Mobile-PDC: High-Accuracy Plant Disease Classification for Mobile Devices

Samiiha Nalwooga[1,2], Luis Quintero[1], Sindri Magnússon[1] and Jeremy Francis Tusubira[2]

[1] Department of Computer and Systems Sciences, Stockholm University

[2] College of Computing and Information Sciences, Makerere University

Abstract

Cassava is a staple crop in Africa, but has a high sensitivity to diseases. Traditional disease diagnosis is cumbersome, requiring experts to visually assess crop health. Current deep transfer learning models can help automate disease diagnosis, but they are not compatible with mobile devices due to limited memory and computational capacity. To address this issue, knowledge distillation is proposed as a technique to build accurate plant disease classification models compatible with mobile devices. The Mobile-PDC (Plant Decision Classification) models have the MobileNet structure, making them compatible with multiple mobile devices. Experiments show that these models can compress 91.2% of the original PDC models without losing accuracy.

Dataset

We use a dataset from Kaggle. The data set consists of 21,367 labelled images collected during a crowdsourcing surveillance survey in Uganda by the Makerere University Artificial Intelligence Research Lab. It has 5 attributes which are the different cassava disease classes plus the healthy class. The different classes include **Cassava Mosaic Disease (CMD), Cassava Brown Streak Disease (CBSD), Cassava Bacterial Blight (CBB), Cassava Green Mite (CGM)** and the **healthy (disease free)** class.

CMD **CBSD** **CBB** **CGM**

Knowledge Distillation

The idea of Knowledge Distillation (KD) is that we have a high-accuracy "teacher" model with large feature space that teaches a smaller "student" model high-accuracy prediction. When a trained teacher (denoted as t) is given an input image X, it produces a vector of scores:

$$S^t(X) = \left[s_1^t(X), s_2^t(X), \ldots\ldots, s_i^t(X)\right]$$

that the Softmax layer transforms into probabilities:

$$P_k^t(X) = \frac{\exp\left(S_i^t(X)\right)}{\sum_j \exp\left(S_j^t(X)\right)}$$

Hinton et al suggested that the small probabilities could be softened by using a hyperparameter for temperature scaling T > 1

$$\tilde{P}_k^t(X) = \frac{\exp\left(S_i^t(X)/T\right)}{\sum_j \exp\left(S_j^t(X)/T\right)}$$

On the other hand, the student (denoted as s) model also produces a softened probability distribution P^s_k (X). The student loss is a summation of the normal cross entropy loss L_{ls} and the distillation loss L_{KD} weighted by a factor α.

$$L = \alpha L_{ls} + (1-\alpha)L_{KD} \text{ where } L_{KD} = -T^2 \sum_k \tilde{P_k^t}(x) \log P_k^{\tilde{s}}(x)$$

α is a weight and T is the temperature that is used to soften the logits.

Methodology

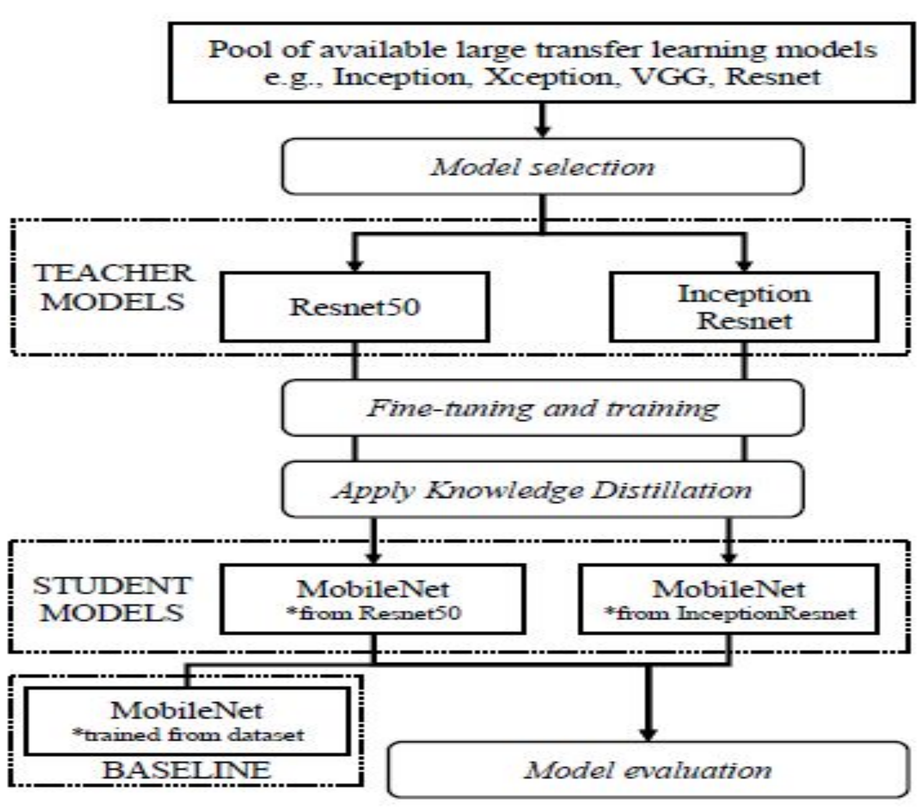

Figure 1: Steps for finding the Mobile-PDC models

Results

Table 1: Comparison of the accuracies and the Cross Entropy losses between mobile-PDC models **D1** and **D2** distilled at T=5, their respective teacher models (T_M) and baseline model trained from scratch (S_M).

T_M	D_M	CE(T_M)	CE (D_M)	CE(S_M)	Accuracy (T_M)	Accuracy (D_M)	Accuracy (S_M)
ResNet 50	**D1**	0.3399	**0.1800**	0.5229	0.8900	**0.9051**	0.7172
Inception ResNet V2	**D2**	0.2961	**0.1234**		0.9120	**0.9125**	

Table 2: The compression rate of the PDC models. T_M is the teacher model and D_M is the mobile-PDC model.

T_M	D_M	D_M parameters	$T_M$params	$CR_{TM/DM}$	Parameter savings	$CR_{DM/TM}$
ResNet 50	D1	4,805,829	23,770,205	**4.95**	79.78	**20.22**
Inception ResNet V2	D2		54,344,421	**11.31**	91.16	**8.84**

Conclusion

We have presented Mobile-PDC models as potential models that can be used by smallholder farmers to perform in-field disease diagnosis. These models have been built by transferring knowledge from the large state-of-the-art models such as ResNet 50 and Inception ResNet V2 to a lightweight model MobileNet using Knowledge Distillation.

Results from our experiments show that these models are able to achieve the classification performance of these large models and yet they are only a proportion of their size.

Furthermore, these models have been built using MobileNet which is a lightweight model that was designed for mobile platforms making them compatible with mobile device capabilities. Since these model have the MobileNet architecture, they can further be optimized through post training quantization (Jacob et al. 2018) to make them much faster during the inference hence better to run on mobile devices.

Future work

- ➢ Explore the use of Generative Adversarial Networks using this approach.
- ➢ Evaluate the performance of the distilled models using the ROC_AUC.

References

[1] Hinton, G.; Vinyals, O.; and Dean, J. 2015. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531.*

[2] Yang, Z.; Cui, Y.; Chen, Z.; Che, W.; Liu, T.; Wang, S.; and Hu, G. 2020. Textbrewer: An open-source knowledge distillation toolkit for natural language processing. *arXiv preprint arXiv:2002.12620.*

[3] Ai, Y.; Sun, C.; Tie, J.; and Cai, X. 2020. Research on recognition model of crop diseases and insect pests based on deep learning in harsh environments. I*EEE Access 8:171686– 171693.*

Acknowledgement

Special thanks to the Swedish International Development cooperation Agency (SIDA) and College of Computing and Information Sciences, Makerere University for the financial support.
Special thanks to my supervisors from the Department of Computer and Systems Sciences at Stockholm University.
Utmost gratitude to the Deep Learning Indaba and GIZ for the financial support to attend the Deep Learning Indaba 2023 conference.