Applied Intelligence (2023) 53:20828–20843
https://doi.org/10.1007/s10489-023-04583-8

An augmented attention-based lightweight CNN model for plant water stress detection

Mohd Hider Kamarudin[1] · Zool Hilmi Ismail[1,2] · Noor Baity Saidi[3] · Kousuke Hanada[4]

Accepted: 19 March 2023 / Published online: 24 April 2023

Abstract
Recently, deep learning techniques specifically the Convolutional Neural Networks (CNNs) have reported outstanding results from the application for plant water stress detection based on computer vision system compared to other machine learning methods. However, the size of the conventional CNN models is generally too large for its deployment on resource-limited devices such as mobile smartphone or embedded devices. In this study, a lightweight CNN is proposed by incorporating attention mechanism as an augmentation module into the model. The model was trained, validated, and tested using plant images of *Setaria* grass undergone three water stress treatments. Experimental results show that the proposed method improved the interclass precision, recall, F1-score, and the overall accuracy by more than 9%. Compared to the established lightweight CNN models, the proposed lightweight CNN achieved faster computational time with comparable parameters. In addition, the proposed lightweight model is also efficient when trained on small plant dataset with limited overfitting.

Keywords Plant water stress · Computer vision · Lightweight convolutional neural network · Attention mechanism · Small dataset

1 Introduction

Water stress is the main factor that limits agricultural productions worldwide [1]. Climate change, global warming, increasing drought occurrence, and worldwide water shortage has instantly put global food security under alarming condition. The mechanisms of water stress and the effects on plant are highly complex and extremely influential on the growth and yield [2]. Measuring water stress in plant can not only improve the knowledge of the vegetation wellbeing but also provide information required for precision irrigation management [3]. Numerous methods for plant water stress identification have been developed over the years based on the measurement of soil moisture, meteorological variables, and leaf water potential. These measurements have been shown to be effective indicators of water stress in plant however, the measurement process is slow, destructive, and unsuitable for real-time water stress detection.

Research on computer vision for rapid and non-destructive plant water stress detection have been continuously progressing for over two decades now [4–6]. Traditional approach for image-based plant water stress detection had been based on hand-engineered features such as color, texture, and structure-based features. This leads to slow image processing limited to expert's knowledge and requires identifying the most relevant features that can give the best water stress interpretation [7]. Along with the development of computer vision technology, deep learning as part of machine learning (ML) techniques have been widely applied to plant water stress detection. [8]. Deep learning (DL) method enables faster water stress detection thanks to its ability to automatically extract features from the plant images [9]. Several models based on deep convolutional neural network (CNN) have been used for plant water stress identification with outstanding results. An, et al. [10] identified plant water stress in maize using

✉ Mohd Hider Kamarudin
hider@graduate.utm.my

✉ Zool Hilmi Ismail
zool@utm.my

1 Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, Jalan Sultan Yahya Petra, 54100 Kuala Lumpur, Malaysia

2 Centre for Artificial Intelligence and Robotics, Universiti Teknologi Malaysia, Jalan Sultan Yahya Petra, 54100 Kuala Lumpur, Malaysia

3 Department of Cell and Molecular Biology, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 Selangor, Malaysia

4 Department of Bioscience and Bioinformatics, Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology, Iizuka, Fukuoka 820-8502, Japan

pretrained Resnet50 and Resnet120 deep CNN models with between 91 and 98% accuracy. Soffer, et al. [11] used pre-trained VGG16 model with Long-Short-Term-Memory (LSTM) concatenation for classification of five water stress levels (no stress, low stress, middle stress, high stress, and very high stress) of corn. The results showed excellent 92% accuracy. Chandel, et al. [12] evaluated three different classical CNN models; AlexNet, GoogLeNet and Inception V3 for plant water stress identification of three different plants. GoogLeNet was found to be more superior compared to the others with an accuracy of 98.3, 97.5 and 94.1% for maize, okra, and soybean respectively. The technology can help farmers especially in the rural areas to detect water stress levels without the need for expert knowledge and with the ease usage of a smartphone [3]. Nevertheless, implementation of such CNN models on mobile and embedded devices can be challenging. The deep CNN architecture may have contributed to the higher accuracy but at the same time requires significant storage space and high processing capability. Cloud computing might be a viable option to overcome the storage and computation conundrum [13, 14] however, it requires a consistent internet connection in which some farming area is hardly to come by.

These days, lightweight CNN based solutions have become popular with many computer vision deployment in various applications including vehicle color monitoring [15], traffic sign recognition [16] and remote sensing image classification [17]. The MobileNet series [18–20] designed by the Google's team are examples of publicly available lightweight CNNs that can be directly implemented on any low-powered devices. The accuracy of these models has been compared to the classical CNNs such as AlexNet, VGG16 and Inception v3 based on the public datasets with good results. Several works have been published highlighting the applicability of the lightweight CNNs in agricultural plant monitoring application. Kamal, et al. [21] constructed lightweight model based on depthwise separable convolution from the MobileNet architectures for plant disease classification form leaf images. The accuracy of the proposed model were comparable to the conventional CNN with faster convergence time. Khaki, et al. [22] used pruned MobileNet as the lightweight base structure for the model WheatNet to detect and count wheat heads from the input images. The model outperformed state-of-the-art lightweight CNN models with the highest accuracy. In the most recent study, Kamarudin and Ismail [23] evaluated several lightweight CNN models including the MobileNets for identifying drought stress plants using visible and near-infrared images. The results showed promising utilization of lightweight CNN models for plant water stress detection that can be applied to mobile terminals.

In this research, a new lightweight CNN model is proposed to classify three classes of plant water stress (drought stress, mild stress, and no stress). However, small CNN network usually has weaker performance on few classes' classification due to limited computational ability to extract features more extensively. The challenge increase when plants undergone water stress condition having similar appearance on different stress levels. To improve the water stress detection ability, a simple attention mechanism was introduced into the lightweight model architecture to enhance feature representation without increasing the network layers considerably. Several other studies have used attention mechanism to boost the representation capability of lightweight CNN by giving emphasis on the region of interest or important features of object of interest. For example, Bao, et al. [24] used convolutional block attention module (CBAM) [25] with the proposed lightweight CNN model called SimpleNet to better differentiate between background and the diseased region of the plant. Tang, et al. [26] implemented squeeze-and-excitation network (SE) [27] blocks in a lightweight ShuffleNet model to increase features concentration of the diseased grape leaves. Bhujel, et al. [28] integrated CBAM module with lightweight CNN to increase network complexity for improved plant disease detection.

In addition, the study also evaluates the ability of the lightweight CNN to be trained on small dataset. In general, deep learning application in agriculture has always been constrained by the limited dataset availability [29]. However, conventional research was mostly using generic deep models designed for large number of image classification datasets, which is computationally inefficient for a small number of many plant datasets [30]. Training a deep CNN network with significant parameters over small dataset with few classes can make the model easily overfit [31]. One of the methods used to overcome the problem of model overfitting during training due to small dataset is dataset augmentation [32]. In this study, we try to see the efficiency of the lightweight CNN model trained on a plant dataset without data augmentation. We also provide mechanism to reduce the effect of overfitting based on the slight modification to the lightweight model structure.

In summary, the contributions of this research are as follows:

a) A novel lightweight CNN model designed for plant water stress detection, suitable for mobile devices deployment with high accuracy compared to other prevailing lightweight models.
b) Simple attention mechanism that can be implemented into lightweight CNN model to improve the classification performance without substantially increase the size and downgrading the performance of the model
c) An efficient lightweight CNN that can be trained on small training plant dataset with no data augmentation and less overfitting achieved.

The subsequent of the paper is organized as follows. Section 2 presents the dataset and methods adopted in this study. Section 3 provides the experimental results and analysis. Finally, Section 4 summarizes the conclusion and points to future directions.

2 Dataset and methods

2.1 Plant water stress dataset

Plant images were obtained from the Donald Danforth Plant Science Centre's publicly available dataset [33]. The dataset includes images of plant shoot of ten *Setaria* grass lines (*S. viridis* (accession A10), *S. italica* (accession B100), and eight randomly selected recombinant inbred line (RIL) populations derived from a cross of *S. viridis* and *S. italica*). *Setaria* grass is a model plant that has been used in many drought-related studies to analyse the plant phenotypes [34]. In this study, all ten genotype variations were trained together to consider the general features of water stress extracted based on the different genotypes.

Four water treatments were performed based on the soil full-water capacity (FC): 100% FC, 66% FC, 33% FC, and 0% FC imposed 17 days after planting (DAP) and maintained for another 17 days. Images prior to 17 DAP were excluded from the sample due to low to none biomass production. Plants that received no water after 17 DAP died within 7 days thus 0% FC images were also omitted from the sample. Because early treatment has no discernible effect on the plants due to available soil moisture, the sample images were taken after 2 days of treatment, from 21 DAP until 33 DAP. One top view and four side view images were taken for each plant. All four-sided images were combined, and the top view images were omitted from the dataset sample. These images were classified into three stress levels of drought stress (33% FC) with label 0, mild stress (66% FC) with label 1, and no stress (100% FC) with label 2.

The dataset originally contains both Red-Green-Blue visible light (RGB) and Near Infrared (NIR) images. In this work, RGB images were used instead being the fact that RGB cameras are inexpensive and widely available. The objective of the research was to develop a lightweight CNN model suitable for low-cost system that can be used in either consumer grade cameras or smart mobile phones. RGB image also contains colour information that is one of the main features that is used to detect water stress in plant [35]. The original RGB plant image resolution is 2454 × 2056 pixels and samples of the images are shown in Fig. 1. The number of total images in each class are shown in Table 1. Training was done on the whole plant shoot as opposed to only leaf as we believe that the plant structure would provide morphological information that aids in detecting the drought-stressed plants.

Table 1 Number of images in each sample class

Water stress severity	Number of images			
	Total	Training	Validation	Test
Drought stress	2281	1597	616	68
Mild stress	2286	1600	617	69
No stress	2382	1667	643	72

2.2 Proposed lightweight CNN

The proposed structure of a lightweight CNN for plant water stress detection is shown in Fig. 2. The model structure primarily consists of basic layers and attention mechanism module. Basic CNN was used to extract overall the image features. CNNs are by design a very efficient algorithm to be used with sparse perception data and still resulted a reasonable performance. Several attention-mechanism modules were added in between the convolutional layers to enhance the features concentration extracted from the plant region.

2.2.1 Basic layers

The structure of the proposed lightweight CNN are based on the conventional feed forward network such as the AlexNet [36] structure but with reduced layers and size. The basic layers used include input layer, convolutional layer, fully connected layer, dropout layer, Rectified Linear Unit (ReLU) activation layer and finally output layer. Input layer is taking 224 × 224 image as the model input. In this study, 5 convolutional layers were used as the base of the lightweight CNN structure. The first four convolutional layers used 3 × 3 kernel size to obtain the global features from the plant image. The fifth convolutional layer used 5 × 5 kernel size to focus on the local features of the plant water stress. The filter number for the first convolutional layer is 32, for the second convolutional layer is 48, for the third and fourth convolutional layer is 64, and the fifth convolutional layer is 104. All convolutional layers are using stride 1. Output from the fifth convolutional layer were then feed into ReLU activation layer before goes onto the fully connected layer with 64 neurons. Dropout layer was placed before the output layer of three classes of stress classification with Softmax probability function. Table 2 shows the full description of the parameters of basic layers used in the proposed lightweight CNN.

2.2.2 Attention mechanism

An attention mechanism based on the spatial attention module introduced in CBAM [25] was incorporated into the lightweight CNN model. Basically, the spatial attention

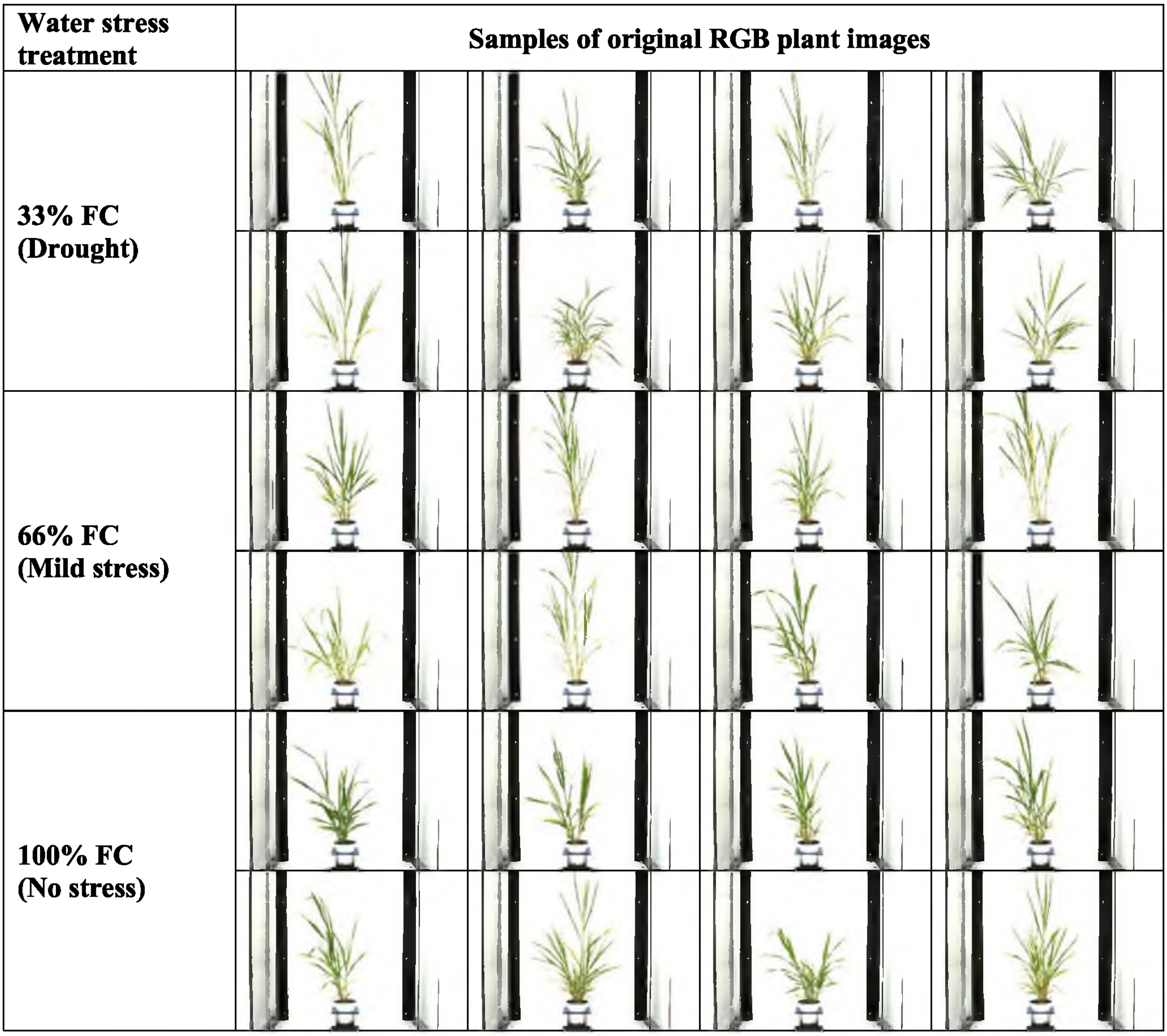


Fig. 1 Process flow for plant water stress detection using lightweight CNN

module was simplified to reduce the computational complexity of the model appropriate for embedded system integration. The simplified attention module was used to enhance the plant water stress's features representation giving emphasis on the shape and pixel intensities interpretation. The module consists of both average-pooling and

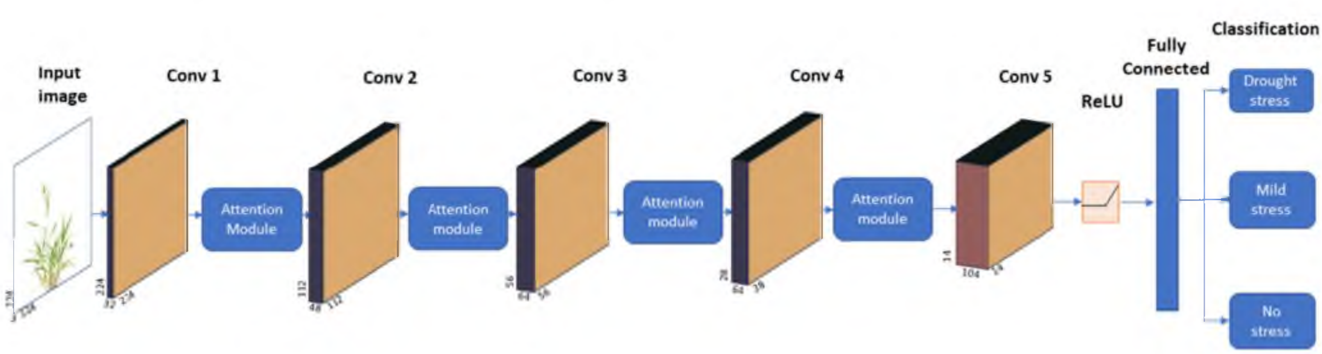


Fig. 2 Proposed lightweight CNN architecture

Table 2 Basic model parameters

Layers	Operation	Input shape	Output shape	Parameters
Input	–	224, 224, 3	224, 224, 3	0
Convolution 1	3×3, 32	224, 224, 3	224, 224, 32	896
Convolution 2	3×3, 48	112, 112, 64	112, 112, 48	27,696
Convolution 3	3×3, 64	56, 56, 96	56, 56, 64	55,360
Convolution 4	3×3, 64	28, 28, 128	28, 28, 64	73,792
Convolution 5	5×5, 104	14, 14, 128	14, 14, 104	326,502
ReLU	–	14, 14, 104	14, 14, 104	0
Fully connected	64	19,992	64	1,279,552
Dropout	0.4	64	64	0
Output	3	64	3	195

maximum-pooling operation taking the feature maps output given as $\mathbf{F} \in \mathrm{R}^{H \times W \times C}$ from the previous convolutional layer as an input matrix. The output from the average and maximum pooling was then feed into ReLU activation represented by ϕ before concatenated and used as input to the next convolutional layer. The attention mechanism computation is shown in eq. (1).

$$Attention = [\phi(AveragePool(\boldsymbol{F})); \phi(MaximumPool(\boldsymbol{F}))] \quad (1)$$

In total, four attention modules were utilized in the proposed lightweight CNN model. The visual structure of a single attention module is shown in Fig. 3.

2.3 Plant water stress detection method

The specific steps of plant water stress detection using the proposed lightweight CNN model are described as follows and the process flow of the methods used in this study is shown in Fig. 4.

2.3.1 Pre-processing image data

Image pre-processing steps which include cropping, resizing, and standardization were performed on the plant images. Image cropping was performed to eliminate unnecessary noises such as unrelated pixels of boxes and pot from the original image to focus only on the plant features. The cropping process also reduces the size of the images which facilitate in the deep training process. Images for this study were taken in a controlled environment setting with a white background and adequate lighting. Due to the consistent background, the plant can be seen clearly from the image thus no plant segmentation is required. Image resizing was done to further reduce the size to 224 × 224 pixels to lower the computational cost. We also standardized the image to improve computation by rescaling the pixel values to the [0,1] range.

2.3.2 Image data augmentation

Image data augmentation was executed to increase the volume of the training dataset as deep learning requires large varieties in samples during training to avoid overfitting. It is also used

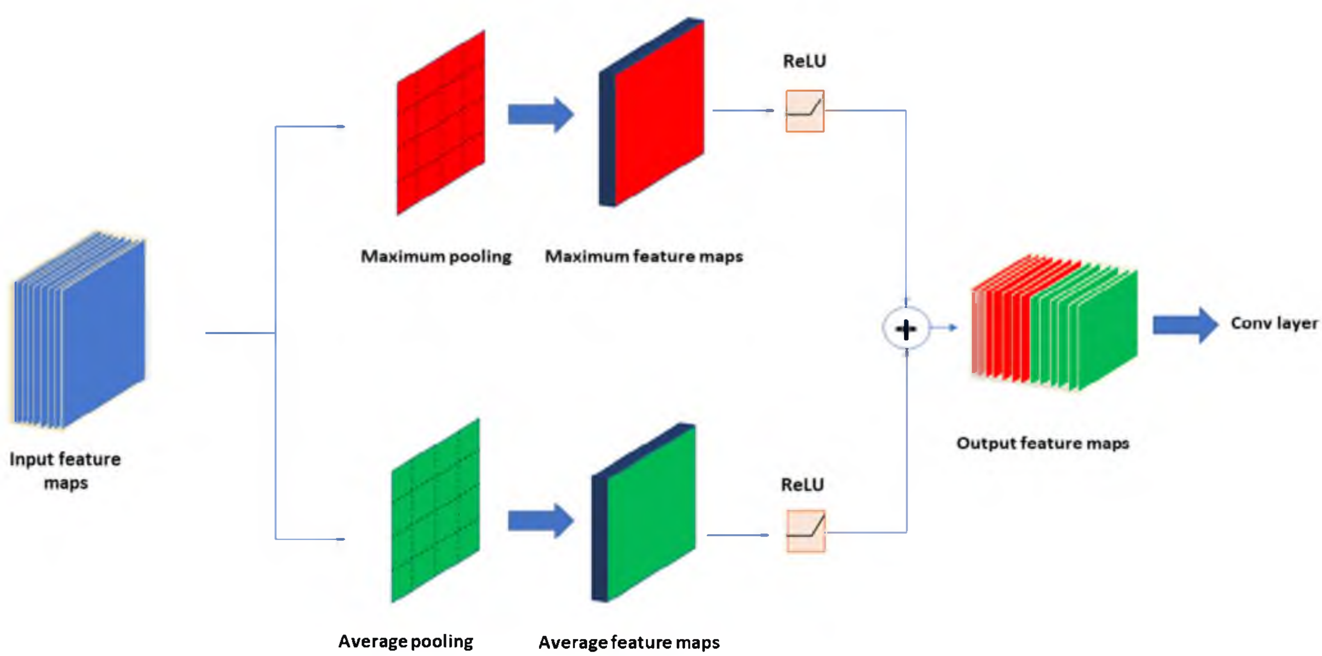


Fig. 3 Proposed simple attention mechanism used in the study

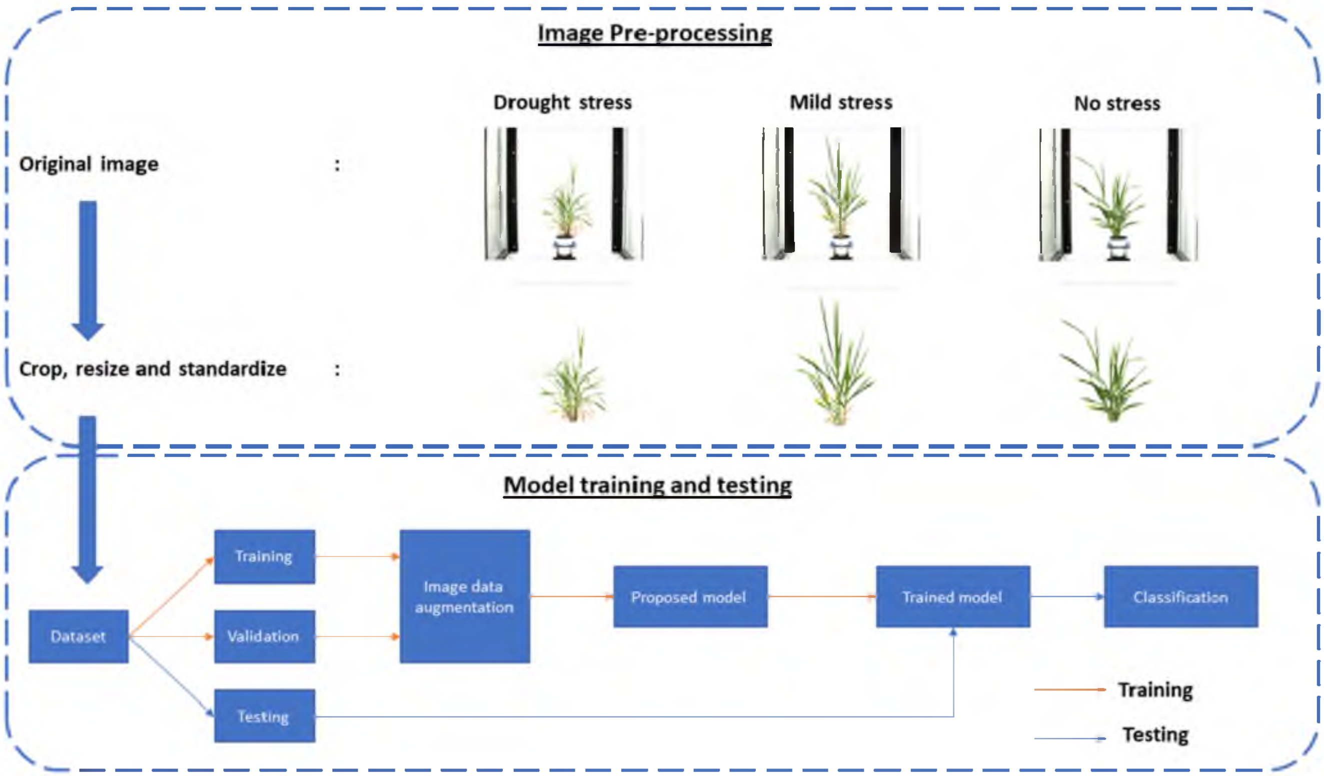


Fig. 4 Confusion matrixes of lightweight CNN models

to increase the generalization ability of the model making it more robust to irregularities. In this study, we have adopted only horizontal flip as an augmentation technique on the sample images to increase the size of the dataset. The reason for this is because we considered the plant morphology to be one of the important features of water stress [37] in addition to the plant's texture and color. Distorting the image heavily for example by rotating or vertically flipping will intuitively eliminate the structural features from the network learning for water stress representation. The augmentation technique was performed on the training and validation dataset resulting in double the volume of plant image samples. The test dataset however was not augmented. We adopted real-time data augmentation which loops over images in batches.

2.4 Training parameters

In the experiment, 70% of the sample images were used for training, and 30% were used for cross-validation. To obtain a generalized measure of classification accuracy, 10% of the validation dataset were used as unseen test dataset. For training parameters, Adam optimizer with default learning rate (0.0001) was used to optimize the hyperparameters [38]. The batch size was set to 16, and the number of epochs was set to 50. The epochs number were set up arbitrarily due to the small size of the model and the use of aggressive dropout. The loss function used was Sparse Categorical Cross Entropy. To reduce the risk of model overfitting, callbacks function was used to perform early stopping with the patience value set to 5 as shown in Table 3 in the hyperparameter settings.

2.5 Setup configuration

The control experiment was carried out in a Windows 10 environment (processor (CPU): Intel core i5 7300HQ; memory: 16G; NVIDIA Graphic Processing Unit (GPU): GeForce GTX 1060). The deep learning framework TensorFlow 2 was used in combination with Cuda10.2 for training in python language platform.

Table 3 Hyperparameter settings

Hyperparameter	Setting
Training and Validation ratio	70:30
Batch size	16
Optimizer	Adam
Learning rate	0.001 (default)
Epochs	100
Loss function	Sparse Categorical Cross Entropy
Early stopping (patience)	5

3 Results and discussion

3.1 Evaluation metrics

To evaluate the proposed lightweight CNN performance for plant water stress detection, the identification accuracy, precision, recall and F1 score were adopted as the evaluation metrics. Accuracy is the ratio of the number of samples correctly predicted to the total number of test samples and reflects the overall performance of the classification. Precision is defined as the ratio of accurately predicted positive samples to all predicted positive samples. Recall is defined as the ratio of accurately predicted positive samples to total positive samples. The F1 score is a comprehensive precision and recall rate index defined by the harmonic mean of precision and recall rates. These standard classification metrics were computed based on Eqs. (2) to (4) [39]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$F1\ score = \frac{2TP}{2TP + FP + FN} \tag{4}$$

where TP is the number of true positive samples, TN is the number of true negative samples, FP is the number of false positive samples, and FN is the number of false negative samples.

Additionally, several metrics, including trainable parameters, model storage size, training time, floating-point operations (FLOPs), and average forward process time (AFT) were measured and compared to analyze the computational efficiency of the proposed lightweight model. The trainable parameters represent the computational ability of the model and is proportional to the storage size of the trained model. The training time is referred to the time it takes to train the model at a specific number of training steps. FLOPs describe the amount of calculation used in a model to measure the computational complexity and AFT represents the time taken to predict a certain number of images on the same hardware (CPU or GPU).

3.2 Water stress identification results

The proposed lightweight CNN for plant water stress identification was compared with other established lightweight CNN models, namely, MobileNets (version 1, version 2, version 3 small, and version 3 large) and NasNet mobile [40] as the baselines. The MobileNet V1 uses depthwise separable convolution to reduce the size and complexity of the model. MobileNet V2 architecture was developed based on inverted residual structure and linear bottleneck with lightweight depthwise convolution to increase the efficiency. Mobilenet V3 is the improved version from the previous MobileNet versions that utilizes the mobile Neural Architecture Search optimization for mobile platform application. NasNet mobile was also developed using the Neural Architecture Search optimization algorithm based on small dataset. These models are the best models available currently for immediate application for mobile and embedded applications. All the models were trained on the same dataset used in this study with the same training hyperparameters. For fair comparison, no pre-training strategy was adopted to train all the lightweight CNN models.

Table 4 demonstrate the performance comparison of all the lightweight CNN models trained from the same plant water stress dataset. The results show that the proposed lightweight CNN with embedded attention mechanism has better identification performance than the established lightweight CNN models. The test accuracy of the proposed model was the highest with 87.02%, more that 9% improvement than the lowest accuracy achieved from the MobileNet V3 small model. The accuracy of the proposed model is in agreement with the study by [41] in which a lightweight CNN was used to detect nitrogen stress of 30 genetically diverse Sorghum plants with images captured in the same setting as the images in this study. It was presented that the proposed model achieved 84% accuracy when trained on two different views. It can be noted all the lightweight CNN models were able to detect drought stress plant with high precision, recall and F1 score values achieved. Nevertheless, the proposed lightweight model achieved the best overall classification rate of drought stress, mild stress, and no stress plants with the highest values of precision, recall and F1 score. The high recall values of the proposed model suggested that the attention mechanism can improve the performance of the small model substantially.

Figure 5 shows the confusion matrixes for all the lightweight CNN models. The results verify the effectiveness of the proposed lightweight CNN to classify drought stress, mild stress, and no stress plants in comparison to other models. It was noted that the Mobilenets (V1, V2, V3 small, V3large) and NasNet mobile networks misclassified mostly the mild stress plants as the no stress plants. The mistake may be due to the high similarity in appearance between the two stress conditions. This has been the main challenge faced for water stress detection in plant. The transpiration rate fluctuated between several parts of the *Setaria* plant

Table 4 Water stress identification results

Models	Accuracy	Labels	Precision	Recall	F1-score
MobileNet V1	0.7212	Drought	0.8311	0.8913	0.8601
		Mild	0.5811	0.6515	0.6143
		No stress	0.7583	0.6233	0.6842
MobileNet V2	0.7019	Drought	0.7866	0.9348	0.8543
		Mild	0.5484	0.6439	0.5923
		No stress	0.8041	0.5342	0.6419
MobileNet V3 small	0.6442	Drought	0.7625	0.8841	0.8188
		Mild	0.4874	0.4394	0.4622
		No stress	0.6423	0.6027	0.6219
MobileNet V3 large	0.6538	Drought	0.7564	0.8551	0.8027
		Mild	0.5258	0.4621	0.4919
		No stress	0.6458	0.6370	0.6414
NasNet mobile	0.7115	Drought	0.8333	0.9058	0.8681
		Mild	0.5616	0.6212	0.5899
		No stress	0.7417	0.6096	0.6692
Proposed model	0.8702	Drought	0.9493	0.9357	0.9424
		Mild	0.8189	0.7879	0.8031
		No stress	0.8411	0.8819	0.8610

making it difficult to differentiate. By using the attention mechanism, the proposed model was able to focus on the area of which the plant from the mild and no stress condition varied.

3.3 Computational performance of lightweight CNN models

Table 5 summarized the computational results of the proposed lightweight CNN model in comparison with other lightweight CNN models. The FLOPs value of the proposed model is the highest among the counterparts even though the parameters is 3 times less than the NasNet mobile which has over 4 M parameters. This shows that the proposed model is much more complex compared to the other lightweight CNN models. The complexity may have been contributed from proposed attention mechanism inserted in the model architecture. However, the training time or the time it takes for the model to converge is much less than the NasNet model which takes almost 5 times longer to train. In addition, AFT of the proposed lightweight models is 3 times better than the smallest model of the MobileNet V3 small. This performance demonstrates the ability of the proposed model to achieve near-real-time speed on the embedded devices. In addition, the size of the proposed model is also comparable with the MobileNet series which are particularly designed for low-powered mobile application. Although the proposed model was not the smallest, it still has the best results compared to the other lightweight models when trained end-to-end on the plant dataset.

3.4 Effectiveness of the attention mechanism

An ablation experiment was performed on the component of the attention mechanism. The experiment was done basically by replacing the attention module with only maximum pooling layer and average pooling layer with ReLU activation layer. The used of single type of pooling operation is of general practise in most conventional CNN models such as in AlexNet that has used maximum pooling and shown good performance in extracting important features from the spatial information. However, combining both features gathered from the maximum and average pooling operation has greatly improved the representation prowess of the small size model showing the effectiveness of the attention mechanism. Table 6 shows the comparison performance in detail.

Model with attention module being implemented achieved higher detection accuracy than model without attention module almost by 10% with better precisions, recalls and F1 scores on each label. The attention mechanism which utilized both maximum pooling and average pooling operation has proven to be more effective in considering the important water stress features in plant than using either maximum pooling or average pooling operation alone. The lowest performance achieved was from using maximum pooling with 78.85% accuracy including the lowest

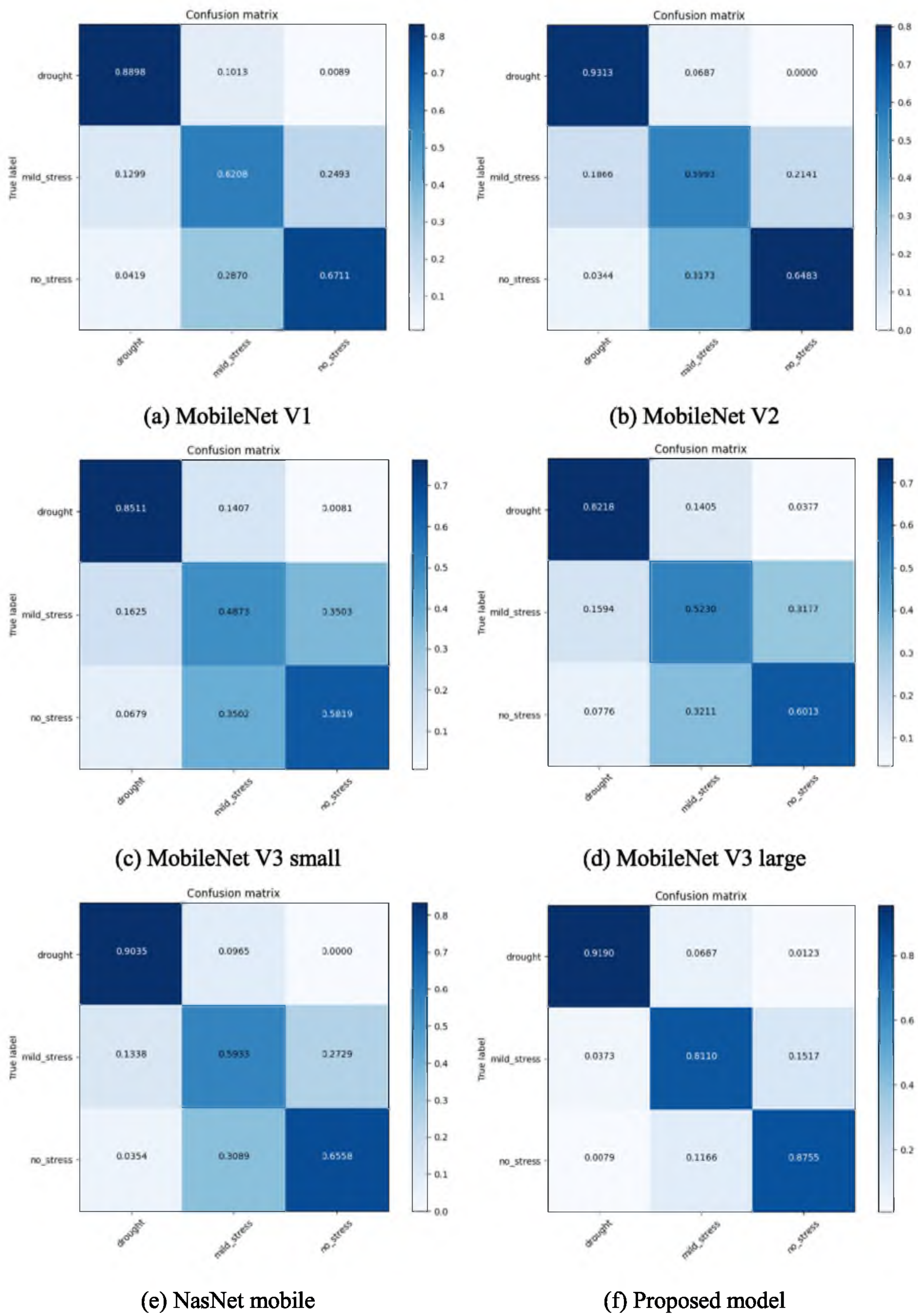


Fig. 5 Different convolutional feature maps of model with attention module, max pooling and average pooling based on the sample plant image

Table 5 Computational performance of different lightweight CNNs

Models	Parameters	Model size MB	BFLOPs	Training time (s)	AFT (ms)
MobileNet V1	3,231,939	12.33	1.14	1380.92	2568.37
MobileNet V2	2,261,827	8.63	0.59	2448.18	3374.54
MobileNet V3 small	940,851	3.59	0.11	338.88	2772.04
MobileNet V3 large	2,999,235	11.44	0.43	1863.05	3770.62
NasNet mobile	4,272,887	16.30	0.55	5515.56	8271.25
Proposed model	1,795,483	6.85	1.38	1203.57	1849.58

of average precision, recall and F1 score value. The number of epochs for model with attention module was also lower than the models without the attention module. This shows that that the model with attention mechanism converge faster than models without the attention module suggesting that the attention module can extract the image features more efficiently. With attention mechanism applied, the number of model parameters increased to 1.79 M which was still however in the category of lightweight model suitable for mobile application.

Figure 6 shows the visualization of the output feature maps of all the convolution layers from the proposed model with attention mechanism, average pooling layer and maximum pooling layer. The figure shows that all models extracted image feature information, such as plant texture, edge, and color in the shallower layer. The visual information of the feature maps decreases and abstract information increases after the third convolutional layer. As can be seen from the feature maps of third and fourth convolutional layers, the model with attention mechanism started to less consider the background features and giving emphasis on the plant features. Models that used maximum pooling and average pooling has more emphasis on the background features than the plant. Apart from focusing on where to get the features, attention module also help enhances the plant features even more based on the warm color of the plant area in the features maps.

3.5 Grad-CAM visualization

To better understand the learning capacity of the attention mechanism in augmenting the lightweight CNN model, visualization of activation regions from the convolutional feature maps in response to the plant sample images and the corresponding stress conditions are shown in Fig. 7. The visualization results represented as heat-maps was generated using 'Gradient-weighted Class Activation Mapping' or Grad-CAM algorithm [42] used to understand more clearly on how the model can speculate the plant stress conditions with high confidence. It is clear from the Grad-CAM images shown that the proposed lightweight attention model was able to localize the region of the water stress descriptors from the plant itself. This also shows that the invariant image background had little influence on the identification results as the model was able to focus on the plant features even though the plant size can be small. It is also worth to note that the features of branches from the plant played a vital role in determining the stress conditions based on the highlighted region of the images. The results were in agreement with the results from Fahlgren, et al. [33] that showed the phenotypic characteristic of plant architecture represented as tiller count was able to differentiate between *Setaria* plant undergone treatment of 100% FC (no stress) and 33% FC (drought stress).

Table 6 Results of attention mechanism ablation experiment

	Parameters	Accuracy	Labels	Precision	Recall	F1-score	Epochs
Attention module	1,795,483	0.8702	Drought	0.9493	0.9357	0.9424	39
			Mild	0.8189	0.7879	0.8031	
			No stress	0.8411	0.8819	0.8610	
Maximum Pooling	1,550,747	0.7885	Drought	0.9385	0.8714	0.9037	50
			Mild	0.7018	0.6061	0.6504	
			No stress	0.7326	0.875	0.7975	
Average Pooling	1,550,747	0.8173	Drought	0.9178	0.9571	0.9370	57
			Mild	0.7479	0.6742	0.7092	
			No stress	0.7748	0.8125	0.7932	

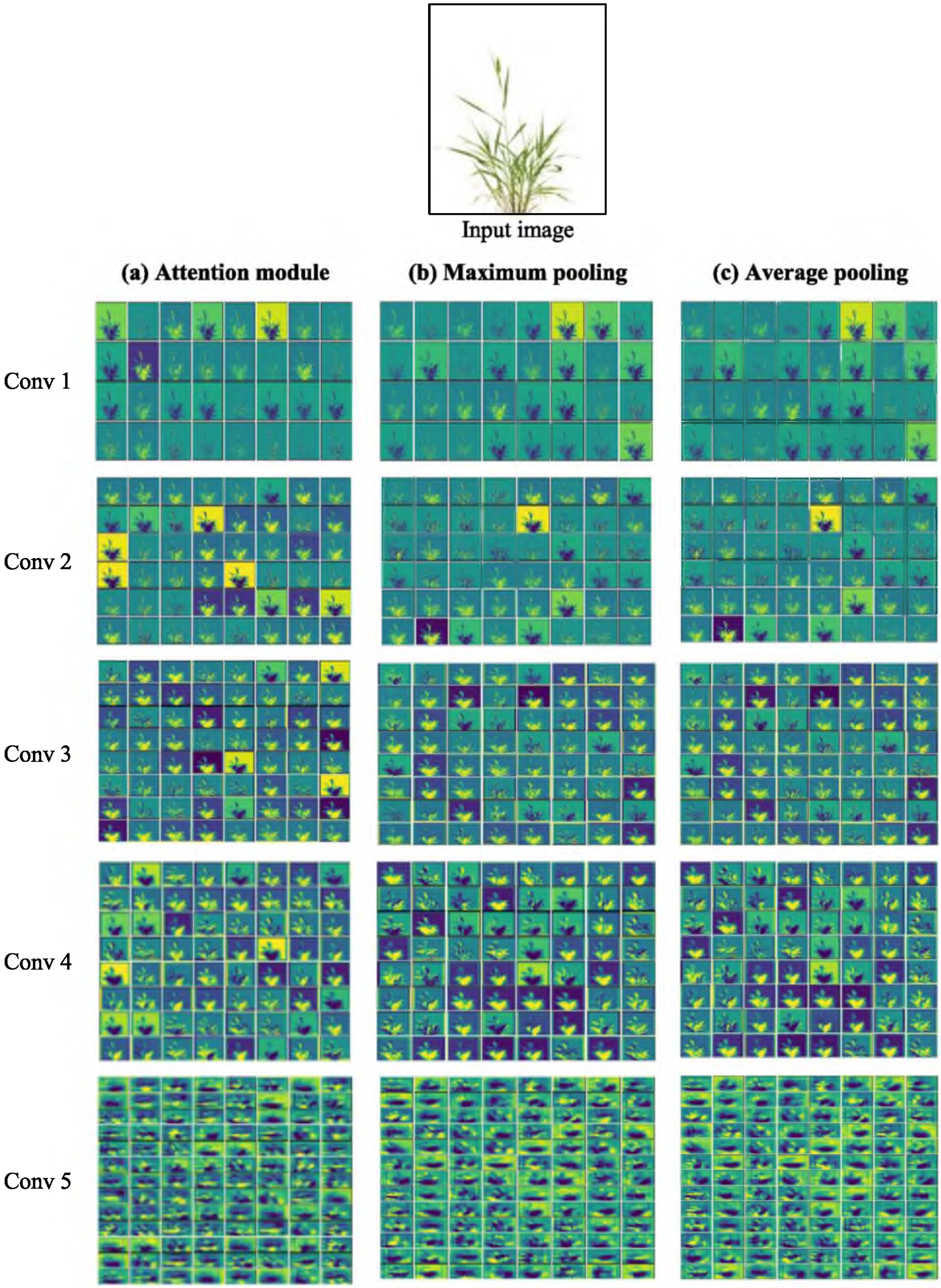


Fig. 6 Gad-CAM visualization of activation maps of the proposed lightweight CNN with attention module based on sample plant images of no stress, mild stress, and no stress label

3.6 Training efficiency on small dataset

To verify the efficacy of the proposed lightweight CNN model to be trained on limited dataset, the image data augmentation part was removed from the process flow. Basically, the dataset used in this study was already considered a small volume dataset with just over 6 thousand of total images and 3 classes. In comparison, standard ImageNet dataset [43] used in most pre-trained network is much larger with 1.2 million images and 1000 classes. The proposed model performance was compared between training with augmented dataset and training without the augmented dataset. The results are shown in Table 7.

The experimental results showed that when the model trained on dataset without data augmentation, accuracy drop by 10% from 0.84 to o.78. From the learning graph shown in Fig. 8, we can see that the training was slightly overfitting due to reduced number of samples in training dataset. Although, the model can still converge at the middle of the epochs and stop when accuracy no longer increased by means of callbacks function. This indicate that our model is not too complex to handle the features from the plant images. Even with small dataset the model can still generalized and give good accuracy. However, an adjustment to the dropout value can reduced the overfitting as shown in Fig. 9. Dropout prevents a layer from seeing twice the exact same pattern. These findings show

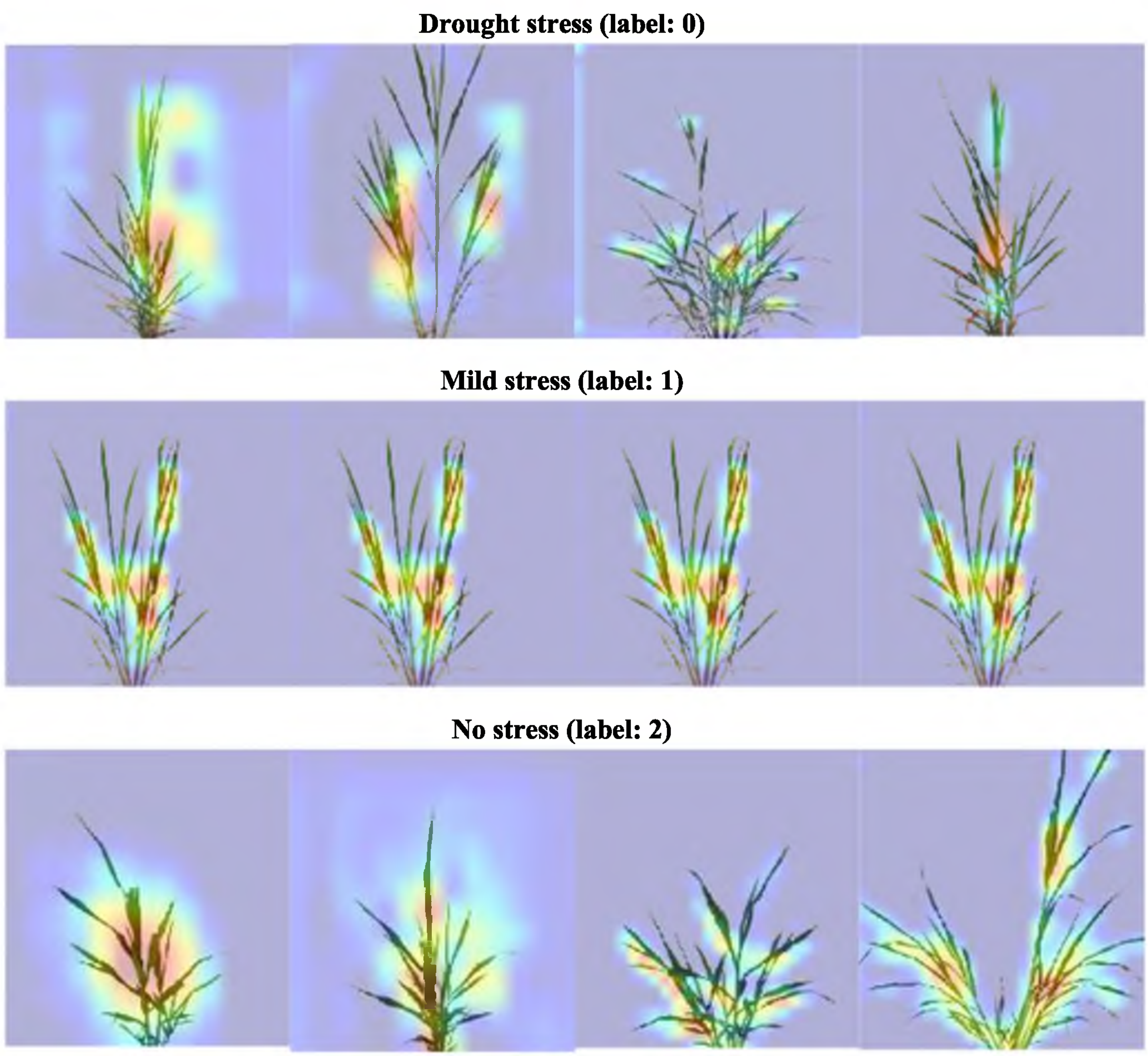


Fig. 7 Learning graph of accuracy and loss versus epochs for lightweight CNN with attention module trained on plant water stress dataset. (**a**) with image data augmentation. (**b**) without image data augmentation

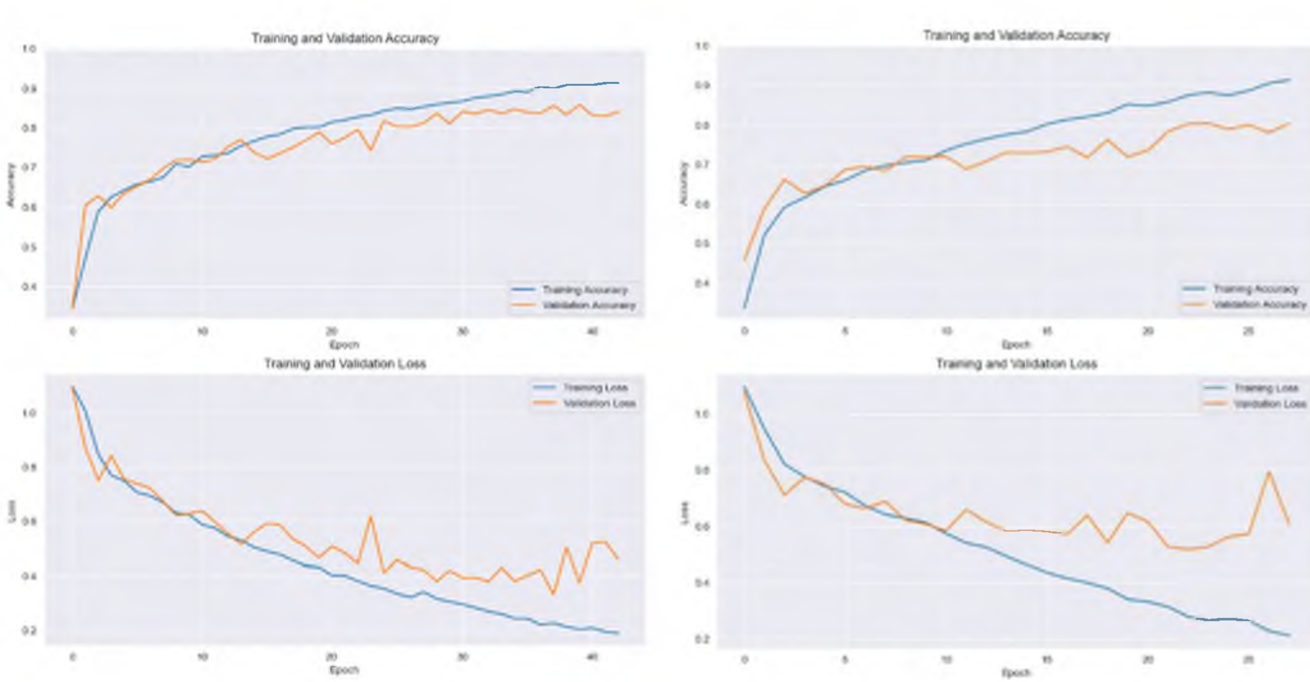
Table 7 Comparison of model trained with and without data augmentation

	Accuracy	Loss	Labels	Precision	Recall	F1-score
With data augmentation	0.8702	0.4060	Drought	0.9493	0.9357	0.9424
			Mild	0.8189	0.7879	0.8031
			No stress	0.8411	0.8819	0.8610
Without data augmentation	0.7260	0.6407	Drought	0.7447	1.0000	0.8537
			Mild	0.6160	0.5833	0.5992
			No stress	0.8252	0.5903	0.6883

that it is possible to have a small size CNN model, to be trained on small dataset, and to have good performance than the deeper structure network. As reported in the study by [44], most lightweight CNNs developed in the area of plant disease detection have been used for small plant datasets with comparable performance to state-of-the-art deep CNN models. In other words, deep learning application with computer vision system for water stress detection is still relevant even with small plant image data if lightweight model is to be utilized.

4 Conclusion

In the light of recent technologies, detecting water stress in plant using smart mobile devices is the rising trend that is highly appealing for smart agricultural application. Although lightweight CNNs have started to get the attention for application in other fields, conventional CNNs are still being used for most plant water stress detection. In this paper, a new lightweight CNN model with the inclusion of simple attention mechanism has been proposed for effective plant water stress detection. Comparison with other lightweight CNN models suggested that the proposed model has faster training and processing time than other models even though the complexity is higher. Classification results have shown that the proposed model has the highest accuracy compared to the established lightweight CNN models of MobileNets and NasNet mobile with parameters comparable MobileNets. The simple attention mechanism has been proven to be efficient to improve the performance of the model while maintaining the small size. Furthermore, the proposed model can be trained on a small plant dataset with limited effect on overfitting.



(a) (b)

Fig. 8 Dropout variation effects on model overfitting

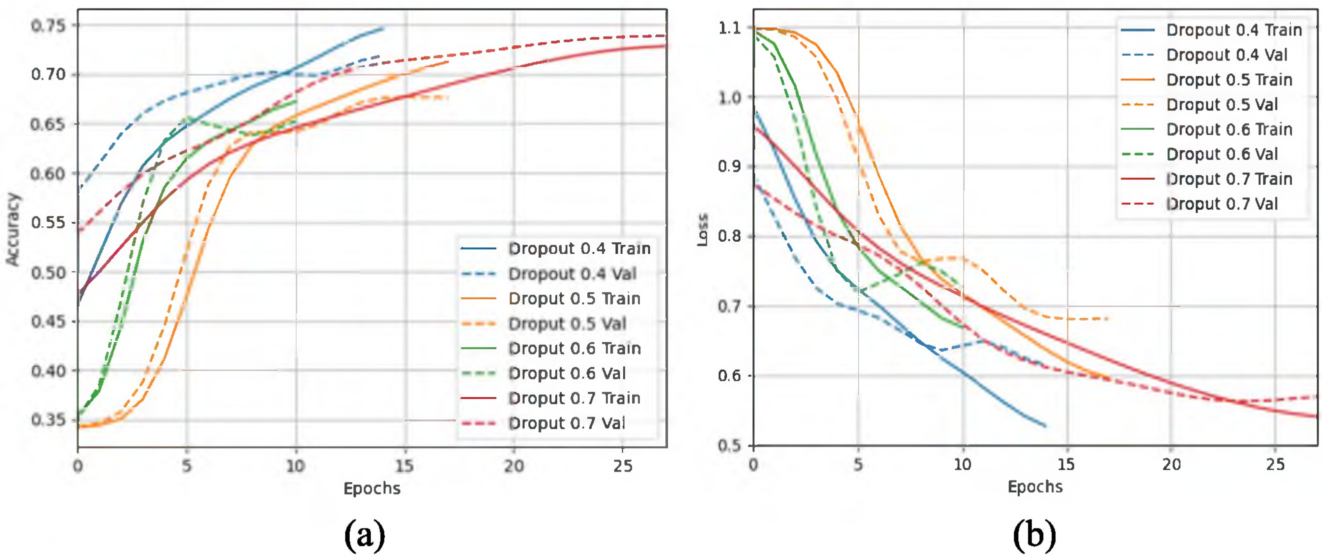


Fig. 9 Learning graph (for training and validation) of accuracy and loss versus epochs for the proposed model based on different configurations of the dropout value ranging from 0.4 to 0.7 which shows effect on model overfitting

Future work will involve the actual mobile deployment of the proposed model to further assess the robustness and effectiveness of the method for easy and fast plant water stress detection under field condition.

Acknowledgements This research was funded by Ministry of Higher Education and Universiti Teknologi Malaysia under research grant no. R.K130000.7843.5F348 and Q.K130000.2543.19H97.

Data and codes availability All original data and code that support the findings of this study are available at https://github.com/hider11/Lightweight-CNN-water-stress.git

Declarations

Conflict of interest All authors have no conflicts of interest to disclose.

References

1. Bertolli S, Mazzafera P, Souza G (2014) Why is it so difficult to identify a single indicator of water stress in plants? A proposal for a multivariate analysis to assess emergent properties. Plant Biol 16(3):578–585
2. Akıncı Ş, Lösel DM (2012) Plant water-stress response mechanisms. Water Stress:15–42
3. Wakamori K, Mizuno R, Nakanishi G, Mineno H (2019) Multimodal neural network with clustering-based drop for estimating plant water stress. Computers and Electronics in Agriculture. p. 105118
4. Seginer I, Elster R, Goodrum J, Rieger M (1992) Plant wilt detection by computer-vision tracking of leaf tips. Trans ASAE 35(5): 1563–1567
5. Kacira M, Ling PP, Short TH (2002) Machine vision extracted plant movement for early detection of plant water stress. Trans ASAE 45(4):1147
6. Shibata S, Kaneda Y, Mineno H (2017) Motion-Specialized Deep Convolutional Descriptor for Plant Water Stress Estimation. In: International Conference on Engineering Applications of Neural Networks. Springer. pp. 3–14
7. Hendrawan Y, Murase H (2011) Neural-intelligent water drops algorithm to select relevant textural features for developing precision irrigation system using machine vision. Comput Electron Agric 77(2):214–228
8. Ramos-Giraldo P, Reberg-Horton C, Locke AM, Mirsky S, Lobaton E (2020) Drought stress detection using low-cost computer vision systems and machine learning techniques. IT Prof 22(3): 27–29
9. Biabi H, Mehdizadeh SA, Salmi MS (2019) Design and implementation of a smart system for water management of lilium flower using image processing. Comput Electron Agric 160: 131–143
10. An J, Li W, Li M, Cui S, Yue H (2019) Identification and Classification of Maize Drought Stress Using Deep Convolutional Neural Network. Symmetry 11(2):256
11. Soffer M, Hadar O, Lazarovitch N (2021) Automatic Detection of Water Stress in Corn Using Image Processing and Deep Learning. In: International Symposium on Cyber Security Cryptography and Machine Learning, Springer. pp. 104–113
12. Chandel NS, Chakraborty SK, Rajwade YA, Dubey K, Tiwari MK, Jat D (2020) Identifying crop water stress using deep learning models. Neural Comput & Applic 33:1–15
13. Zhang W et al (2021) A cloud computing-based approach using the visible near-infrared spectrum to classify greenhouse tomato plants under water stress. Comput Electron Agric 181:105966
14. Freeman D et al (2019) Watson on the Farm: Using Cloud-Based Artificial Intelligence to Identify Early Indicators of Water Stress. Remote Sens 11(22):2645
15. Zhang Q, Zhuo L, Li J, Zhang J, Zhang H, Li X (2018) Vehicle color recognition using multiple-layer feature representations of lightweight convolutional neural network. Signal Process 147: 146–153
16. Haque WA, Arefin S, Shihavuddin A, Hasan MA (2021) DeepThin: a novel lightweight CNN architecture for traffic sign recognition without GPU requirements. Expert Syst Appl 168: 114481

17. Chen L, Wei Z, Xu Y (2020) A lightweight spectral–spatial feature extraction and fusion network for hyperspectral image classification. Remote Sens 12(9):1395
18. Howard AG et al. (2017) Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv: 1704.04861
19. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L-C (2018) Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4510–4520
20. Howard A et al. (2019) Searching for mobilenetv3. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 1314–1324
21. Kamal K, Yin Z, Wu M, Wu Z (2019) Depthwise separable convolution architectures for plant disease classification. Comput Electron Agric 165:104948
22. Khaki S, Safaei N, Pham H, Wang L (2022) Wheatnet: a lightweight convolutional neural network for high-throughput image-based wheat head detection and counting. Neurocomputing 489: 78–89
23. Kamarudin MH, Ismail ZH Lightweight deep CNN models for identifying drought stressed plant. IOP Conf Series: Earth Environ Sci 1091(1):012043. https://doi.org/10.1088/1755-1315/1091/1/012043
24. Bao W, Yang X, Liang D, Hu G, Yang X (2021) Lightweight convolutional neural network model for field wheat ear disease identification. Comput Electron Agric 189:106367
25. Woo S, Park J, Lee J-Y, Kweon IS (2018) Cbam: Convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV). pp. 3–19
26. Tang Z, Yang J, Li Z, Qi F (2020) Grape disease image classification based on lightweight convolution neural networks and channelwise attention. Comput Electron Agric 178:105735
27. Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7132–7141
28. Bhujel A, Kim N-E, Arulmozhi E, Basak JK, Kim H-T (2022) A lightweight Attention-based convolutional neural networks for tomato leaf disease classification. Agriculture 12(2):228
29. Kamilaris A, Prenafeta-Boldú FX (2018) Deep learning in agriculture: a survey. Computers and electronics in agriculture, Review vol. 147, pp. 70–90. https://doi.org/10.1016/j.compag.2018.02.016
30. Barbedo JGA (2018) Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. Comput Electron Agric 153:46–53
31. Raghu M, Zhang C, Kleinberg J, Bengio S (2019) Transfusion: Understanding transfer learning for medical imaging. Advances in neural information processing systems, vol. 32
32. Perez L, Wang J (2017) The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv: 1712.04621
33. Fahlgren N, Feldman M, Gehan MA, Wilson MS, Shyu C, Bryant DW, Hill ST, McEntee CJ, Warnasooriya SN, Kumar I, Ficor T, Turnipseed S, Gilbert KB, Brutnell TP, Carrington JC, Mockler TC, Baxter I (2015) A versatile phenotyping system and analytics platform reveals diverse temporal responses to water availability in Setaria. Mol Plant 8(10):1520–1535
34. M. J. Feldman et al., "Time dependent genetic analysis links field and controlled environment phenotypes in the model C4 grass Setaria," PLoS Genet, vol. 13, no. 6, p. e1006841, 2017
35. Nadafzadeh M, Mehdizadeh SA (2018) Design and fabrication of an intelligent control system for determination of watering time for turfgrass plant using computer vision system and artificial neural network. Precis Agric:1–23
36. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. pp. 1097–1105
37. Foucher P, Revollon P, Vigouroux B, Chasseriaux G (2004) Morphological image analysis for the detection of water stress in potted forsythia. Biosyst Eng 89(2):131–138
38. Kingma Diederik P, Adam JB (2014) A method for stochastic optimization. arXiv preprint arXiv:1412.6980
39. Grandini M, Bagli E, Visani G (2020) Metrics for multi-class classification: an overview. arXiv preprint arXiv:2008.05756
40. Zoph B, Vasudevan V, Shlens J, Le QV (2018) Learning transferable architectures for scalable image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 8697–8710
41. Azimi S, Kaur T, Gandhi TK (2021) A deep learning approach to measure stress level in plants due to nitrogen deficiency. Measurement 173:108650
42. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017) Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision. pp. 618–626
43. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC, Fei-Fei L (2015) Imagenet large scale visual recognition challenge. Int J Comput Vis 115(3):211–252
44. Thakur PS, Khanna P, Sheorey T, Ojha A (2022) Trends in vision-based machine learning techniques for plant disease identification: A systematic review. Expert Systems with Applications. p. 118117

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Mohd Hider Kamarudin is presently pursuing his doctoral studies at the Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia. His research focuses on various areas, including lightweight neural networks, artificial intelligence, agriculture monitoring systems, system modeling, plant-based water stress, and smart sensors.

Dr. Noor Baity Saidi obtained her PhD in Molecular Plant Science from the University of Edinburgh, United Kingdom, on the interaction between the model plant, Arabidopsis thaliana, and the bacterial pathogen Pseudomonas syringae pv. tomato DC3000. She is currently a lecturer at the Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia. Dr. Noor Baity uses molecular tools and next-generation sequencing to examine plant-pathogen interactions and develop biocontrol formulations targeting various plant diseases. She has published more than 30 scientific articles in peer-reviewed journals, including Scientific Reports, Agronomy, Frontiers in Plant Science, and Microbial Ecology. She has received numerous government and industry grants at the national and international levels, including the Australian Centre for International Agricultural Research.

Dr. Zoo IHilmi Ismail who received his Ph.D. degree in Electrical Engineering from Heriot-Watt University in Edinburgh, United Kingdom in 2011, has an impressive academic background that includes obtaining his B. Eng and M. Eng degrees in mechatronics engineering from Universiti Teknologi Malaysia, Skudai, Johor, Malaysia in 2005 and 2007 respectively, and being appointed as a lecturer at the same university in 2011. Currently, he serves as a research member at the Malaysia-Japan International Institute of Technology and Center for Artificial Intelligence and Robotics at UTM Kuala Lumpur, and has previously held positions as a visiting researcher at Kyoto University and Jordan University of Science & Technology in 2014 and 2016, respectively. Additionally, he is a registered professional engineer under the Board of Engineers Malaysia, a member of various professional organizations, including the Society for Underwater Technology, The Institution of Engineering and Technology, and the Institute of Electrical and Electronics Engineers - Oceanic Engineering Society, and a registered chartered marine engineer of the Institute of Marine Engineering, Science and Technology. Dr. Ismail's current research interests include learning control, digital-twin simulation, edge-computing, multi-agent system, and unmanned vehicles, and he has received numerous awards from International RoboCup Competitions (Service Robot Category). Recently, he was appointed as one of the technical committee members for the International RoboCup@Home Competition and the World Robot Summit.

Dr. Kousuke Hanada is currently affiliated with the Department of Bioscience and Bioinformatics at Kyushu Institute of Technology, Japan. He obtained his Ph.D. from the National Institute of Genetics, Japan, where he worked in Dr. Gojobori's laboratory on comparative genomics. After completing his Ph.D., he moved to the United States and conducted research at the University of Chicago and Michigan State University, exploring the evolutionary origins of differential expression and morphological traits in plant evolution. He then returned to Japan and secured an independent position at RIKEN Institute and Kyushu Institute of Technology. Dr. Hanada's research focuses on identifying plant genes using various -omics data, such as transcriptome, epigenome, proteome, metabolome, and phenome, both by comparing different plant species and within a given plant species.