

Maize Leaf Disease Identification with Large and Lightweight Convolutional Neural Models

Mohd Haris Lye^{a,*}, Mohammad Faizal Ahmad Fauzi^a, Lim Kian Ming^b

^a Faculty of Engineering, Multimedia University, Persiaran Multimedia, Cyberjaya, Malaysia

^b Faculty of Information Science and Technology, Multimedia University, Jalan Ayer Keroh Lama, Melaka, Malaysia

Corresponding author: *haris.lye@mmu.edu.my

Abstract— To minimize yield losses in maize plantations, control measures that include early leaf disease detection are essential. In this study, we evaluated extensive and lightweight convolutional neural network (CNN) models to accurately classify maize diseases from leaf images. To achieve a high image classification performance, existing deep learning approaches often use large models that require substantial computational resources. Simpler and lightweight models provide faster inferences but at the expense of lower accuracy in prediction performance. To improve maize leaf disease classification performance on the lightweight SqueezeNet model, the response-based knowledge distillation method was evaluated for model training. In response-based knowledge distillation, the logit output from the last layer of the large model is used in the loss function to train the lightweight model. This enables the lightweight model to learn from the knowledge of large and complex models, thereby improving its predictive accuracy while maintaining a simpler architecture and faster inference. A six-class maize disease dataset was prepared using two publicly available datasets. The dataset was used to train and evaluate the selected large and lightweight models. The large and lightweight model demonstrated high classification accuracy when trained till 40 epochs. The trained SqueezeNet model showed promising performance for accurately identifying various maize leaf diseases with an accuracy of 96.68%. When the model is trained with the response-based knowledge distillation method, the test accuracy improves to 97.13%. Such lightweight models with high accuracy can facilitate the deployment on resource-constrained devices.

Keywords— Knowledge distillation; image classification; lightweight model; plant disease; deep neural network.

Manuscript received 10 Oct. 2024; revised 17 Dec. 2024; accepted 5 Mar. 2025. Date of publication 31 Mar. 2025. International Journal on Informatics Visualization is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.



I. INTRODUCTION

Maize is one of the top three most important food crops, along with rice and wheat [1]. It is a significant food source for humans and livestock. However, maize and other crops are easily infected by diseases that reduce their yield and thus threaten food security [2]. A recent study showed that pests and diseases contribute to a mean yield loss of 22.6% in maize plantations [3]. Current practice requires farmers to manually inspect their crops for signs of disease, which usually manifest through visual patterns on the plant's leaves. This tedious process is not practical for large farms.

Many studies have been conducted to perform automated disease identification in maize crops [4], [5]. The most effective way is through a deep learning approach that leverages high-performing models for image recognition, such as the convolutional neural network (CNN) and recent vision transformer. Survey paper [5] reviewed 70 recent

papers on deep-learning applications for plant disease identification. Many techniques with good accuracy have been proposed. However, most rely on complex deep-learning models. In addition, when the model was trained and evaluated on different datasets involving field images, the performance dropped significantly.

Several papers have shown the effectiveness of convolutional neural networks in maize disease identification. The Inception CNN model was evaluated on a field-collected maize image dataset and obtained an accuracy of 96% [6]. Data augmentation using brightness change was applied. Recent methods have proposed a lightweight CNN model with attention blocks [7], [8]. The smaller CNN model outperforms the mainstream larger models for predicting maize leaf diseases. A vision transformer with reduced model complexity is proposed for maize disease identification [9], [10].

The paper by [11] proposes to solve maize disease severity estimation by two-stage segmentation procedure. In the first stage, the leaves were segmented. The identified region of

interest was then segmented to identify the lesion region. Severity was predicted based on the percentage of the leaf-covered by the lesion associated with the predicted disease. Three types of maize diseases were considered: the gray leaf spot (gls), northern leaf blight (nlb) and northern leaf spot (nls). The proposed method was evaluated using field images.

Applying deep learning models on low-cost edge devices in agricultural settings necessitates reducing model complexity. This requirement arises from the resource-constrained nature of these devices, which are characterized by limited computational capabilities and stringent memory constraints. However, a simple model provides lower accuracy due to its limited representation capacity. Consequently, techniques such as knowledge distillation for model compression [12] and [13] have garnered significant attention for improving the predictive accuracy of small models as a way to reduce the computational burden imposed by complex deep learning architectures.

Knowledge distillation (KD) is a technique used to transfer knowledge from a larger deep-learning model to a smaller lightweight model [14]. In this scheme, the teacher is the large model, and the student is the smaller model. Knowledge distillation typically involves the following steps. Firstly, the teacher model is trained based on the target dataset to achieve high generalization performance. The trained teacher generates soft targets from its output or intermediate features for teaching the student model. This has been shown to improve student performance because the soft target contains useful information that complements the hard target obtained from the ground truth label. This way, the student is trained to imitate the teacher's behavior.

Knowledge distillation differs from transfer learning, where the weights of a model trained on a large dataset are used to initialize a similar model for related tasks [15]. The model is then fine-tuned on a smaller target dataset. Many studies have shown that this results in performance improvement compared to a randomly initialized model.

There are two major approaches to knowledge distillation [16], [14]. This depends on information sources leveraged from the pre-trained teacher model for training the student model. One approach taps into the teacher's output probabilities, also known as the logit-based or response-based method [17]. The other approach relies on the teacher's intermediate representations, often referred to as the hint-based method [18]. Although many new knowledge distillations have been proposed, recent research shows that the standard or vanilla knowledge distillation method still performs well if strong data augmentation and a large dataset is used during training [16].

Knowledge distillation has been applied in various fields of plant disease identification. In [19], a small CNN model, namely the tiny mobileNet-v2, which has 25% of the original model size, is used for plant disease recognition on the 39 classes of the Plant Village dataset. The model trained with standard response-based knowledge distillation from a large CNN model showed improved classification performance. The large CNN model EfficientNet and Xception model were used as the teacher. However, this was only demonstrated on a clean lab-prepared dataset. Paper by [20] with student CNN model composed of just four layers with 3.71 million parameters, standard KD can improve accuracy ranging from

3.5% to 4.73%. The student model was distilled from various large models namely Resnet50, DenseNet121 and Xception. In the work [21], the maize disease identification on the Plant Village dataset is performed with object detection approach. The YOLOv5s model is simplified to reduce its parameter by 15%. To enhance its performance, channel-wise knowledge distillation [22] is used. After training with knowledge distillation from the bigger YOLOv5m model, the mean average precision mAP (0.5) improved by 3.8%, and the parameter size decreased by 15.5%. Similar results on knowledge distillation with object detection methods for plant disease have also been reported [13]. In a related work [23] Knowledge distillation with data augmentation was applied to the Plant Doc dataset. Literature studies show that there is still limited work that applies knowledge distillation to plant disease detection problems. Most of the papers only use to clean image datasets that are captured in the lab. However, plant diseases should ideally be identified in the field. Such images are challenging to process due to challenging background objects. This study aims to solve the problem by testing a small model in field-collected images.

In this paper, we aim to evaluate the performance of the large and the small, lightweight CNN models and assess the impact of using the basic response-based knowledge distillation method [24] for training a small CNN model. The large models include DenseNet-121 [25], Resnet-50 [26] While the two lightweight models MobileNet [27] and SqueezeNet [28] are selected for evaluation. Knowledge distillation is used with the SqueezeNet model.

DenseNet-121 and ResNet-50 each have 121 and 50 layers, respectively. Both models use skip connections to construct deep layers that mitigate the vanishing gradient problem. In DenseNet, each layer has skip connections to all other layers in a feed-forward manner. These large and deep models have been used successfully in many computer vision applications and demonstrate impressive performance [29], [30], [31] However, these models are not suitable for low-end edge devices because they require large memory space and incur high latency during inference. To address this issue, a number of simple models, such as MobileNet and SqueezeNet, have been proposed.

SqueezeNet is a type of convolutional neural network with 18 layers [28]. The model's size is less than 0.5 MB. This significant complexity reduction is achieved by using the squeeze and expansion layers. The squeeze layer, consisting of 1x1 convolution filters, reduces the number of channels. The expansion layer, using 1x1 and 3x3 convolutional filters, increases the number of channels. The squeeze and expansion layers are grouped in a module. These modules, known as fire modules, are stacked to form the architecture of the SqueezeNet model.

MobileNet introduces a layer that uses depthwise separable convolutions to reduce model parameters [27]. This layer uses depthwise convolution followed by (1x1) pointwise convolution to replace the standard convolution operation. This strategy has been shown to significantly reduce model parameters and increase computation speed with a minor impact on predictive accuracy [27], [32]. A comparison of the size of parameters, model file size, and inference time per image is shown in Table 1. The size of the lightweight model is significantly smaller than that of the large model, such as

Resnet-50 [33], and thus, it is feasible for usage on low-end edge devices such as Raspberry Pi, where the processor speed is low and memory size is restricted. The architectural visualization of the large and small lightweight models used are shown in Figure (1) – (4).

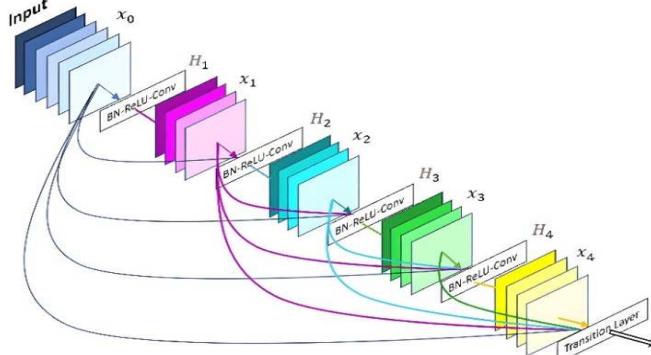


Fig. 1 DenseNet architecture diagram [25]

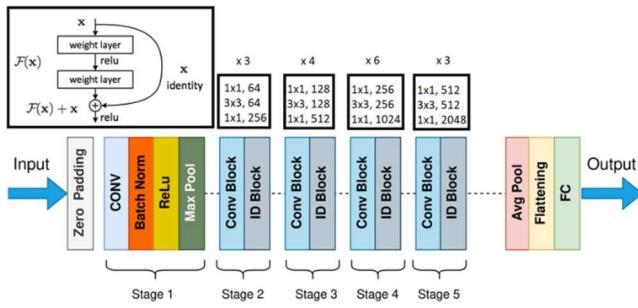


Fig. 2 ResNet50 architecture diagram [34]

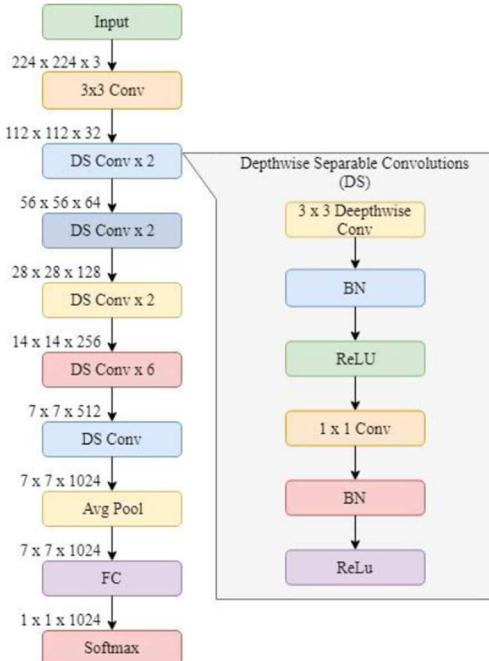


Fig. 3 MobileNet architecture diagram [33]

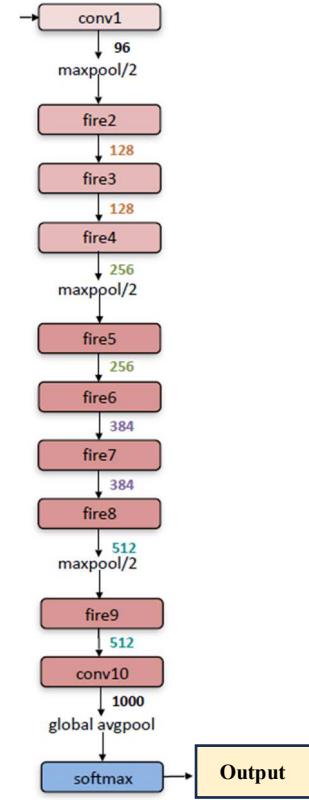


Fig. 4 SqueezeNet architecture diagram [28]

TABLE I
NUMBER OF PARAMETERS, FILE SIZE AND MODEL INFERENCE TIME FOR THE CNN MODELS EVALUATED IN THE EXPERIMENT

Model	Number of Parameters (million)	File Size (MB)	Model Inference Time (seconds)
MobileNet-V2	3.50 M	13.9	0.066
SqueezeNet1.1	1.24 M	4.8	0.066
Resnet-50	25.56 M	100.1	0.074
DenseNet-121	7.98 M	31.6	0.082

II. MATERIALS AND METHOD

A. Method

We use the standard knowledge distillation approach [24] to improve the small lightweight CNN model. This enables the lightweight model to learn from the actual label and complex model response. In this case the lightweight model is named as the student model while the deep complex model is named as the teacher.

The student model is trained with both the student classification loss and distillation loss. The distillation loss function uses Kullback-Leibler (KL) [35] the metric function is shown in equation (2). It is used to measure the distribution difference between the soft targets generated from the teacher model output p_t and the predicted probabilities of the student model p_s . The SoftMax function parameterized by the temperature T is used to obtain class probability estimates for both the teacher (soft targets) and student models. Higher temperature leads to smoother class probability distribution and this can aid in transferring the teacher's knowledge to the student model.

Let z_i be the logit output for class i . The class probability $p_i(T)$ of the image is given by equation (1). When all classes are considered, the class predictions form the vectorial input \mathbf{p}^s and \mathbf{p}^t used in the Kullback-Leibler metric function shown in equation (2).

$$p_i(T) = \frac{\exp\left(\frac{z_i}{T}\right)}{\sum_j \exp\left(\frac{z_j}{T}\right)} \quad (1)$$

$$L_{KL}(\mathbf{p}^t, \mathbf{p}^s) = \sum_j p_j^t(T) \log \left(\frac{p_j^t(T)}{p_j^s(T)} \right) \quad (2)$$

In standard knowledge distillation, the student model is trained based on the loss function that is the weighted sum of the Kullback–Leibler divergence (KL) loss and cross-entropy (CE) loss, as shown in equation (3). The training scheme is illustrated in Figure 5. Training involves two stages. In the first stage, the teacher is trained on the selected dataset. In the second stage, the teacher’s output from the training images is used as soft target. The hard target is obtained from the ground truth label y and this is jointly used with the soft target for training the student model.

$$L(\mathbf{y}, \mathbf{p}^t, \mathbf{p}^s) = \alpha T^2 L_{KL}(\mathbf{p}^t, \mathbf{p}^s) + (1 - \alpha) L_{CE}(\mathbf{y}, \mathbf{p}^s) \quad (3)$$

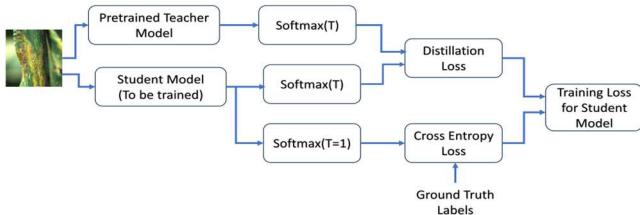


Fig. 5 Visualization of the student model training scheme based on the standard response-based knowledge distillation.

B. Dataset

To evaluate the models, we prepared the maize image dataset by combining the publicly available CDS [36] and Makerere maize plant disease dataset [37] to form a new six-class dataset named the CDS Makerere maize disease dataset. The CDS dataset contains images from three classes of disease: Gray Leaf Spot (GLS), Northern Leaf Blight (NLB), and Northern Leaf Spot (NLS). The Makerere dataset has two classes of maize disease: Maize Leaf Blight (MLB) and Maize Streak Virus (MSV), with one healthy class. Since the number of images from the Makerere dataset is much more significant than the CDS dataset, an approximately equal number of images from each class are sampled to match the class distributions in the CDS dataset. Five classes of maize plant diseases and one normal class are used in the six classes dataset. The detailed distribution of the classes is presented in Table 2. Sample images in each class are shown in Figure 6. The field-collected images are challenging due to high interclass similarities, low intraclass differences, and cluttered backgrounds.

TABLE II

DISTRIBUTION OF TRAINING AND TEST IMAGES FOR THE 6 CLASSES CDS MAKERERE DATASET

Class Name	#Training Images	#Test Images
gls	262	261
healthy	250	250
mlb	250	250
msv	250	250
nlb	249	248
nls	275	276
Total	1536	1535

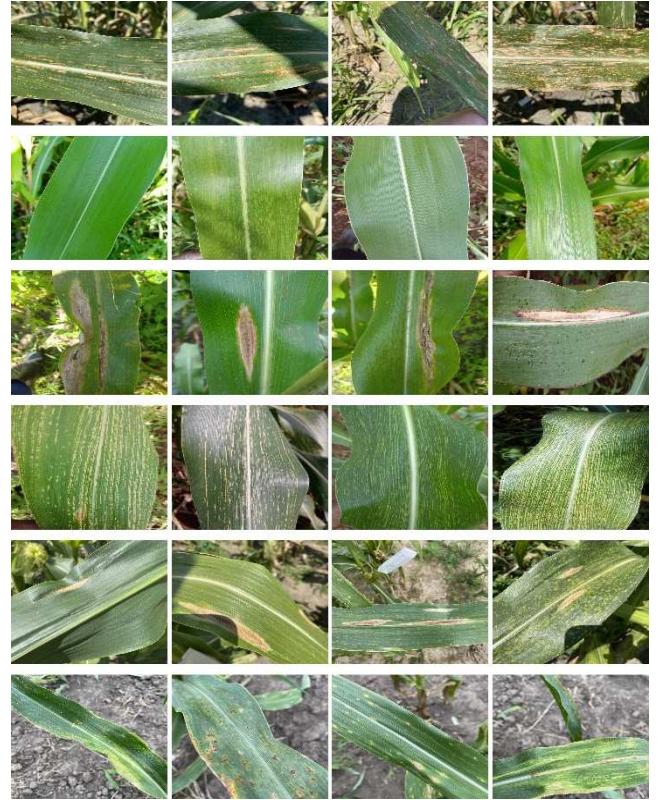


Fig. 6 Shows sample images from the training set for the 6 classes of the CDS Makerere dataset. Each row shows 3 sample images for one class. The classes from the top row are gls, healthy, MLB, MSV, nlb, and nls.

III. RESULTS AND DISCUSSION

A. Experiments Setup

The experiment aims to evaluate the performance of large and lightweight models in maize disease identification from leaf images. The DenseNet-121 and the Resnet-50 models are selected for the large model. The MobileNetV2 and SqueezeNet1.1 are the selected versions of the light models. The models are all developed with Pytorch 1.1 software. The experiment was conducted on an x86 Intel-based desktop computer with 32 GB memory and was equipped with Nvidia GPU RTX3070 8GB RAM. The images from the CDS Makerere dataset are used in this study. The training images are augmented using a transformation pipeline comprising of resizing to 224x224 pixels, random horizontal flipping, and random rotations within ± 15 degrees. All experiments are executed with the random seed 41 to ensure reproducible results.

All the models are initialized with pre-trained weights from the ImageNet dataset with the transfer learning method. The trained weight is obtained from the Pytorch Torchvision library. Training involves finetuning all the layers of the model. The AdamW optimizer is used with a learning rate of 0.0001, batch size sixty-four, and trained for 40 epochs.

B. Results

The first experiment trains the DenseNet-121, Resnet-50, MobileNet (version 2), and SqueezeNet (version 1.1) based on the CDS Makerere training images. The models are then evaluated on the test images. The classification accuracy for all the evaluated models on the CDS_Makere test set is shown in Table 3. The DenseNet-121 model shows the highest

accuracy of 99.22%. It is worth noting that the lightweight model MobileNetV2 manages to attain higher accuracy than the larger Resnet50 model.

TABLE III
CLASSIFICATION ACCURACY FOR VARIOUS CNN MODELS ON THE CDS MAKERERE MAIZE DISEASE DATASET.

CNN Model	Classification Accuracy (%)
DenseNet-121	99.22
Resnet-50	97.20
MobileNet (mobilenet_v2)	98.70
SqueezeNet (squeezeNet1_1)	96.68
SqueezeNetKD	97.13

The name in brackets shows the exact version used. squeezeNetKD is a squeeze net trained with standard knowledge distillation. In the second experiment, the best large model, DenseNet121, is used to train the SqueezeNet model with the standard knowledge distillation method [25]. DenseNet-121 is used as the teacher model, and SqueezeNet is used as the student model. The same CNN training parameters are used as in the first experiment. The aim is to evaluate whether knowledge distillation can help increase the accuracy of the SqueezeNet model.

In knowledge distillation, logit output from the teacher is utilized as a soft target to train the student model with a custom training loss function indicated in equation (3). The training loss depends on two parameters, namely temperature T and distillation weight loss, as shown in equation (3). The paper [24] recommends the use of T in the range [2.5, 4]. To achieve a smoother soft target for aiding student model learning, the temperature value of 4 ($T=4$) is used. Knowledge distillation with three values of α is compared to evaluate the effect of distillation loss weight. The result is tabulated in Table 4.

TABLE IV
IMPACT OF DISTILLATION LOSS WEIGHT α ON STUDENT MODEL CLASSIFICATION ACCURACY ON CDS MAKERERE DATASET.

Alpha	Student Accuracy (%)
0.1	95.11
0.5	96.35
0.9	97.13

DenseNet-121 serves as the teacher model, and SqueezeNet as the student model. Higher distillation loss weights improve classification performance, demonstrating the advantage of stronger model logits in training weaker models during knowledge distillation. Using the optimal loss weight enhances SqueezeNet's accuracy over standard cross-entropy training. The SqueezeNet model trained with knowledge distillation ($\alpha=0.9$) and DenseNet-121 will be referred to as the SqueezeNetKD model.

To assess the performance of SqueezeNetKD for each class, the model performance is visualized in a confusion matrix. The confusion matrix shows the number of predicted labels belonging to true and incorrect labels. This way, the model's confusion concerning its predicted class can be analyzed. The prediction accuracy for each class can be quantified as well. This is achieved by computing the class-wise precision, recall, and F1 score. The multi-class classification is framed as a separate binary classification in this setup. Thus, for each sample in the test set, the model

predicts a positive class or unfavorable class. This is repeated for each class in the test set. For each binary classification, the predicted output is labeled as either true positive, true negative, false positive, or false negative. The total true positive, true negative, false negative, and false positive labels are named TP, TN, FN, and FP, respectively. These scores are used to compute the class-wise Precision, Recall, and F1 values for the model performance metrics, as shown in equations (4), (5), and (6).

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (6)$$

Figure 7 displays the confusion matrix of SqueezeNetKD on CDS Makerere test images, showing highly accurate predictions across all classes. Analysis indicates strong Precision, Recall, and F1 score metrics for all classes (see Table 5). Figure 8 shows training and testing results for SqueezeNetKD across various epochs, demonstrating convergence to stable training and test loss values. Figure 9 illustrates SqueezeNetKD's test accuracy, which rises rapidly in the first five epochs and stabilizes after 30 epochs.

TABLE V
REPORT ON PRECISION, RECALL AND F1 SCORE FOR SQUEEZE NETKD MODEL TESTED ON THE CDS MAKERERE DATASET

Class	Precision	Recall	F1-Score
gls	0.9731	0.9693	0.9712
healthy	0.9718	0.9640	0.9679
mlb	0.9683	0.9760	0.9721
msv	0.9801	0.9840	0.9820
nlb	0.9606	0.9839	0.9721
nls	0.9741	0.9529	0.9634

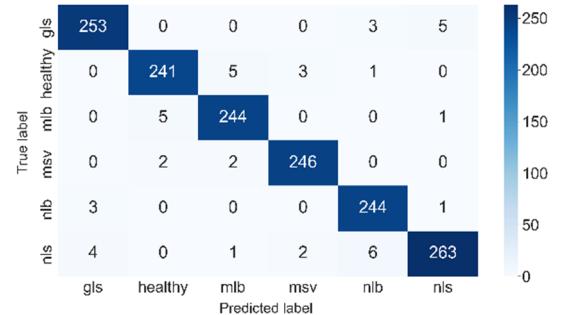


Fig. 7 Confusion matrix for SqueezeNetKD prediction on CDS Makerere dataset test set.

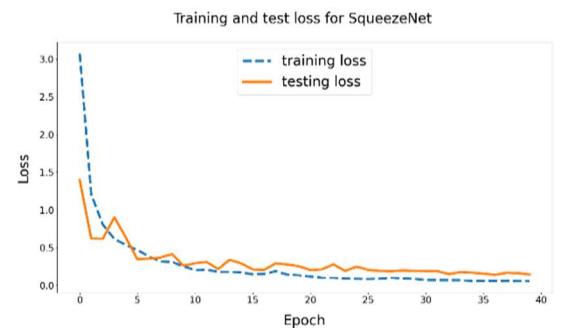


Fig. 8 Training and test loss for SqueezeNetKD model when evaluated after various training epochs. The model is evaluated with the CDS Makerere dataset

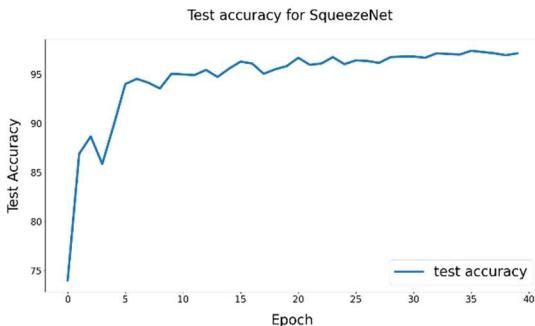


Fig. 9 Test accuracy for SqueezeNetKD model when evaluated after various training epochs. Model is tested on the CDS Makerere dataset

IV. CONCLUSION

In this paper, we evaluate the application of large and lightweight convolutional neural network models to recognize maize leaf diseases from leaf images accurately. Our experiments demonstrated the promising performance of the deep and large CNN models, DenseNet-121 and ResNet-50, in accurately classifying various maize leaf diseases. However, the deployment of these large models on resource-limited edge devices in agricultural settings is hindered by their substantial computational requirements and memory footprint. To address this challenge, we evaluated lightweight models and contrasted the performance with the larger model's. The result shows that small model can attain comparable accuracy to the large model on the proposed maize leaf disease dataset.

Knowledge distillation is used to train the SqueezeNet model to improve the performance of the small model. This enables knowledge transfer from the best-performing complex model, DenseNet-121, to the smaller SqueezeNet model. Specifically, we trained the lightweight SqueezeNet model using the standard response-based knowledge distillation method, with DenseNet-121 as the teacher model. The results revealed that knowledge distillation improves the classification accuracy of the SqueezeNet model, bridging the performance gap with the larger models. The SqueezeNet model, trained with knowledge distillation, achieved an accuracy of 97.13% on the CDS Makerere dataset, outperforming the standard SqueezeNet model trained solely on ground truth labels.

The successful application of knowledge distillation in this study highlights its potential for developing accurate and efficient lightweight models for plant disease identification in the field with resource-constrained edge devices. By leveraging the knowledge of large and complex models, lightweight models can achieve comparable performance while maintaining a smaller memory footprint and faster inference times. Although our study focused on maize leaf diseases, the knowledge distillation method can be applied to other plant species and disease categories, fostering the development of practical and affordable solutions for sustainable agriculture.

ACKNOWLEDGMENT

The authors acknowledge the research support from Multimedia University management.

REFERENCES

- [1] B. Shiferaw, B. M. Prasanna, J. Hellin, and M. Bänziger, "Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security," *Food Sec.*, vol. 3, no. 3, pp. 307–327, Sep. 2011, doi: 10.1007/s12571-011-0140-5.
- [2] R. N. Strange and P. R. Scott, "Plant disease: a threat to global food security," *Annu Rev Phytopathol*, vol. 43, pp. 83–116, 2005, doi:10.1146/annurev.phyto.43.113004.133839.
- [3] J. B. Ristaino *et al.*, "The persistent threat of emerging plant disease pandemics to global food security," *Proc Natl Acad Sci U S A*, vol. 118, no. 23, p. e2022239118, Jun. 2021, doi:10.1073/pnas.2022239118.
- [4] H. NunooMensah, S. Kuseh, J. Yankey, and F. Acheampong, "A Survey of Deep Learning Techniques for Maize Leaf Disease Detection: Trends from 2016 to 2021 and Future Perspectives," *J. Electr. Comput. Eng. Innovations*, no. Online First, Feb. 2022, doi:10.22061/jecce.2022.8602.531.
- [5] A. Ahmad, D. Saraswat, and A. El Gamal, "A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools," *Smart Agricultural Technology*, vol. 3, p. 100083, Feb. 2023, doi: 10.1016/j.atech.2022.100083.
- [6] Md. A. Haque *et al.*, "Deep learning-based approach for identification of diseases of maize crop," *Sci Rep*, vol. 12, no. 1, p. 6334, Dec. 2022, doi: 10.1038/s41598-022-10140-z.
- [7] S. Cui, Y. L. Su, K. Duan, and Y. Liu, "Maize leaf disease classification using CBAM and lightweight Autoencoder network," *J Ambient Intell Human Comput*, vol. 14, no. 6, pp. 7297–7307, Jun. 2023, doi: 10.1007/s12652-022-04438-z.
- [8] J. Hu, X. Jiang, J. Gao, and X. Yu, "LFMNet: a lightweight model for identifying leaf diseases of maize with high similarity," *Front. Plant Sci.*, vol. 15, p. 1368697, Apr. 2024, doi: 10.3389/fpls.2024.1368697.
- [9] P. Gole, P. Bedi, S. Marwaha, Md. A. Haque, and C. K. Deb, "TrIncNet: a lightweight vision transformer network for identification of plant diseases," *Front. Plant Sci.*, vol. 14, p. 1221557, Jul. 2023, doi: 10.3389/fpls.2023.1221557.
- [10] X. Qian, C. Zhang, L. Chen, and K. Li, "Deep Learning-Based Identification of Maize Leaf Diseases Is Improved by an Attention Mechanism: Self-Attention," *Front. Plant Sci.*, vol. 13, p. 864486, Apr. 2022, doi: 10.3389/fpls.2022.864486.
- [11] L. G. Divyanth, A. Ahmad, and D. Saraswat, "A two-stage deep-learning based segmentation model for crop disease quantification based on corn field imagery," *Smart Agricultural Technology*, vol. 3, p. 100108, Feb. 2023, doi: 10.1016/j.atech.2022.100108.
- [12] F. Wu, R. Gazo, B. Benes, and E. Havíarová, "Deep BarkID: a portable tree bark identification system by knowledge distillation," *Eur J Forest Res*, vol. 140, no. 6, pp. 1391–1399, Dec. 2021, doi:10.1007/s10342-021-01407-7.
- [13] Q. Huang *et al.*, "Knowledge Distillation Facilitates the Lightweight and Efficient Plant Diseases Detection Model," *Plant Phenomics*, vol. 5, p. 0062, Jan. 2023, doi: 10.34133/plantphenomics.0062.
- [14] J. Gou, B. Yu, S. J. Maybank, and D. Tao, "Knowledge Distillation: A Survey," *Int J Comput Vis*, vol. 129, no. 6, pp. 1789–1819, Jun. 2021, doi: 10.1007/s11263-021-01453-z.
- [15] J. Chen, J. Chen, D. Zhang, Y. Sun, and Y. A. Nanehkaran, "Using deep transfer learning for image-based plant disease identification," *Computers and Electronics in Agriculture*, vol. 173, p. 105393, Jun. 2020, doi: 10.1016/j.compag.2020.105393.
- [16] Z. Hao, J. Guo, K. Han, H. Hu, C. Xu, and Y. Wang, "Revisit the power of vanilla knowledge distillation: from small scale to large scale," in *Proceedings of the 37th International Conference on Neural Information Processing Systems*, in NIPS '23. Red Hook, NY, USA: Curran Associates Inc., 2024.
- [17] T. Huang, S. You, F. Wang, C. Qian, and C. Xu, "Knowledge distillation from a stronger teacher," *Advances in Neural Information Processing Systems*, vol. 35, pp. 33716–33727, 2022.
- [18] H. Chen, Y. Wang, C. Xu, C. Xu, and D. Tao, "Learning Student Networks via Feature Embedding," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 32, no. 1, pp. 25–35, Jan. 2021, doi:10.1109/TNNLS.2020.2970494.
- [19] A. Ghofrani and R. Mahdian Toroghi, "Knowledge distillation in plant disease recognition," *Neural Comput & Applic*, vol. 34, no. 17, pp. 14287–14296, Sep. 2022, doi: 10.1007/s00521-021-06882-y.
- [20] S.-V. Echim, I.-M. Tăiatu, D.-C. Cercel, and F. Pop, "Explainability-Driven Leaf Disease Classification Using Adversarial Training and Knowledge Distillation;," in *Proceedings of the 16th International Conference on Agents and Artificial Intelligence*, Rome, Italy:

SCITEPRESS - Science and Technology Publications, 2024, pp. 782–791. doi: 10.5220/0012392900003636.

[21] Y. Hu, G. Liu, Z. Chen, J. Liu, and J. Guo, “Lightweight One-Stage Maize Leaf Disease Detection Model with Knowledge Distillation,” *Agriculture*, vol. 13, no. 9, p. 1664, Aug. 2023, doi:10.3390/agriculture13091664.

[22] C. Shu, Y. Liu, J. Gao, Z. Yan, and C. Shen, “Channel-wise Knowledge Distillation for Dense Prediction,” in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada: IEEE, Oct. 2021, pp. 5291–5300. doi:10.1109/ICCV48922.2021.00526.

[23] H. Zhang and M. Wang, “MixKd: Mix Data Augmentation Guided Knowledge Distillation for Plant Leaf Disease Recognition,” in *Green, Pervasive, and Cloud Computing*, C. Yu, J. Zhou, X. Song, and Z. Lu, Eds., Cham: Springer International Publishing, 2023, pp. 169–177.

[24] G. Hinton, O. Vinyals, and J. Dean, “Distilling the Knowledge in a Neural Network,” Mar. 09, 2015, *arXiv: arXiv:1503.02531*. Accessed: Apr. 10, 2024. [Online]. Available: <http://arxiv.org/abs/1503.02531>

[25] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI: IEEE, Jul. 2017, pp. 2261–2269. doi:10.1109/CVPR.2017.243.

[26] 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)*, Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[27] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT: IEEE, Jun. 2018, pp. 4510–4520. doi:10.1109/CVPR.2018.00474.

[28] F. Iandola and K. Keutzer, “Small neural nets are beautiful: enabling embedded systems with small deep-neural-network architectures,” in *Proceedings of the Twelfth IEEE/ACM/IFIP International Conference on Hardware/Software Codesign and System Synthesis Companion*, Seoul Republic of Korea: ACM, Oct. 2017, pp. 1–10. doi:10.1145/3125502.3125606.

[29] S. Chang *et al.*, “Recognition of wheat rusts in a field environment based on improved DenseNet,” *Biosystems Engineering*, vol. 238, pp. 10–21, Feb. 2024, doi: 10.1016/j.biosystemseng.2023.12.016.

[30] W. Bao *et al.*, “An improved DenseNet model to classify the damage caused by cotton aphid,” *Computers and Electronics in Agriculture*, vol. 203, p. 107485, Dec. 2022, doi: 10.1016/j.compag.2022.107485.

[31] S. M. Hassan and A. K. Maji, “Pest Identification Based on Fusion of Self-Attention With ResNet,” *IEEE Access*, vol. 12, pp. 6036–6050, 2024, doi: 10.1109/ACCESS.2024.3351003.

[32] J. Chen, D. Zhang, M. Suzauddola, and A. Zeb, “Identifying crop diseases using attention embedded MobileNet-V2 model,” *Applied Soft Computing*, vol. 113, p. 107901, Dec. 2021, doi:10.1016/j.asoc.2021.107901.

[33] S. Phiphatphaisit and O. Surinta, “Food Image Classification with Improved MobileNet Architecture and Data Augmentation,” in *Proceedings of the 2020 The 3rd International Conference on Information Science and System*, Cambridge United Kingdom: ACM, Mar. 2020, pp. 51–56. doi: 10.1145/3388176.3388179.

[34] R. Gomes *et al.*, “A Comprehensive Review of Machine Learning Used to Combat COVID-19,” *Diagnostics*, vol. 12, no. 8, p. 1853, Jul. 2022, doi: 10.3390/diagnostics12081853.

[35] T. Kim, J. Oh, N. Y. Kim, S. Cho, and S.-Y. Yun, “Comparing Kullback-Leibler Divergence and Mean Squared Error Loss in Knowledge Distillation,” in *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, Montreal, Canada: International Joint Conferences on Artificial Intelligence Organization, Aug. 2021, pp. 2628–2635. doi:10.24963/ijcai.2021/362.

[36] A. Ahmad, A. E. Gamal, and D. Saraswat, “Toward Generalization of Deep Learning-Based Plant Disease Identification Under Controlled and Field Conditions,” *IEEE Access*, vol. 11, pp. 9042–9057, 2023, doi: 10.1109/ACCESS.2023.3240100.

[37] M. Banadda, N. K. Aloysius, S. Nakazzi, O. B. Ernest, M. S. Owekwiibwa, and G. Marvin, “Explainable Artificial Intelligence for Maize Disease Diagnostics,” in *2023 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, Nadi, Fiji: IEEE, Dec. 2023, pp. 1–6. doi: 10.1109/CSDE59766.2023.10487660.