Transfer Learning-based Weed Classification and Detection for Precision Agriculture

Nurul Ayni Mat Pauzi, Seri Mastura Mustaza*, Nasharuddin Zainal, Muhammad Faiz Bukhori
Department of Electrical, Electronic & Systems Engineering, Universiti Kebangsaan Malaysia (UKM), Bangi, Selangor, Malaysia

***Abstract*—Artificial intelligence (AI) technologies, including deep learning (DL), have seen a sharp rise in application in agriculture in recent years. Numerous issues in agriculture have led to crop losses and detrimental effects on the environment. Precision agriculture tasks are becoming increasingly complicated; however, AI facilitates huge improvement in learning capacity brought about by the advancements in deep learning techniques. This study examined how CNN and VGG16 (transfer learning) were used for weed classification for the application of spraying herbicides selectively in palm oil plantations based on the type of optimizer, values of learning rate and weight decay used on the models. The result shows that the VGG 16 BN model with Adagrad optimizer, learning rate value of 0.001 and weight decay of 0.0001 shows the average accuracy of 97.6 percent and highest accuracy of 99 percent.**

***Keywords*—Artificial intelligence; deep learning; CNN; transfer learning; VGG16**

I. Introduction

The integration of artificial intelligence (AI) in agriculture has been a subject of ongoing research and application; however, recent years have seen a substantial escalation in the adoption and advancement of AI technologies in this domain. Various challenges such as weed infestation and uncontrolled use of herbicide have resulted in crop losses and negative environmental repercussions. Innovative solutions that leverage AI's adaptability, precision, affordability, and overall higher efficiency are required to overcome these obstacles. In recent years, advances in deep learning techniques have led to a significantly better learning capacity, enabling the approach to handle increasingly complex tasks in the field of precision agriculture.

Deep learning, a branch of AI, is a rapidly evolving field which has seen increased adoption in various fields. Deep learning emerged as a powerful tool used in many applications such as image recognition and classification, and it has extended its impact in vital areas such as agriculture, medicines, finance and more. This revolution has been primarily driven by the availability of vast amounts of data (big data) and the advancement of technology in computing power, some of which can be accessed for free such as in Google Colaboratory.

Deep learning utilized the use of algorithm to learn from data enabling it to perform predictions or decisions with high accuracy and efficiency. Deep learning is a subset of machine learning, which involves the use of neural networks which are used to analyze large dataset, with the aim to simulate the structure and function of human brain. These neural networks are highly effective in solving complex problems such image and speech recognition due to its ability to learn from unstructured data [1], [2], [3]. Deep learning has become an indispensable tool in the development of intelligent systems, paving the way for innovations across diverse industries.

The most prominent deep learning approach is the Convolutional Neural Networks (CNNs). It was first introduced by LeCun et al. [4], [5] for the purpose of handwritten digits classification. In a CNN, there is an input layer, hidden layers and an output layer. The two core structures in CNN are the convolutional layer and the pooling layer. The convolutional layer share weights, and the pooling layer lowers the data rate from the layer below by subsampling the convolutional layer's output [6]. In building a CNN algorithm, the most frequently used hidden layers in the CNN algorithm are, convolutional layers, fully connected layers, normalization layers, and pooling layers.

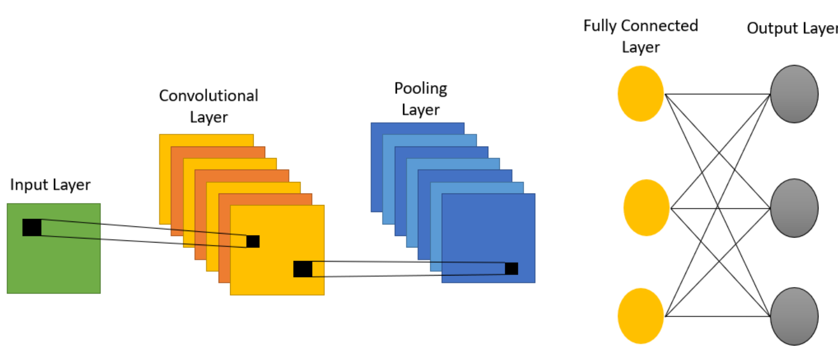


Fig. 1. CNN architecture.

*Corresponding Author.

In order to form a more complex CNN models, additional layers can be added to the CNN architecture. The CNN architecture has proven to be an outstanding solution in most computer vision problems. Since 2011, CNN layers have enhanced deep learning models for tasks involving images, and at this point, CNN layers are used in the majority of DLs [7], [8]. Fig. 1 shows the basic CNNs architecture. Fig. 2 and Fig. 3 shows an example of the operation carried out in convolution and pooling layer.

Transfer learning allows a model to be taught and refined for one task, then adapted for a related task, leveraging prior knowledge to enhance performance and efficiency in the new context. This technique exemplifies the ability to use insights gained in one domain to improve outcomes in another. Data sets smaller than the real training datasets were fed into pre-trained models [9], [10].

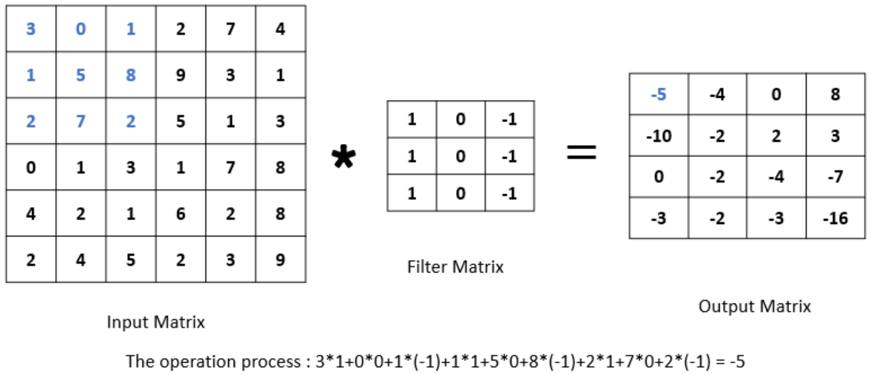


Fig. 2. Operation carried out by convolution layer.

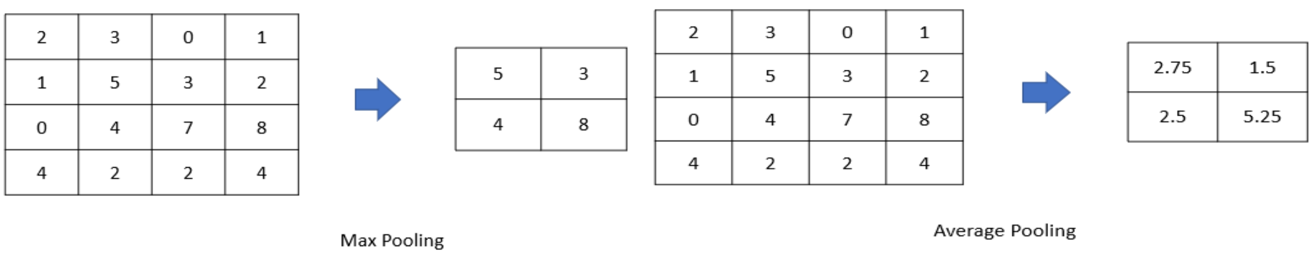


Fig. 3. Operation carried out by Max Pooling and Average Pooling.

Max-pooling layer picks the maximum number from the layer input within a selected window, while an average-pooling layer computes average values over the selected window.

The VGG16 and VGG16_BN model was trained using Image Net, and it was repurposed to learn (or shift) features so that it can become proficient on a new dataset for this project (weed images). Rather than starting from scratch with random weight initialization, initial training can be carried out using the Image Net dataset and Transfer Learning, which enables to better fit the new dataset/task utilizing the learned features and model structure. The network architecture and dataset attributes need to be tested and adjusted to determine which factors affect classification accuracy.

The VGG concept was first presented by the University of Oxford's Visual Geometry Group [6], [9]. Their extremely complex ConvNet is made up of sixteen weight layers, comprising three fully connected layers and thirteen convolutional layers with a 3x3 filter size. Both the padding and the convolution stride are set at one pixel. Five max pooling layers, which come after some of the convolutional layers, handle spatial pooling. There is no Local Response Normalization (LRN) in the network, and all weight layers have ReLU nonlinearity. Fig. 4 shows the architecture of the VGG16 algorithm.

In this study, two transfer learning methods, VGG16 and VGG16_BN, were applied to classify weed images for precision agriculture. The use of several optimizers were explored and the impact of hyperparameters, such as learning rate and weight decay were investigated, on enhancing classification accuracies. Additionally, a detailed comparison of the chosen transfer learning methods with traditional Convolutional Neural Networks (CNNs) were provided, focusing on their respective performances and the implications for weed classification. The findings highlight the potential of transfer learning to improve weed classification accuracy, thereby contributing to more efficient and sustainable agricultural practices. Furthermore, this research underscores the importance of hyperparameter tuning and optimizer selection in optimizing model performance for agricultural applications. This study is divided into seven sections. Section I is the introduction, while in Section II, related works are discussed. Dataset is given in Section III. The methodology adopted in this study were discussed in Section IV. In Sections V and VI, the result and discussion are presented, respectively. Lastly, the study concludes in Section VII with recommendations for future studies.

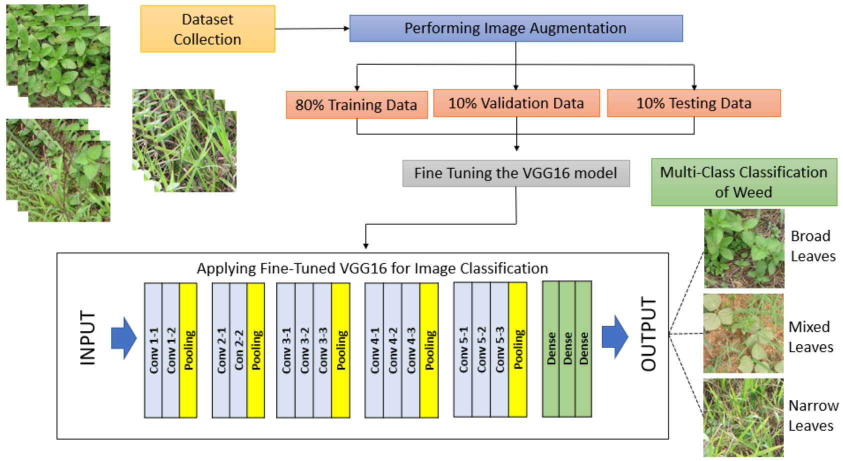


Fig. 4. VGG16 architecture.

II. Related Works

By 2050, there will likely be 9.8 billion people on the planet [10], meaning that more food will be required to feed this population expansion. In contrast, one might expect droughts and natural disasters to increase the value of agricultural food products, as well as reductions in the amount of land and resources available. Conversely, because there is a limited supply of arable land, the rising demand for these items will mostly be satisfied by increasing the usage of agricultural inputs like pesticides, fertilizers, and water [11].

While mechanical weed removal necessitates a significant amount of labour and a huge workforce, the overuse of herbicides poses grave risks to human health and the environment. By selectively spraying herbicides, for example, AI technology can counteract those tendencies and reduce the use of herbicides while increasing the effectiveness of weed management.

In the year 2019, Adhikari et al. [12] proposed semantic graphic learning with convolutional encoder-decoder network for crop line and weed detection in paddy fields. The system shows higher performances compared to bounding box-based object detection.

In 2020, You et al. [13] developed a segmentation model based on deep neural network in order to differentiate weed and crop in complex environment condition. The model is tested on the challenging Bonn and Stuttgart dataset and showed promising potential. In the same year, Arun et al. [14] developed a reduced U-Net integrated with pixel-wise segmentation for weed and crop classification to aid in automated weed removal. The proposed approach managed to achieve a segmentation accuracy of ~95%.

Hussain et al. [15] in 2021 investigated the application of deep convolutional neural network to detect the lamb quarter weeds in potato fields in Canada under various conditions. The models used were GoogLeNet, VGG16 and EfficientNet in both Tensorflow and Pytorch frameworks. The results show excellent performance with more than 90% accuracy with the Pytorch frameworks showing better performance for all models. Jin et al. [16] in the same year proposed the combination of deep learning (CenterNet) and image processing for weed identification in vegetable plantation. The method managed to achieve a precision of 95.6%, recall of 95% and F1-score of 0.953. Then, Ofori and El-Gayar [17] proposed the use of CNN with transfer learning for weed detection among plant seedlings. In the study, the mobile sized EfficientNet was combined with transfer learning and managed to achieve 95.44% classification accuracy on plant seedlings.

Kamath et al. [18] in 2022 investigated the use of semantic segmentation models (SegNet, Pyramid Scene Parsing Network (PSPNet) and UNet) for the segmentation of paddy crop and two types of weeds. The models managed to achieve accuracies of over 90% which shows promising potential to be used for site-specific weed management. In the same year, Mustaza et al. [19] classified weeds using a multilayer perception neural network (MLPNN) with 50 hidden layers as the classifier. Mustaza et al. utilized a modified line filter technique in directional shape feature extraction and the proposed method achieved an accuracy rate of over 97%. Then, G C et al. [20] compared the use of Support Vector Machine (SVM) and VGG16 in performing the classification of four weeds and six crops species. The results shows that the VGG16 classification models outperformed the SVM classifiers. After that, Nasiri et al. [21] utilized deep learning model, UNet as a deep encoder-decoder convolutional neural network (CNN) for the use of pixel-wise semantic segmentation of weed, soil and sugar beet. The results showed an accuracy of 96% and an intersecting over union (IoU) of 84%.

In 2023, Jiang et al. [22] proposed the use of weeding method where herbicides are applied to injured weed tissue. Jiang also designed an intra-row weeding robot to evaluate the performance of the method. The experimental results show better weed removal rate than purely mechanical weeding and shows that it can reduce mechanical weeding operations. The results also show that, compared to normal chemical weeding, the proposed method achieved good weed removal rate while minimizing the use of herbicides. In the same year, Farooque et al. [23] study the performance of a variable rate sprayer for application of agrochemicals such as herbicides. The system utilized CNN for target detection and agrochemicals targeted application in potato field infested with lamb's quarters weed and corn spurry weeds. The system managed to reduce the spray volume up to 47% when compared with conventional chemical application which is applied at a constant rate.

III. Dataset

In this study, the dataset used is made from the images of weed captured using camera from an oil palm plantation in Selangor, Malaysia. The two main weeds found in the oil palm plantations are the broadleaves weed and the narrow grass weed. The weed images are captured using natural light on a sunny day. The images were captured at a close distance approximately 1.5m height at 45° camera viewing angle. This is because it is aligned with the current height and design of sprayer boom tractor frequently used in palm oil plantations in Malaysia. A total of 2377 dataset images is created with image size of 224 x 224 pixels from the images obtained from the plantation. The dataset consists of three classes namely, broad weed, mixed weed and narrow weed. The dataset for each class is divided into approximately 80% train dataset, 10% validation dataset and 10% test dataset. Fig. 5 shows the example of dataset used in the study.

(A) (B) (C)

Fig. 5. The example of dataset used where (A) Broadleaves weed, (B) Mixed weed, (C) Narrowleaves weed.

IV. Methodology

This study evaluates the performance of simple CNNs and transfer learning models which are VGG16 and VGG16 Batch Normalizations (BN) for weed image classification. Firstly, the models were tested with different optimizers with constant parameters which are, learning rate=0.001, weight decay=$1e^{-4}$ and momentum=0.9 for the optimizers that used momentum. Each model was trained for 50 epochs under these conditions. Next, the top three optimizers were identified based on their performance across the three models. Subsequent experiment focused on fine-tuning the parameters of these top optimizers, particularly adjusting the learning rate (LR) and weight decay. The second set of experiment involved training the model for 20 epochs to determine the optimal parameter values. Then, using the optimum parameter found from the experiments, the average value of the accuracy is calculated to determine the best model and parameters.

The algorithm first starts with pre-processing and loading the dataset for train, test, and validation process. Then, the model is loaded, either by initializing the VGG architecture or defining a custom CNN function. Next, the model is trained. The model needs to be set to training mode during the training phase while during the validation phase, it needs to be set to evaluate mode. Then, the input and corresponding labels are applied to the model generating output predictions and calculating the associated loss. During training phase, the model is optimized. Then, the loss and correct predictions are calculated. In training phase, the LR was systematically decayed by a factor of 0.1 every 7 epoch. Next, the accuracy and loss for the current epoch is calculated and the model is saved. This iterative process continues for the total number of epochs specified in the configuration. The final step involves generating a confusion matrix and a comprehensive performance report to assess the model’s classification accuracy, precision, recall and other relevant metrics. This thorough evaluation ensure that the model’s efficacy is rigorously validated, providing critical insights into its predictive capabilities and areas for potential improvement.

A. Pre-Processing

The dataset images for VGG16 and VGG16_BN are resized to 224 x 224 pixels to fit the requirements for the VGG architecture. While the dataset images for CNN are resized to 32 x 32 pixels. Next, the datasets are subjected to random horizontal flip. Then, the datasets are subjected to random affine transformation of the images keeping centre invariant. Lastly, the images are transformed to tensor and are normalized with mean and standard deviation.

B. Platform and Library

The platform utilized for this study is the Google Colaboratory, a cloud-based Python coding environment provided by Google. Using Google Colaboratory is more convenient as most common library is provided inside the platform and they just need be imported. One of the significant advantages of using Google Colaboratory is the availability of online GPUs, ensuring consistent performance irrespective of the user’s local hardware capabilities, such as when using PC without dedicated GPUs. The library that is used in this study is the PyTorch library due to its extensive range of feature and flexibility making it well suited for deep learning task. The dataset used in this research is stored in a folder on Google Drive, which is mounted onto the Colaboratory environment. This allows seamless access to the dataset, with the folder path specified to receive images for training, testing and validation purposes. Leveraging the resources of Google Colaboratory, combined with the robust capabilities of PytTorch facilitates efficient model development and experimentations.

C. Optimizers

To obtain the optimal model for integration with the sprayer boom system in a weed control system in palm oil plantation, certain features can be changed to optimize the performance. The study compared the performance of CNN with VGG-16 and VGG-16 BN focusing specifically on optimizers utilization. The study used the optimizer that can be used for image dataset and supported by the PyTorch library. Optimizers are essential algorithms in deep learning which dynamically adjust a model's parameters during training with the goal of minimizing a predetermined loss function. The optimizer main role is to improve performance by minimizing the error or loss function. By iteratively fine-tuning the weights and biases in response to feedback from the data, these specialized algorithms help in the neural networks learning process. The number of epochs is fixed at 50 epochs based on preliminary tests indicating that accuracy frequently improves up to the 50^{th} epoch and rarely increases beyond this point. This decision also helps mitigate the risk of overfitting. Then, the test accuracy is compared for each optimizer for each model. Both the VGG-16 and VGG-16 BN models used the pre-trained weights from ImageNet to maximize its accuracies. The type of optimizers experimented in this study are Adam, Adadelta, Adagrad, AdamW, Adamax, ASGD, NAdam, RAdam, RMSprop, Rprop and SGD.

D. Learning Rate and Weight Decay

For the top three optimizers which produces the best accuracy, the models were tested with different value of learning rate and weight decay. Learning rate (LR) is the hyperparameter that dictates the extent of adjustments made to the model in response to the predicted error after each update of the model weights. In context of weed image classification, the LR determines the speed at which the model adapts to the classification task. Generally, smaller LR requires more training time due to the smaller weight changes with each update while larger LR provides rapid changes and requires less epoch. A model may converge too soon if the LR is too large while the process may become stuck if LR is too small. Meanwhile, Weight Decay is a regularization technique which penalizes large weights in the network. Weight decay keeps the weights from getting too big by reducing their magnitude. Weight decay helps prevent overfitting and maintain generalization. Keeping weights small will also prevent exploding gradient. The learning rate and weight decay are changed with ten times increment. The learning rate used are between 0.00001 and 10 while the weight decay used are between $1e^{-9}$ and $1e^{1}$. For this experiment, the training is only run for 20 epochs to see the trend of the accuracy performance.

Since the accuracy for each runs differs, the average accuracy is used to clearly analyse the accuracy of the top three optimizers. The learning rate and weight decay used are the optimum values found from the learning rate and weight decay experiment in this same paper. The accuracy reading is taken for 10 times and the average is calculated. Table I shows the pseudocode of the algorithm used in this study.

TABLE I. PSEUDOCODE OF THE STUDY'S ALGORITHM

Algorithm Pseudocode

```
Function Pre-processing
START
Input IMG: Weed images
For i in range IMG
| IMG_aug = transform(IMG)
| return(IMG_aug)
End for
For VGG16/BN: Load model with pretrained weights from ImageNet
For CNN: Define CNN model function
Function VGG16/BN or CNN
START
Input IMG_aug: Augmented Weed Images
Input hyperparameter
Define loss function, optimizer and LR Scheduler
Load model
For epoch in range total epoch
| For i in range IMG_aug:
|   if phase == train
|   | Perform Backward + Optimize
|   End if
|   Running_loss+=loss*input.size(0)
|   Running_corrects+=sum(predictions=labels)
|   if phase == train
|   | Decay LR
|   End if
| Epoch_loss = running_loss/len(train_dataset)
| Epoch_acc = running_corrects/len(train_dataset)
| Save checkpoint and model
Show time elapse
End for
Load model and test model:
 Show classification report
 Show confusion matrix
Load model checkpoint
END
```

V. RESULT

During processing, an input image is presented to the system and subjected to pre-processing stages. Once the images have been transformed according to Section IV (A), the images are then used to train the algorithms. The classification performance is based on the accuracy which is obtained based on Eq. (1).

The result for the first experiment which utilizes the three types of models with different optimizers is as tabulated in Tables II, III and IV.

TABLE II. ACCURACY OF THE VGG16 MODEL

DL Type	Optimizer	Test Accuracy	Time taken for training
VGG16	Adam	194/238 82%	46m 52s
	Adadelta	226/238 95%	48m 6s
	Adagrad	229/238 96%	46m 12s
	AdamW	94/238 39%	44m 28s
	Adamax	227/238 95%	47m 28s
	ASGD	231/238 97%	47m 14s
	NAdam	188/238 79%	49m 13s
	RAdam	231/238 97%	51m 2s
	RMSprop	94/238 39%	47m 46s
	Rprop	120/238 50%	49m 7s
	SGD	231/238 97%	55m 12s

TABLE III. ACCURACY OF THE VGG16 BN MODEL

DL Type	Optimizer	Test Accuracy	Time taken for training
VGG16 BN	**Adam**	230/238 97%	52m 10s
	Adadelta	222/238 93%	54m 60s
	Adagrad	234/238 98%	55m 12s
	AdamW	230/238 97%	53m 2s
	Adamax	232/238 97%	58m 45s
	ASGD	224/238 94%	51m 13s
	NAdam	231/238 97%	53m 7 s
	RAdam	232/238 97%	51m 23s
	RMSprop	95/238 40%	53m 2s
	Rprop	222/238 93%	55m 41s
	SGD	228/238 96%	50m 15s

TABLE IV. ACCURACY OF THE CNN MODEL

DL Type	Optimizer	Test Accuracy	Time taken for training
CNN	**Adam**	157/238 66%	19m 21s
	Adadelta	107/238 45%	19m 41s
	Adagrad	129/238 54%	20m 43s
	AdamW	146/238 61%	14m 24s
	Adamax	140/238 59%	19m 48s
	ASGD	105/238 44%	21m 38s
	NAdam	156/238 66%	19m 20s
	RAdam	151/238 63%	20m 11s
	RMSprop	94/238 39%	19m 51s
	Rprop	125/238 53%	22m 35s
	SGD	136/238 57%	20m 54s

In the beginning of this study, the performance of the models was evaluated with accuracies. However, aside from accuracy, there are several other performance metrics that need to be used in order to accurately measuring the performance of a deep learning model. The performance metrics of the models are calculated by using the formulas in Eq. (1) [19], [24].

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN}$$

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$F-Score = 2 * \left(\frac{Precision * Recall}{Precision - Recall}\right) \quad (1)$$

Where:

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

The other performance metrics are used on top three (3) optimizers among the models. The top three are from the VGG16_BN model which utilized the RAdam, Adamax and Adagrad optimizers.

Another experiment was conducted to identify the optimum learning rate and weight decay to find the best model among top three. The results were as tabulated in Table V and Table VI.

TABLE V. THE ACCURACY OF TOP THREE OPTIMIZER WITH DIFFERENT LEARNING RATE (WEIGHT DECAY :$1e^{-4}$)

Learning Rate	RAdam (%)	Adamax (%)	Adagrad (%)
0.00001	94	92	85
0.0001	96	97	97
0.001	98	97	98
0.01	94	52	48
0.1	40	39	41
1	38	39	48
10	38	45	40

TABLE VI. THE ACCURACY OF TOP THREE OPTIMIZER WITH DIFFERENT WEIGHT DECAY (LEARNING RATE: 0.001)

Weight Decay	Radam (%)	Adamax (%)	Adagrad (%)
$\mathbf{1e^{-9}}$	97	97	97
$\mathbf{1e^{-8}}$	96	97	98
$\mathbf{1e^{-7}}$	97	96	97
$\mathbf{1e^{-6}}$	97	97	98
$\mathbf{1e^{-5}}$	97	97	97
$\mathbf{1e^{-4}}$	98	97	98
$\mathbf{1e^{-3}}$	96	98	97
$\mathbf{1e^{-2}}$	97	97	97
$\mathbf{1e^{-1}}$	96	95	96
$\mathbf{1e^{0}}$	96	96	96
$\mathbf{1e^{1}}$	61	39	60

Fig. 6 and Fig. 7 are the graph of accuracy vs. learning rate and accuracy vs. weight decay shown to better analyse the effect of learning rate and weight decay to accuracy.

For the average accuracy for the top three optimizers, the highest average accuracy is achieved by the Adagrad optimizer with 97.6% accuracy. However, based on Table VII, the highest accuracy achieved by the Adamax and RAdam optimizer is 98% while the highest accuracy for the Adagrad optimizer is 99%.

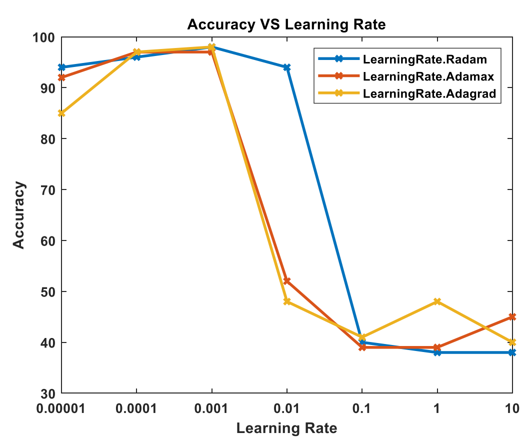


Fig. 6. The graph of Accuracy (%) vs. Learning Rate.

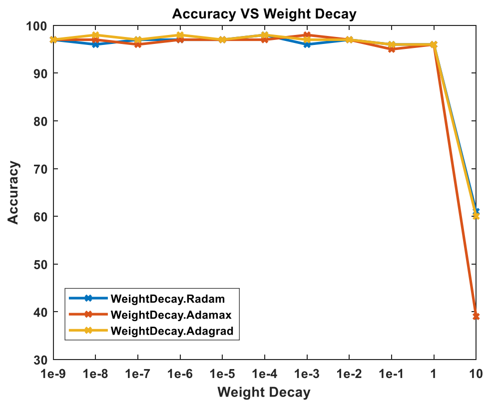


Fig. 7. The graph of Accuracy (%) vs. Weight Decay.

TABLE VII. AVERAGE ACCURACY FOR TOP THREE OPTIMIZERS

	RAdam	Adamax	Adagrad
Highest Accuracy achieved (%)	98	98	99
Average Accuracy (%)	97.3	97.3	97.6

The performance in term of accuracy for the VGG16 BN used in this study was compared with previously reported study on weed classification. The following Table VIII depict the comparisons.

TABLE VIII. STUDY COMPARISON

Study	Dataset	Architecture	Highest Accuracy
Arun et al. [14]	Crop/Weed Field Image Dataset (CWFID)	Reduced U-Net	(Segmentation accuracy) 95.34%
Hussain et al. [15]	Potato crop and Lamb Quarters weed	GoogleNet, VGG16, EfficientNet	92-97% accuracy in every growth stage (EfficientNet)
Jin et al. [16]	Weed and Vegetable	CenterNet	(Precision) 95.6%
Ofori & El Gayar [17]	Weed and Plant Seedlings	EfficientNet	95.44%
This study	Weed (Broad, mixed narrow)	CNN, VGG16, VGG16 BN	97.6% (VGG16 BN)

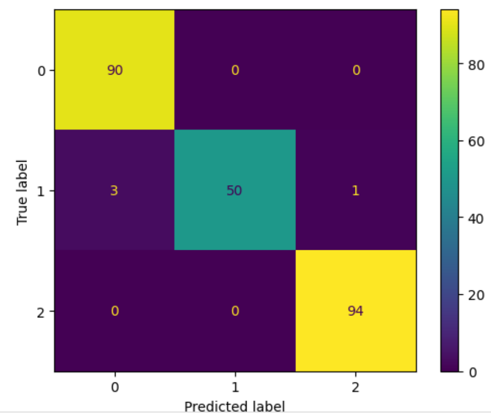
However, aside from accuracy, other performances metrics such as precision, recall and f1-score are also important. Fig. 8, 9 and 10 show the other performance metrics and the confusion matrix for the highest accuracy achieved for the top three optimizers. Based on the figure, the top three optimizers achieved precision, recall and f1-score more than 90% (0.93 to 1) which shows good performance. The confusion matrix of Fig. 8, 9 and 10 shows that the mixed weed labelled 1 (broad - 0, mixed - 1, narrow - 2), have the highest number of misclassifications compared to broad and narrow.

```
Confusion Matrix
Classification Report
              precision    recall  f1-score   support

           0       0.97      1.00      0.98        90
           1       1.00      0.93      0.96        54
           2       0.99      1.00      0.99        94

    accuracy                           0.98       238
   macro avg       0.99      0.98      0.98       238
weighted avg       0.98      0.98      0.98       238
```

(A)



(B)

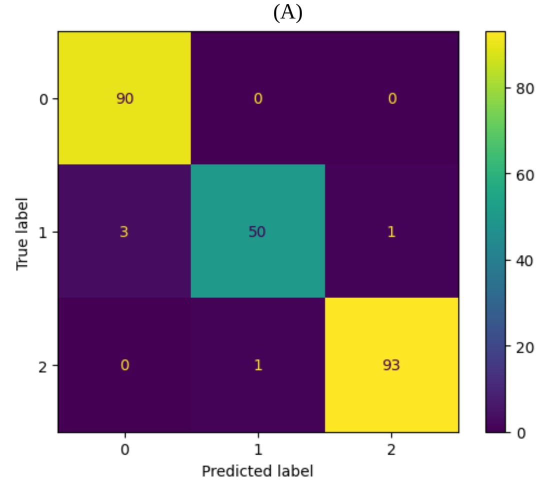
Fig. 8. (A) Classification report and (B) Confusion matrix for Adamax Optimizer.

```
Classification Report
              precision    recall  f1-score   support

           0       0.97      1.00      0.98        90
           1       0.98      0.93      0.95        54
           2       0.99      0.99      0.99        94

    accuracy                           0.98       238
   macro avg       0.98      0.97      0.98       238
weighted avg       0.98      0.98      0.98       238
```

(A)



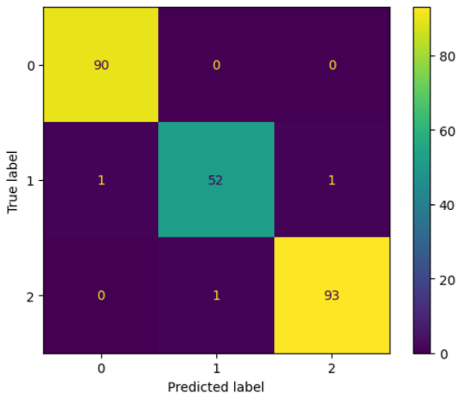
(B)

Fig. 9. (A) Classification report and (B) Confusion matrix for RAdam Optimizer.

Confusion Matrix
Classification Report

	precision	recall	f1-score	support
0	0.99	1.00	0.99	90
1	0.98	0.96	0.97	54
2	0.99	0.99	0.99	94
accuracy			0.99	238
macro avg	0.99	0.98	0.99	238
weighted avg	0.99	0.99	0.99	238

(A)



(B)

Fig. 10. (A) Classification report and (B) Confusion matrix for Adagrad Optimizer.

These results show that the VGG16 BN model using the Adagrad optimizer and learning rate of 0.001 with weight decay of $1e^{-4}$ is an effective model to distinguish between broadleaves, mixed and narrowleaves weed.

VI. Discussion

In this study, the weed classification involves assigning weeds to broadleaves, mixed and narrowleaves categories. 2377 images consisting of 905 broadleaves, 538 mixed, and 934 narrow leaves were used to test the performance of the proposed image classification model.

During preprocessing, in order to vary the dataset images to generalize the model classification, some study will change the hue/color of the image. However, for this particular study, it was crucial to maintain the original hue and color of the images to ensure that the model accurately detects living weeds, which are green rather than misclassifying twigs or dead weeds as live weeds requires spraying. Changing the hue or color could undermine the efficiency of the model and its ability to perform selective herbicide spraying.

Poorer performance was observed for the CNN algorithm used due to the simplicity in the architecture however CNN has the least training time compared with the VGG16 models. This is because the CNN doesn't need to load pre-trained weights like the transfer learning VGG models thus allowing quicker training time. The VGG16 and VGG16_BN achieve higher accuracies because they are transfer learning-based algorithms where the algorithm uses pre trained weights trained on thousands of images from ImageNet.

For the three models used, RMSprop consistently shows low accuracies compared to the other optimizers. This may be because for RMSprop, the learning rate must be defined manually, and the suggested learning rate does not work for every application. Since during the experiment, the study used the predefined learning rate, and this may not be suitable for the study's application.

Based on the experiment, the optimal learning rate for the three optimizers are at 0.001. The accuracy will decrease the further away the value of the learning rate from the optimal value, backward and forward as shown in Fig. 6.

For the weight decay, the optimal value for RAdam and Adagrad optimizer is a $1e^{-4}$ t, while Adamax is at $1e^{-3}$. However, the accuracies remain almost the same from $1e^{-9}$ to 1 and reduce significantly after that as shown in Figure 7. This may be because no matter how much the training epoch is, if the weight decay value is set too big, the model will never quite fit well enough; on the other hand, if the weight decay value is too little, the model can still train well; but the training needs to stop a little early.

The model also produces a number of misclassifications due to several reasons. First, it is possible that some of the photos of broadleaves have stems and green tree branches or very little number of narrowleaves overlapping, leading to misclassification to mixed leaves. The misclassification may also result from dead grasses which does not require herbicide but is still classified as narrowleaves.

VII. Conclusion

In this study, the comparison of the CNN, VGG-16 and VGG-16 BN for weed classification task has been performed. A dataset of images obtained from a local palm oil plantation was used to train, validate and test the algorithms. Based on the result, it can be concluded that the VGG (transfer learning) algorithm shows better accuracy compared to the simple CNN algorithm. Between the two VGG model, VGG 16 and VGG16 BN, the VGG 16 BN with Adagrad optimizer and with learning rate of 0.001 and weight decay of $1e^{-4}$ shows better accuracy. The best model is intended to be used with herbicide spraying system on the sprayer boom tractor. The results obtained indicate that the proposed model is highly reliable and can perform weeds classification with an average accuracy of 97.6% and highest accuracy of 99%. The model can assist in the implementation of an automated weed management system for precision agriculture. For future work, the algorithm can be further improved with attention mechanism to improve performance and robustness of the technique.

Acknowledgment

The work described in this article was supported by the Research University Grant (GUP), of Universiti Kebangsaan Malaysia under grant no GUP-2021-024.

References

[1] K. Sharifani and M. Amini, "Machine Learning and Deep Learning: A Review of Methods and Applications," World Information Technology and Engineering Journal, vol. 10, no. 07, pp. 3897–3904, 2023, [Online]. Available: https://ssrn.com/abstract=4458723

[2] M. Amini, N. S. Safavi, R. M. Bahnamiri, M. M. Omran, and M. Amini, "Development of an Instrument for Assessing the Impact of Environmental Context on Adoption of Cloud Computing for Small and

Medium Enterprises," Aust J Basic Appl Sci, vol. 8, no. 10, pp. 129–135, 2014, [Online]. Available: http://ssrn.com/abstract=2483091Electroniccopyavailableat:http://ssrn.com/abstract=2483091Electroniccopyavailableat:http://ssrn.com/abstract=2483091www.ajbasweb.com

[3] M. Amini, N. S. Safavi, and A. Tolooie, "The Role of Top Manager Behaviours on Adoption of Cloud Computing for Small and Medium Enterprises," Article in AUSTRALIAN JOURNAL OF BASIC AND APPLIED SCIENCES, 2014, [Online]. Available: https://www.researchgate.net/publication/260479568

[4] Y. Lecun, E. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition," in Proceedings of the IEEE, 1998, pp. 2278–2324.

[5] D. Ballard et al., "Backpropagation Applied to Handwritten Zip Code Recognition," Neural Comput, vol. 1, pp. 541–551, 1989.

[6] M. Pak and S. Kim, "A Review of Deep Learning in Image Recognition," in 2017 4th International Conference on Computer Applications and Information Processing Technology (CAIPT), International Conference on Learning Representations, ICLR, 2017.

[7] M. Iman, H. R. Arabnia, and K. Rasheed, "A Review of Deep Transfer Learning and Recent Advancements," Technologies (Basel), vol. 11, no. 2, Apr. 2023, doi: 10.3390/technologies11020040.

[8] M. Iman, H. R. Arabnia, and R. M. Branchinst, "Pathways to Artificial General Intelligence: A Brief Overview of Developments and Ethical Issues via Artificial Intelligence, Machine Learning, Deep Learning, and Data Science," in ICAI 2020 - The 22nd International Conference on Artificial Intelligence, H. R. Arabnia, K. Ferens, D. de la Fuente, E. B. Kozerenko, J. A. Olivas Varela, and F. G. Tinetti, Eds., in Transactions on Computational Science and Computational Intelligence. Cham: Springer International Publishing, 2020. doi: 10.1007/978-3-030-70296-0.

[9] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in ICLR 2015 - International Conference on Learning Representation, Sep. 2015, pp. 1–14. [Online]. Available: http://arxiv.org/abs/1409.1556

[10] "World population projected to reach 9.8 billion in 2050, and 11.2 billion in 2100 | United Nations." Accessed: Jan. 05, 2023. [Online]. Available: https://www.un.org/en/desa/world-population-projected-reach-98-billion-2050-and-112-billion-2100

[11] R. P. Sishodia, R. L. Ray, and S. K. Singh, "Applications of remote sensing in precision agriculture: A review," Remote Sens (Basel), vol. 12, no. 19, pp. 1–31, Oct. 2020, doi: 10.3390/rs12193136.

[12] S. P. Adhikari, H. Yang, and H. Kim, "Learning Semantic Graphics Using Convolutional Encoder–Decoder Network for Autonomous Weeding in Paddy," Front Plant Sci, vol. 10, Oct. 2019, doi: 10.3389/fpls.2019.01404.

[13] J. You, W. Liu, and J. Lee, "A DNN-based semantic segmentation for detecting weed and crop," Comput Electron Agric, vol. 178, Nov. 2020, doi: 10.1016/j.compag.2020.105750.

[14] R. A. Arun, S. Umamaheswari, and A. V. Jain, "Reduced U-Net Architecture for Classifying Crop and Weed using Pixel-wise Segmentation," in 2020 IEEE International Conference for Innovation in Technology, INOCON 2020, Institute of Electrical and Electronics Engineers Inc., Nov. 2020. doi: 10.1109/INOCON50539.2020.9298209.

[15] N. Hussain et al., "Application of deep learning to detect Lamb's quarters (Chenopodium album L.) in potato fields of Atlantic Canada," Comput Electron Agric, vol. 182, Mar. 2021, doi: 10.1016/j.compag.2021.106040.

[16] X. Jin, J. Che, and Y. Chen, "Weed identification using deep learning and image processing in vegetable plantation," IEEE Access, vol. 9, pp. 10940–10950, 2021, doi: 10.1109/ACCESS.2021.3050296.

[17] M. Ofori and O. El-Gayar, "An Approach for Weed Detection Using CNNs and Transfer Learning," in 53rd Hawaii International Conference on System Sciences (HICCS), online, January 5-8, 2021, University of Hawai'i at Manoa, 2021.

[18] R. Kamath, M. Balachandra, A. Vardhan, and U. Maheshwari, "Classification of paddy crop and weeds using semantic segmentation," Cogent Eng, vol. 9, no. 1, 2022, doi: 10.1080/23311916.2021.2018791.

[19] S. M. Mustaza, M. F. Ibrahim, M. H. M. Zaman, N. Zulkarnain, N. Zainal, and M. M. Mustafa, "Directional Shape Feature Extraction Using Modified Line Filter Technique for Weed Classification," International Journal of Electrical and Electronics Research, vol. 10, no. 3, pp. 564–571, 2022, doi: 10.37391/IJEER.100326.

[20] S. G C, Y. Zhang, C. Koparan, M. R. Ahmed, K. Howatt, and X. Sun, "Weed and crop species classification using computer vision and deep learning technologies in greenhouse conditions," J Agric Food Res, vol. 9, Sep. 2022, doi: 10.1016/j.jafr.2022.100325.

[21] A. Nasiri, M. Omid, A. Taheri-Garavand, and A. Jafari, "Deep learning-based precision agriculture through weed recognition in sugar beet fields," Sustainable Computing: Informatics and Systems, vol. 35, Sep. 2022, doi: 10.1016/j.suscom.2022.100759.

[22] W. Jiang, L. Quan, G. Wei, C. Chang, and T. Geng, "A conceptual evaluation of a weed control method with post-damage application of herbicides: A composite intelligent intra-row weeding robot," Soil Tillage Res, vol. 234, Oct. 2023, doi: 10.1016/j.still.2023.105837.

[23] A. A. Farooque et al., "Field evaluation of a deep learning-based smart variable-rate sprayer for targeted application of agrochemicals," Smart Agricultural Technology, vol. 3, Feb. 2023, doi: 10.1016/j.atech.2022.100073.

[24] J. Pardede, B. Sitohang, S. Akbar, and M. L. Khodra, "Implementation of Transfer Learning Using VGG16 on Fruit Ripeness Detection," International Journal of Intelligent Systems and Applications, vol. 13, no. 2, pp. 52–61, Apr. 2021, doi: 10.5815/ijisa.2021.02.04.