI research has expanded significantly every year regarding the application of Machine Learning (ML) which resulted in the development of a new model called Deep Learning [1].

This can be seen that the AI system such as Deep Learning [2] are capable to provide services that deem to be intelligent and creative. One of the subgroups of Artificial Intelligence is deep learning which lately frequently been used to carry out tasks such as image categorization, voice identification and many more [3].

For the past few years, deep learning has gained a lot of attention from the researchers due to the increasing request for learnable machines that can solve various complex problem. Deep learning has a technique that able to solve the feature selection issue by do not require the pre-selected features but extracting the important features from unprocessed input automatically [4]. There are a group of processing layers that can be found in deep learning which able to learn many features of data through various levels of abstractions [5]. The network can learn the recognizable features due to the multiple levels.

It appears that the deep learning has accomplished a promising result in several applications such as image classification [6], speech to text [7], semantic segmentation [8], object detection [9], Natural Language Processing (NLP) [10], vehicle recognition [11] and many more.

The Deep Learning architecture that usually used for image processing is Convolutional Neural Network (CNN) [12]. CNN is one of the deep learning models that inspired by the structure of the animal visual cortex which organizes data with a grid pattern [13]. CNN also known as ConvNet [14] is built with deep feed-forward and better ability compared to other networks with fully connected layers for generalizing.

Next, in between various of deep learning approaches, CNN has shown excellent accomplishment in image recognition tasks [15]. CNN has made a series of development in image recognition and classification area for the past years [16]. CNN is the most establish algorithm amidst several deep learning models, which has been a dominant technique in computer vision tasks since the surprising outcome on the object recognition in an ImageNet Large Scale Visual Recognition Competition (ILSVRC) in 2012 [17]. The CNN has been using in various tasks that obtain an impressive performance in several applications.

Besides that, due to the CNN special process, it can operate

This manuscript is submitted on 23 February 2021 and accepted on 23 November 2021. Siti Maisarah Zainorzuli, Syahrul Afzal Che Abdullah, Husna Zainol Abidin, and Fazlina Ahmat Ruslan are with the School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor (e-mail: 2019683242@student.uitm.edu.my).

Syahrul Afzal Che Abddullah, is now with the Centre for Computer Engineering Studies (CCES), School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA (e-mail: bekabox181343@uitm.edu.my).

Husna Zainol Abidin, is now with the Centre for Computer Engineering Studies (CCES), School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA (e-mail: husnaza@uitm.edu.my).

Fazlina Ahmat Ruslan, is now with the Centre for System Engineering Studies (CSES), School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA (e-mail: fazlina419@uitm.edu.my).

*Corresponding author
Email address: fazlina419@uitm.edu.my

well than the other deep neural network. Rather than focus on the picture pixel one at a time, CNN gathers some of the pixels together to recognize the temporary pattern [18]. The common types of CNN that been use for object detection and classification from pictures are AlexNet, GoogleNet and ResNet-50 [19].

A study by Rangajaran [20] which used AlexNet and VGG 16 ro classify tomato diseases concluded that within the minimum execution time, AlexNet were able to obtain higher accuracy compared to the VGG 16. One of the AlexNet disadvantages is it has less depth compared to the newest model thus it has difficulty grasping the image features. The advantage of ResNets is its fast convergence [21]. ResNet has a great approach in which once it learn the features it will not relearn again but will focus on the new feature instead.

This paper compare the performance of different types of CNN architectures to be used on agriculture images. The remainder of this paper is organized as follows: Section 2 is a summary that review several background studies that use image classification technique by the past researchers. Next, Section 3 will explain methods that was used in this research to obtain the expected result. Then, Section 4 will clarify the obtained result by three types of CNN models used in this research which is AlexNet, GoogleNet and ResNet-50. Lastly, Section 5 will conclude the result obtained by the three types of CNN models that has been used in this research.

II. Literature Review

This part will cover some of the CNN models which is AlexNet, GoogleNet and ResNet-50 that has been used in this study. Besides that, this part also a summary of a few study by the past researcher that using image classification technique in several backgrounds such as agriculture and medical.

A. Image Processing

Image processing used to extract photographic data from the image after processing the digital images. It also used to operate the contrast and brightness of the pictures since the digital camera was not capable to do so. Pictures that were captured by the digital camera also contain dissimilar types of noises which can be reduced by using this technique [22].

B. Convolutional Neural Network (CNN) Models

1) AlexNet

Alex Khrizevsky won an ImageNet Large Scale Visual Recognition Competition (ILSVRC) in 2012 which lead to the development of the CNN structure that was named after him, AlexNet [23]. Countless of researches were focus to improvise the architecture of CNN since AlexNet accomplish astonishing result in ILSVRC-2012 image classification competition. The development of computer hardware and the improvement of data sets are the main contributor to AlexNet achievement [24].

AlexNet is a large neural network that has 60 million parameters and 650,00 neurons [25]. It has 5 convolutional layers and 3 fully connected layers [26]. AlexNet is designed with piles of multiple layers.

2) GoogleNet

A new Deep Learning structure was introduced in 2014 known as GoogleNet [27] by Christian Szegedy to reduce the computation difficulty compared to the traditional CNN [28]. GoogleNet has another name which is Inception [29]. GoogleNet contains the Inception Modules which changed the sizes of convolution and a series of filter for the following layer [30]. Inception modules also helps to improvise localization and object recognition [31]. Besides that, it also contains a convolution layer including 128 filters to decrease the parameters, a fully connected layer, in lines layer with softmax loss as the classifier and a dropout layer [32].

3) ResNet-50

K. He was the one that introduces the Residual Network (ResNet) [21]. The author claimed that by stacking the CNN layers on the top of each other will not improve the performance due to the vanishing gradient problem. To overcome this problem, the author suggested the "residual block".

ResNet layers were reformatted so that it can learn the residual mapping which also known as shortcut connection [33]. This network is mainly designed to analysis a large-scale data and established with a lot of various numbers of layers [34].

C. In Agriculture

A Research by Hyun K. Suh [35] used some of CNN model in agriculture which is to classify the sugar beet and volunteer potato. The focus of the EU (European Union) Smartbot project was to manage above 95% of volunteer potatoes and below 5% of sugar beet plants by using Deep Learning and also very time-consuming thus, transfer learning was the key to this situation. The transfer learning procedure was analyzed first by using three different execution of AlexNet and compute the performance distinction between six network architectures which is AlexNet, Visual Geometry Group 19 (VGG-19), Residual Network (ResNet) 50, ResNet-101 and Inception-v3. All of the nets were pre-trained by using the ImageNet dataset. There were three scenarios, the theory was tried out whether with or without retraining AlaexNet will the classification gain the precision of 95% or further could be attained using the property from Fully-Connected layer (FC) 6 and FC7 and also by using the conventional classifiers. The second scenario, the AlexNet was undergoing some changes to generate binary classification outcome then it was fine-tuned with training pictures of volunteer potato and sugar beet. The modified VGG-19 achieve the highest categorization with 98.7% precision by compared between networks.

Next, this research by Komal Bashir [36] used the image classification to analyze and identify three types of rice crop diseases which have two-stage which is the training stage and disease forecast stage. The disease on the leaf was classified by using trained classifier in this research. To obtain the highest efficiency, the researcher enhances the Support Vector Machine (SVM) parameters which are gamma and nu. Training stage used for the analysis of the training dataset and training SVM.

The disease detection stage detects and forecasts the disease from images that were not used in the training stage. Detection and classification of the diseases using Scale-Invariant Feature Transform (SIFT). The result obtains shows that this approach has obtained 94.16% accuracy which accompanied by 5.83% misclassification rate, 91.6% recall rate and 90.9% precision.

D. *In Medical Field*

Wajahat Nawaz [37] used AlexNet to classify breast cancer histology pictures. In this study, the researcher fine-tunes the AlexNet model by replacing and insert the input layer which is convolutional layers to fully connected (FC) layer. The training set X1 was extended by created mirrored and rotated images in X1. The overlapping patches size 512x512 were extracted from the training images. Rather than using the default input layer size of AlexNet which is 227x227, the researcher resizes to 512x512. Besides that, the FC layer was replaced by the convolutional layer with 256 filter size of 3x3. The activation (ReLU) and max-pooling layer have the same size. The size of FC layer was also been reduced (FC-1:256) and the ReLU (without the dropout layer) also has been resized. After that, to avoid overfitting, the FC layer will be resized (FC2:100). At the last layer, the researcher added another 4 FC layer of various classes. This study obtains 75.73% accuracy for image-wise, 81.25% for the validation set and 57% for the test set.

Next, Lawrence V. Fulton [38] research was about the classification of Alzheimer's disease (AD) using gradient boosted machines (GDM) and ResNet-50 by. The purpose of this study was to predict the existence of AD by using clinical, socio-demographic and magnetic resonance imaging (MRI) data. The file name and associated labels were stored in a data frame due to large size of datasets. The data frame was divided randomly to two different data frames which is 80% for training set and 20% for validation set. The data were rescaled by training image generator so that the value will be between 0 and 1 (divided by 255 grayscale). The images were randomly rotated up to 20 degrees, zoomed up by 10%, the width and height were shifted by 10%, sheared up by 10% and some images were flipped horizontally. By using this method to edit the training images, the machine learning (ML) model were able to learn how to adapt to the deviations by the expected format. By using socio-demographic and mini-mental state exam (MMSE), the prediction accuracy result obtains by the GBM was 91.3% for dichotomous clinical dementia rating (CDR). While for ResNet-50, by using the image generation technique the accuracy obtains 98.99%.

Based on the literature work done that were tabulated in Table I, it can been seen that recently these CNN architectures are widely used in the area of medical and plantation. However, as far as the author knowledge there is no work done in classifying flower images where the structure of the images were different from these work done.

TABLE I
SUMMARY OF RELATED WORK

Ref.	Architecture	Dataset	Year
[35]	AlexNet, VGG-19, GoogleNet, ResNet-50, ResNet-101 snd Inception-v3	Sugar beet and volunteer potato	2018
[36]	SVM (Support Vector Machine)	Broen Spot, Bacterial Leaf Blightand and False Smut	2019
[37]	AlexNet	Breast cancer	2018
[38]	ResNet-50	Alzheimer's Disease	2019

III. METHODOLOGY

In this part, the whole process of this research starting from the beginning to the end will be explain. The process will start with the data collection and end with the CNN model's development.

A. *Data Collection*

This research used datasets from MATLAB and Kaggle to develop the CNN models. The datasets that were used from MATLB and Kaggle is flowers dataset. The datasets have five types of flower which is daisy, dandelion, rose, sunflower and tulip. Based on the past researcher [39], it will be difficult to train the CNN model by using a dataset that only has 1,440 examples. Therefore, the total of images that were used to develop the model in this research is 3,898.

The datasets were separated into two sets which is training and validation. The training set used 70% of the pictures while the remaining 30% used for validation. The first layer of each CNN models explains the input size.

Every types of CNN model can only develop RGB (Red, Green and Blue) pictures according to the CNN input picture requirement. By using the augmentedImageDatastore the input pictures will be resized and change grayscale picture to RGB. Besides that, overfitting and learning the precise features of the training pictures can be avoided by using the data augmentation.

A few layers in CNN that are appropriate for picture characteristic abstraction. The beginning layers of the network will capture simple picture characteristic which is blobs and edges. By using transfer learning, the number of epochs to be trained does not have to be numerous.

B. *Model Development*

The pictures of 5 types of the flower were load in image datastore. By using the function of imageDatastore the images were automatically label based on the folder name where the images were stored. The data in image datastore were divided into two categories which is training and validation.

This study used 2 methods of how the images were divided for training and validation. The first method was the first 70%

of the images were selected as training data sets, while the remaining 30% were used as the validation data sets. The second method was using the randomized function which will select 70% of the images randomly for the training data sets and the remaining were used as the validation data sets.

Next, the pre-trained CNN model which is AlexNet, GoogleNet and ResNet-50 were loaded. The images that were store in image datastore have different sizes thus, by using the augmented image datastore function the training images were automatically resized.

The three models of CNN used the Stochastic Gradient Descent with Momentum (sgdm) as enhancer with two sets of parameters. The first set is used fixed epoch at 30 and various learning rate such as 0.00005, 0.0005, 0.0001 and 0.001. The second set is by using a fixed learning rate which is at 0.0001 and the different sets of epochs at 10, 20 and 30. The parameter used for both of the sets was based on past researcher which is [34], [35] and [40].

The learning process occurs in the transfer layers. After the parameter was set for the CNN model, the training and validation data were loaded into the network. After the network was done trained, the validate images were classified and the network accuracy were calculated.

IV. Result and Discussion

This part will discuss the result that were obtain by AlexNet, GoogleNet and ResNet-50. In this study, the research is carried out on five various flower datasets which obtain from MATLAB and Kaggle database. Every type of flower was split into two groups, training and validation datasets. There were two methods datasets selection. The first method where the first 70% of the first flower for every category were selected as the training data, while the remaining images used to validate. The second method, 70% of all type of flower used as training data while the remaining were used as validation data.

A. *Images selection by using the first 70% of the images*

TABLE II

VALUE OF ACCURACY (%) OBTAIN USING THE FIRST SET OF PARAMETERS

CNN Model	Learning Rate			
	0.001	0.0001	0.0005	0.00005
AlexNet	99.50	98.40	99.54	97.36
GoogleNet	99.74	96.84	99.82	94.64
ResNet-50	99.42	98.80	99.58	97.30

Based on the previous section, it already explains and justifies the value of parameters that have been used in this research. The learning rate determines the learning development of the CNN models while the number of epochs is the full teaching pattern over the whole training datasets.

It can be seen from the data in Table II that the outcome by using the first set of parameters shows that GoogleNet obtains the highest accuracy of 99.74% and 99.82% when the value of learning rate is 0.0005 and 0.001. Now turn to ResNet-50 model that acquire the most accurate classification with 98.80% at 0.0001 learning rate, while at learning rate 0.00005 AlexNet gain the highest accuracy of 97.36%.

TABLE III

VALUE OF ACCURACY (%) OBTAIN USING THE SECOND SET OF PARAMETERS

CNN Model	No. of Epoch		
	10	20	30
AlexNeet	93.96	98.02	98.40
GoogleNet	94.66	98.16	96.84
ResNet-50	94.46	98.60	98.80

From Table III, it Illustrates the breakdown of accuracy obtained by the three CNN models by using the second set of parameters. The parameter used is the learning rate is fixed at 0.0001 while the number of epochs set to 10, 20 and 30. Based on Table 2, the GoogleNet model achieves the highest accuracy which is 94.66% at epoch number is 10. Next, The ResNet-50 attain the highest accuracy at the number of epochs at 20 and 30 with the accuracy of 98.60% and 98.80%.

B. *Images selection by using randomized method*

This method using the same parameter value as the previous image selection method. All of the models using the sgdm as enhancer with two sets of parameters. The parameter used in this method also based on past researchers which is same as the previous image selection technique.

TABLE IV

VALUE OF ACCURACY (%) OBTAIN USING THE FIRST SET OF PARAMETERS

CNN Model	Learning Rate			
	0.001	0.0001	0.0005	0.00005
AlexNet	99.74	97.44	99.78	97.26
GoogleNet	99.62	97.56	99.70	97.66
ResNet-50	99.52	98.54	99.82	97.16

The first set of parameters used is the number of epochs is fixed at 30 while the learning rate is set to several value starting from 0.001 and end at 0.00005. As shown in Table IV, AlexNet attains the highest accuracy with 99.74% at 0.001 learning rate. Next, the highest accuracy gains by the ResNet-50 is 98.54% and 99.2% at 0.0001 and 0.0005 learning rate. Lastly, GoogleNet achieves the highest accuracy with 97.66% at 0.00005 learning rate.

TABLE V

VALUE OF ACCURACY (%) OBTAIN USING THE SECOND SET OF PARAMETERS

CNN Model	No. of Epoch		
	10	20	30
AlexNeet	94.88	98.50	97.44
GoogleNet	94.92	97.56	97.56
ResNet-50	94.02	96.16	98.54

Now turn to the outcome obtained in Table V that gain by the three models by using the parameter of fixed learning rate at 0.0001 while the number of epochs is increment by 10 and the starting value is 10. It can be seen that when the number of epochs set to 10, the highest accuracy gained is 94.92% by the GoogleNet model. Next, by setting the number of the epoch at 20, the highest accuracy attained is 98.50% by the AlexNet. Lastly, the ResNet-50 model achieves the highest accuracy with 98.54% at the number of the epoch is 30.

To summaries the result gained in Table II and IV, the performance was affected by the number of epochs which depend on the CNN architecture. A few of shallow networks (consist only one hidden layer) [41] such as AlexNet, VGG-16 or ResNet-34 [42] only needed a small value of epochs to achieve the highest accuracy. Meanwhile, the deeper and more complex network like Inception-v3 and ResNet-101 [43] required a very large number of epochs.

C. Influences of the learning rate

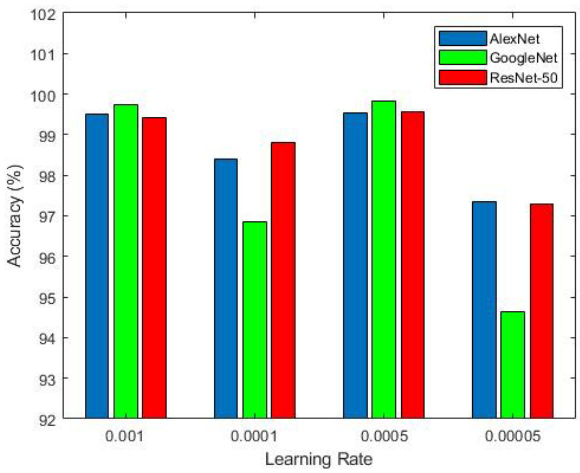

Fig. 1. Variation of accuracy with learning rate using the images selection via the first 70% method.

This part examines the significant parameter such as the number of epochs and learning rate. The proficiency of the CNN model can be affected by the values of learning rate. If the learning process were sped up due to the high value of learning rate which will boost the learning performance loss. Meanwhile, small learning rate values leads to the loss function to reduce gradually. To lessen the loss function for the categorization problem, it is essential to choose the ideal learning rate values. In this research, the CNN models were trained with the value of learning rate at 0.001, 0.0001, 0.0005 and 0.00005. Based on the Figure 1, by using the first 70% image selection method, the best classification accuracy attains when the learning rate value at 0.0005.

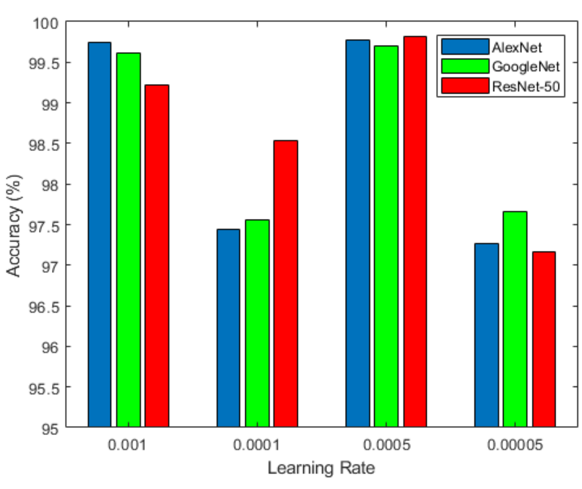

Fig. 2. Variation of accuracy with learning rate using the images selection via the randomized method.

Next, by using the randomized images selection method operating at the same learning rate values as the previous method. The learning rate values that were used is 0.001, 0.0001, 0.0005 and 0.00005. It can be seen in Figure 2, the greatest categorization accuracy achieves when the learning rate at 0.0005.

D. Influence of the number of epochs

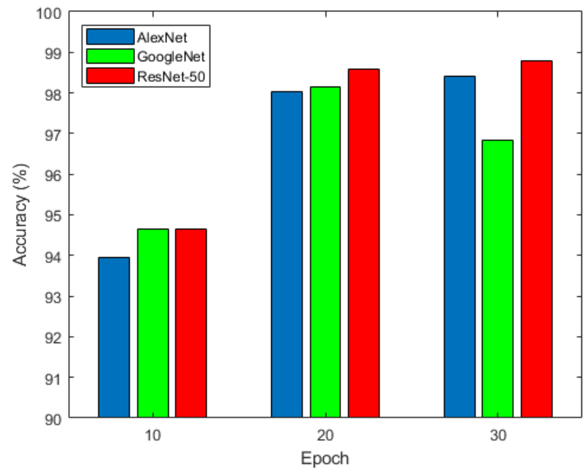

Fig. 3. Variation of accuracy with epoch using the images selection via the first 70% method.

The optimal value of the epochs number was determined concerning categorization accuracy. The three of the CNN models were trained with the number of epochs start from 10 to 30. Firstly, let's take a look at the result obtained when the datasets were trained using the first 70% of images selection method with a various value of epochs. As can be seen in Figure 3, the categorization accuracy growth rapidly from epoch 10 to 20 but slightly increase from epoch 20 to 30. This shows that by increasing the number of epochs it will increase the accuracy.

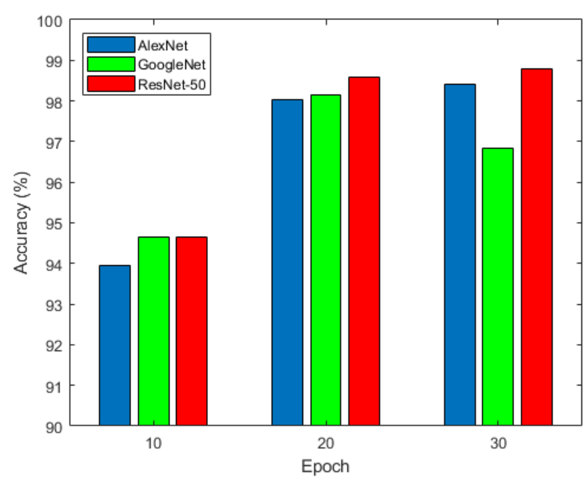

Fig. 4. Variation of accuracy with epoch using the images selection via the randomized method.

By operating at the same epochs value as the previous method, the randomized images selection method undergoes the epochs value start from 10 and end with 30. The outcome shown in Figure 4 indicates that the highest accuracy obtains when the number of epochs was set at 30

V. CONCLUSION

Finally, it may be concluded that the result of this study showed that the optimal parameter configuration is by setting the number of the epoch at 30 while the learning rate set at 0.0005 to obtain the highest accuracy. The CNN model that consistently obtain the highest accuracy was the ResNet-50. ResNet-50 obtains the highest accuracy by using the optimal parameter configuration. Thus, the most appropriate architecture to use for our research in agriculture is ResNet-50 with. In future, this work can be improved more by doing some modification to the CNN models so that the image classification will be more precise.

ACKNOWLEDGMENT

The author would like to give the deepest gratitude to the School of Electrical Engineering, College of Enginerring, Universiti Teknologi MARA (UiTM) Shah Alam campus for the support given to this study.

REFERENCES

[1] J. G. M. Esgario, R. A. Krohling, and J. A. Ventura, "Deep learning for classification and severity estimation of coffee leaf biotic stress," *Computers and Electronics in Agriculture,* vol. 169, p. 105162, 2020, doi: 10.1016/j.compag.2019.105162.

[2] M. Mijwil, *History of Artificial Intelligence*. 2015, pp. 1-8.

[3] V. Panchbhaiyye and T. Ogunfunmi, "Experimental Results on Using Deep Learning to Identify Agricultural Pests," in *2018 IEEE Global Humanitarian Technology Conference (GHTC)*, 2018, pp. 1-2, doi: 10.1109/GHTC.2018.8601896.

[4] S. Indolia, A. K. Goswami, S. P. Mishra, and P. Asopa, "Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach," *Procedia Computer Science,* vol. 132, pp. 679-688, 2018, doi: 10.1016/j.procs.2018.05.069.

[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature,* vol. 521, no. 7553, pp. 436-444, 2015, doi: 10.1038/nature14539.

[6] H. Huang, P. S. Yu, and C. Wang, "An Introduction to Image Synthesis with Generative Adversarial Nets," *ArXiv,* vol. abs/1803.04469, 2018.

[7] C.-C. Chiu *et al.*, *State-of-the-Art Speech Recognition with Sequence-to-Sequence Models*. 2018, pp. 4774-4778.

[8] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 39, no. 4, pp. 640-651, 2017, doi: 10.1109/TPAMI.2016.2572683.

[9] E. Sangineto, M. Nabi, D. Culibrk, and N. Sebe, "Self Paced Deep Learning for Weakly Supervised Object Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 41, no. 3, pp. 712-725, 2019, doi: 10.1109/TPAMI.2018.2804907.

[10] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent Trends in Deep Learning Based Natural Language Processing [Review Article]," *IEEE Computational Intelligence Magazine,* vol. 13, pp. 55-75, 2018, doi: 10.1109/MCI.2018.2840738.

[11] X. Luo, R. Shen, J. Hu, J. Deng, L. Hu, and Q. Guan, "A Deep Convolution Neural Network Model for Vehicle Recognition and Face Recognition," *Procedia Computer Science,* vol. 107, pp. 715-720, 2017, doi: 10.1016/j.procs.2017.03.153.

[12] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing,* vol. 187, pp. 27-48, 2016, doi: 10.1016/j.neucom.2015.09.116.

[13] R. Yamashita, M. Nishio, R. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into Imaging,* vol. 9, 2018, doi: 10.1007/s13244-018-0639-9.

[14] H. S. Abdullahi, R. E. Sheriff, and F. Mahieddine, "Convolution neural network in precision agriculture for plant image recognition and classification," in *2017 Seventh International Conference on Innovative Computing Technology (INTECH)*, 2017, pp. 1-3, doi: 10.1109/INTECH.2017.8102436.

[15] A. Fuentes, S. Yoon, S. C. Kim, and D. S. Park, "A Robust Deep-Learning-Based Detector for Real-Time Tomato Plant Diseases and Pests Recognition," *Sensors,* vol. 17, no. 9, 2017, doi: 10.3390/s17092022.

[16] A. S. B. Reddy and D. S. Juliet, "Transfer Learning with ResNet-50 for Malaria Cell-Image Classification," in *2019 International Conference on Communication and Signal Processing (ICCSP)*, 2019, pp. 0945-0949, doi: 10.1109/ICCSP.2019.8697909.

[17] L. Zhu, Z. Li, C. Li, J. Wu, and J. Yue, "High performance vegetable classification from images based on AlexNet deep learning model," *International Journal of Agricultural and Biological Engineering,* vol. 11, pp. 190-196, 2018, doi: 10.25165/j.ijabe.20181103.2690.

[18] B. Abirami, T. S. Subashini, and V. Mahavaishnavi, "Gender and age prediction from real time facial images using CNN," *Materials Today: Proceedings,* vol. 33, pp. 4708-4712, 2020, doi: 10.1016/j.matpr.2020.08.350.

[19] N. Sharma, V. Jain, and A. Mishra, "An Analysis Of Convolutional Neural Networks For Image Classification," *Procedia Computer Science,* vol. 132, pp. 377-384, 2018, doi: 10.1016/j.procs.2018.05.198.

[20] A. K. Rangarajan, R. Purushothaman, and A. Ramesh, "Tomato crop disease classification using pre-trained deep learning algorithm," *Procedia Computer Science,* vol. 133, pp. 1040-1047, 2018, doi: 10.1016/j.procs.2018.07.070.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[22] V. Agnihotri, "Machine Learning based Pest Identification in Paddy Plants," in *2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2019, pp. 246-250, doi: 10.1109/ICECA.2019.8822047.

[23] H.-T. Nguyen, E.-H. Lee, and S. Lee, "Study on the Classification Performance of Underwater Sonar Image Classification Based on Convolutional Neural Networks for Detecting a Submerged Human Body," *Sensors,* vol. 20, p. 94, 2019, doi: 10.3390/s20010094.

[24] W. Wang, Y. Yang, X. Wang, W. Wang, and J. Li, "Development of convolutional neural network and its application in image classification: a survey," *Optical Engineering,* vol. 58, no. 4, p. 040901, 2019, doi: 10.1117/1.OE.58.4.040901.

[25] T. Lu, F. Yu, C. Xue, and B. Han, "Identification, classification, and quantification of three physical mechanisms in oil-in-water emulsions using AlexNet with transfer learning," *Journal of Food Engineering,* vol. 288, p. 110220, 2021, doi: 10.1016/j.jfoodeng.2020.110220.

[26] M. Li, H. Tang, M. Chan, X. Zhou, and X. Qian, "DC-AL GAN: Pseudoprogression and True Tumor Progression of Glioblastoma Multiform Image Classification Based on DCGAN and AlexNet," *Medical Physics,* vol. 47, 2019, doi: 10.1002/mp.14003.

[27] Y. Li, H. Wang, L. M. Dang, A. Sadeghi-Niaraki, and H. Moon, "Crop pest recognition in natural scenes using convolutional neural networks," *Computers and Electronics in Agriculture,* vol. 169, p. 105174, 2020, doi: 10.1016/j.compag.2019.105174.

[28] M. Z. Alom *et al.*, "A State-of-the-Art Survey on Deep Learning Theory and Architectures," *Electronics,* vol. 8, p. 292, 2019, doi: 10.3390/electronics8030292.

[29] M. A. Kassem, K. M. Hosny, and M. M. Fouad, "Skin Lesions Classification Into Eight Classes for ISIC 2019 Using Deep Convolutional Neural Network and Transfer Learning," *IEEE Access,* vol. 8, pp. 114822-114832, 2020, doi: 10.1109/ACCESS.2020.3003890.

[30] P. Ballester and R. Araujo, *On the performance of GoogLeNet and AlexNet applied to sketches*. 2016, pp. 1124–1128.

[31] V. Paidi, H. Fleyeh, and R. Nyberg, "Deep learning-based vehicle occupancy detection in an open parking lot using thermal camera," *IET Intelligent Transport Systems,* vol. 14, pp. 1295-1302, 2020, doi: 10.1049/iet-its.2019.0468.

[32] C. Szegedy *et al.*, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1-9, doi: 10.1109/CVPR.2015.7298594.

[33] S. Boumaraf, X. Liu, Z. Zheng, X. Ma, and C. Ferkous, "A new transfer learning based approach to magnification dependent and independent classification of breast cancer in histopathological images," *Biomedical Signal Processing and Control,* vol. 63, p. 102192, 2021, doi: 10.1016/j.bspc.2020.102192.

[34] K. Thenmozhi and U. Srinivasulu Reddy, "Crop pest classification based on deep convolutional neural network and transfer learning," *Computers and Electronics in Agriculture,* vol. 164, p. 104906, 2019, doi: 10.1016/j.compag.2019.104906.

[35] H. K. Suh, J. Ijsselmuiden, J. W. Hofstee, and E. J. van Henten, "Transfer learning for the classification of sugar beet and volunteer potato under field conditions," *Biosystems Engineering,* vol. 174, pp. 50-65, 2018, doi: 10.1016/j.biosystemseng.2018.06.017.

[36] K. Junaid, M. Rehman, and M. Bari, "Detection and Classification of Rice Diseases: An Automated Approach Using Textural Features," *Mehran University Research Journal of Engineering and Technology,* vol. 38, pp. 239-250, 2019, doi: 10.22581/muet1982.1901.20.

[37] W. Nawaz, S. Ahmed, M. Tahir, and H. Khan, "Classification Of Breast Cancer Histology Images Using ALEXNET," 2018, pp. 869-876.

[38] Fulton, Dolezel, Harrop, and Yan, "Classification of Alzheimer's Disease with and without Imagery Using Gradient Boosted Machines and ResNet-50," *Brain Sciences,* vol. 9, p. 212, 2019, doi: 10.3390/brainsci9090212.

[39] X. Wu, C. Zhan, Y. Lai, M. Cheng, and J. Yang, "IP102: A Large-Scale Benchmark Dataset for Insect Pest Recognition," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 8779-8788, doi: 10.1109/CVPR.2019.00899.

[40] L. A. da Silva, P. O. Bressan, D. N. Gonçalves, D. M. Freitas, B. B. Machado, and W. N. Gonçalves, "Estimating soybean leaf defoliation using convolutional neural networks and synthetic images," *Computers and Electronics in Agriculture,* vol. 156, pp. 360-368, 2019, doi: 10.1016/j.compag.2018.11.040.

[41] H. Mhaskar, Q. Liao, and T. Poggio, "When and Why Are Deep Networks Better Than Shallow Ones?," *Proceedings of the AAAI Conference on Artificial Intelligence,* vol. 31, no. 1, 2017.

[42] K. Bressem, L. Adams, C. Erxleben, B. Hamm, S. Niehues, and J. Vahldiek, "Comparing different deep learning architectures for classification of chest radiographs," *Scientific Reports,* vol. 10, 2020, doi: 10.1038/s41598-020-70479-z.

[43] L. Mattheuwsen and M. Vergauwen, "Manhole Cover Detection on Rasterized Mobile Mapping Point Cloud Data Using Transfer Learned Fully Convolutional Neural Networks," *Remote Sensing,* vol. 12, no. 22, 2020, doi: 10.3390/rs12223