Leaf Disease Detection using knowledge Distillation

Shilajit Banerjee
M22CS062
m22cs062@iitj.ac.in

Dattatreyo Roy
M22CS060
m22cs060@iitj.ac.in

Devansh Kaushik
M22CS005
m22cs005@iitj.ac.in

Abstract

The main objective of this project is developing a deep learning model for leaf disease detection that can be trained faster with higher accuracy using transfer learning where we will involve teacher-student architecture. The teacher model will be trained on a large dataset and its knowledge will be transferred to smaller student model to improve the accuracy. Our aim is to make an ensemble of student model which has learned from the teacher model.

1. Introduction

The detection of plant diseases has become an important task for crop protection and management. Traditional methods of plant disease detection are expensive, time-consuming and error-prone. However, deep learning models is proving to be effective in the detection of leaf diseases in plants. In this project, we propose to implement the concept of knowledge distillation in deep learning to improve the accuracy of leaf disease detection.

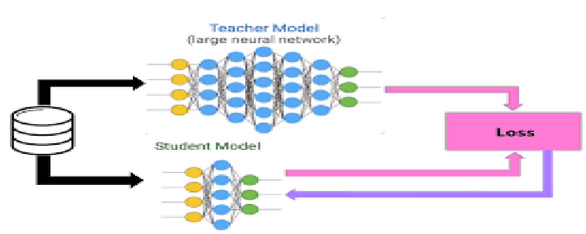

Figure 1. knowledge distillation

2. Literature Review

The paper "Plant disease recognition using knowledge distillation" by Ali Ghofrani and Rahil Mahdian Toroghi proposes a novel approach to improve the efficiency and accuracy of plant disease recognition models. Ghofrani and Toroghi use a client-server system with a large deep CNN model on the server and a small model on the client for plant disease recognition. They apply knowledge distillation to transfer knowledge from the large model to the small model, resulting in improved accuracy on the MobileNet architecture with limited processing power. This teacher-student approach improves the classification rate of the small model. In the paper "Distilled-MobileNet Model of Convolutional Neural Network Simplified Structure for Plant Disease Recognition" by QIU Wenjie, YE Jin, HU Liangqing, YANG Juan2, LI Qili, MO Jianyou, YI Wanmao, they propose a structured model compression method using knowledge distillation to transfer knowledge from a complex teacher model to a lightweight student model, which they apply to the classification of 38 common plant diseases across 14 crops. The resulting Distilled-MobileNet model achieves high accuracy while significantly reducing model size and recognition time, making it a promising approach for disease recognition on devices with limited computing resources.

3. Major Objectives of the project

The main objective of our project is:

1. Visualise the effect of transfer Learning and knowledge distillation.
2. Understanding how a student network learns from the teacher model which can be taken as a powerful model.
3. Give the student model very less data so that it use the knowledge of the teacher model.
4. We will train the student model with less no of parameters.
5. Making an ensemble of student networks which can learn from different teacher model and finding the application in the knowledge Distillation by transfer learning.

4. Dataset Collection

Dataset to be used:
https://www.kaggle.com/datasets/csafrit2/plant-leaves-for-image-classification

The plant leaves dataset on Kaggle consists of 4,503 images of 12 plant species, namely Mango, Arjun, Alstonia Scholaris, Guava, Bael, Jamun, Jatropha, Pongamia Pinnata, Basil, Pomegranate, Lemon, and Chinar. The images in this subset have been labeled into two classes - healthy and diseased. The plants were named from P0 to P11 and the dataset was divided into 22 subject categories ranging from 0000 to 0022. The images labeled with categories 0000 to 0011 represent the healthy class, while those labeled with categories 0012 to 0022 represent the diseased class. The dataset contains 2,278 images of healthy leaves and 2,225 images of diseased leaves. The images were captured using a smartphone camera and collected from various sources including online databases, personal collections, and botanical gardens.

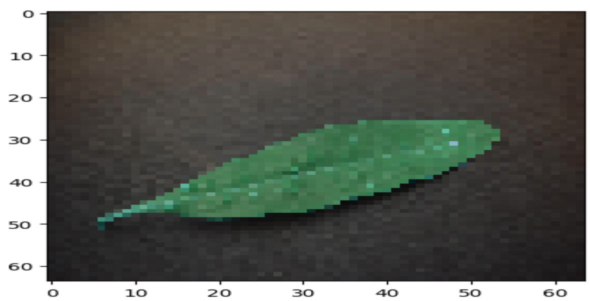

Figure 2. leaf sample.

5. Approach

The experiment aims to boost the performance of two student models using knowledge distillation, which involves transferring knowledge from a more complex teacher model to simpler student models. Soft targets generated by the teacher model are used to train the student models in addition to the hard targets of the dataset. This allows the student models to learn richer information and potentially improve their performance.

We will use DenseNet121 and Resnet101, two popular deep convolutional neural networks, as the teacher model in our leaf image classification experiment. The teacher model will be trained on a dataset for a large number of epochs to achieve high accuracy. We will use knowledge distillation to fine-tune the pre-trained model on this dataset.

Our objective is to use transfer learning and knowledge distillation to transfer the knowledge learned by the teacher model to the two selected student models, ShuffleNet and a custom-made CNN model consisting of 4 convolutional layers, which previously had poor performance. The student models will be trained on a smaller subset of the dataset with fewer epochs than the teacher model. By transferring knowledge, the student models can achieve higher accuracy in a shorter amount of time.

After training the ShuffleNet and custom-made CNN student models using knowledge distillation from the DenseNet121 and ResNet101 teacher models, we have four student models in total, which demonstrated improved performance compared to training them individually. Now, we want to do something which can increase the performance of these students and made them better.

To further enhance the performance of our student models, we can create an ensemble model by combining the predictions of the four student models. Ensemble learning has been shown to improve accuracy and robustness by leveraging the strengths of each individual model. The ensemble model can take the average prediction or use a more sophisticated technique like weighted averaging. In our experiment, the ensemble model made up of the four student models outperformed each individual student model that learned from their teacher model, proving the effectiveness of ensemble learning.

Class imbalance is a common problem in machine learning, where the number of examples in each class is not evenly distributed. Class imbalance can negatively affect neural network training by causing overfitting to majority classes and underfitting to minority classes. Synthetic Minority Over-sampling Technique (SMOTE) is a data augmentation technique that generates synthetic samples of the minority class by interpolating between existing samples. SMOTE helps balance class distributions in datasets, which can improve the neural network's ability to classify minority classes accurately.

In this experiment, SMOTE was applied to the imbalanced dataset to improve neural network performance during training and subsequently enhance the quality of the knowledge distillation process. The imbalanced-learn library in Python was used to implement SMOTE. After applying SMOTE, the resulting dataset had a more balanced class distribution, which was used to train the neural networks and perform knowledge distillation. The application of SMOTE led to improved results, which were recorded and analyzed.

6. Expected Results

As we know that DenseNet121 and Resnet101 models are well known and gives better results. Having applying knowledge distillation, the performance of our student models made significant improvements as compared to the normal student model with it. Furthurmore, the ensemble of the students gave better result.

Then, we have applied smoting and have obtained better results. There was a increase of around 6-8 percent with smoting as compared to when using the imbalanced dataset.

7. Conclusion

In this experiment, we used knowledge distillation to improve the performance of two student models, namely ShuffleNet and a custom-made CNN, by transferring knowl-

Without SMOTING:

	Without Knowledge Distillation				**With Knowledge Distillation (Teacher is DenseNET121)**			
Model	Accuracy	Precision	Recall	F1 Score	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Student Model 1 (DNET)	0.53	0.56	0.46	0.46	**0.58**	**0.60**	**0.58**	**0.56**
Shuffle Net (Student Model 2)	0.45	0.33	0.45	0.37	**0.63**	**0.63**	**0.63**	**0.62**

	Without Knowledge Distillation				**With Knowledge Distillation (Teacher is ResNet101)**			
Model	Accuracy	Precision	Recall	F1 Score	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Student Model 4 (RNET)	0.53	0.56	0.53	0.50	**0.59**	**0.62**	**0.59**	**0.57**
Shuffle Net (Student Model 3)	0.41	0.30	0.42	0.34	**0.63**	**0.61**	**0.63**	**0.61**

Figure 3. knowledge distillation without smoting

Model	Accuracy	Precision	Recall	F1 Score
DenseNet 121	0.75	0.77	0.75	0.69
Resnet 101	0.77	0.81	0.77	0.77
Ensemble Model	0.67	0.65	0.67	0.65

Table 1. Performance metrics for the models without SMOTING

Model	Accuracy	Precision	Recall	F1 Score
DenseNet 121	0.77	0.79	0.77	0.78
Resnet 101	0.74	0.78	0.73	0.73
Ensemble Model	0.74	0.75	0.74	0.74

Table 2. Performance metrics for the models with SMOTING

edge from pre-trained teacher models, DenseNet121 and ResNet101. After the knowledge transfer, we obtained four student models in total, and an ensemble of these models was created to further improve the performance. We found that the ensemble model outperformed each of the individual student models.

With SMOTING:

	Without Knowledge Distillation				**With Knowledge Distillation (Teacher is DenseNET121)**			
Model	Accuracy	Precision	Recall	F1 Score	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Student Model 1 (DNET)	0.53	0.57	0.51	0.51	**0.68**	**0.68**	**0.67**	**0.68**
Shuffle Net (Student Model 2)	0.58	0.64	0.58	0.59	**0.64**	**0.69**	**0.64**	**0.66**

	Without Knowledge Distillation				**With Knowledge Distillation (Teacher is ResNet101)**			
Model	Accuracy	Precision	Recall	F1 Score	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Student Model 4 (RNET)	0.54	0.63	0.55	0.55	**0.61**	**0.63**	**0.59**	**0.57**
Shuffle Net (Student Model 3)	0.53	0.61	0.52	0.52	**0.62**	**0.67**	**0.62**	**0.63**

Figure 4. knowledge distillation with smoting

We also addressed the issue of class imbalance by applying Synthetic Minority Over-sampling Technique (SMOTE) to the dataset. This helped to balance the distribution of the classes in the dataset and improve the robustness of the neural networks.

Overall, our experiment shows that knowledge distillation and ensemble learning can be effective techniques for improving the performance of neural networks. Additionally, addressing the issue of class imbalance can further enhance the performance of the models.

8. Reference

1. Hinton, G., Vinyals, O., Dean, J. (2015). Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531.
2. Ghofrani, A., Toroghi, R. M. (2020). Plant disease recognition using knowledge distillation. Computers and Electronics in Agriculture, 178, 105742.
3. Qiu, W., Ye, J., Hu, L., Yang, J., Li, Q., Mo, J., ... Yi, W. (2020). Distilled-MobileNet model of convolutional neural network simplified structure for plant disease recog-

nition. PloS one, 15(10), e0240602.

4. Yu, Y., Chen, H., Dou, Q. (2020). Discriminative network compression for plant disease recognition. Computers and Electronics in Agriculture, 174, 105510.

.

RESULTS:

Table 3. Comparison of model performance with and without knowledge distillation (teacher is DenseNET121) without smoting

	Without Knowledge Distillation				With Knowledge Distillation			
Model	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score
Student Model 1 (DNET)	0.53	0.56	0.46	0.46	0.58	0.60	0.58	0.56
Shuffle Net (Student Model 2)	0.45	0.33	0.45	0.37	0.63	0.63	0.63	0.62

Table 4. Comparison of model performance with and without knowledge distillation (teacher is ResNET101) without smoting

	Without Knowledge Distillation				With Knowledge Distillation			
Model	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score
Student Model 4 (RNET)	0.53	0.56	0.53	0.50	0.59	0.62	0.59	0.57
Shuffle Net (Student Model 3)	0.41	0.30	0.42	0.34	0.63	0.61	0.63	0.61

Table 5. Comparison of model performance with and without knowledge distillation (teacher is DenseNET121) with smoting

	Without Knowledge Distillation				With Knowledge Distillation			
Model	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score
Student Model 1 (DNET)	0.53	0.57	0.51	0.51	0.68	0.68	0.67	0.68
Shuffle Net (Student Model 2)	0.58	0.64	0.58	0.59	0.64	0.69	0.64	0.66

Table 6. Comparison of model performance with and without knowledge distillation (teacher is ResNET101) with smoting

	Without Knowledge Distillation				With Knowledge Distillation			
Model	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score
Student Model 4 (RNET)	0.54	0.63	0.55	0.55	0.61	0.63	0.59	0.57
Shuffle Net (Student Model 3)	0.53	0.61	0.52	0.52	0.62	0.67	0.62	0.63

.

9. Appendix

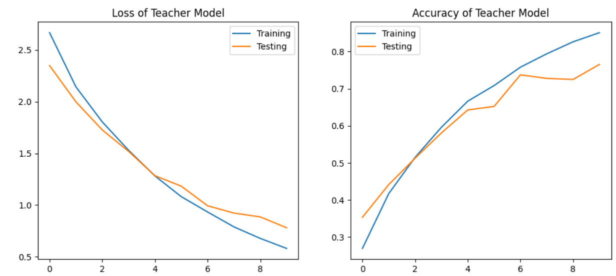

Figure 5. Comparison of knowledge distillation performance without smoting using densenet121.

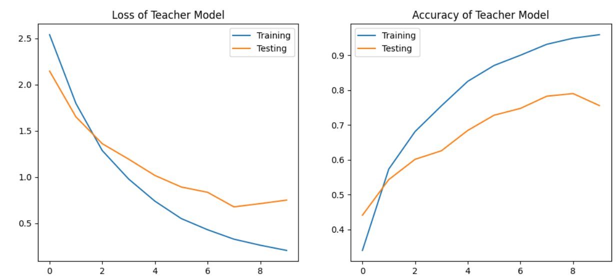

Figure 6. Comparison of knowledge distillation performance with smoting using densenet121.

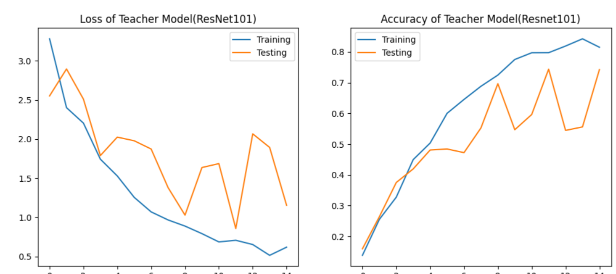

Figure 7. Comparison of knowledge distillation performance without smoting using resnet101

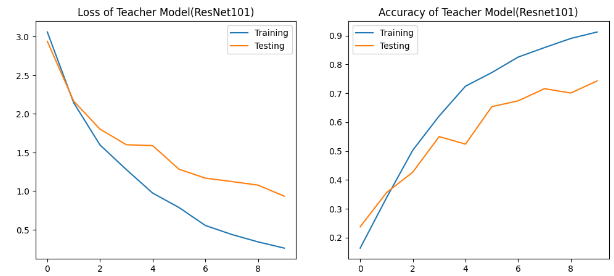

Figure 8. Comparison of knowledge distillation performance with smoting using resnet101.

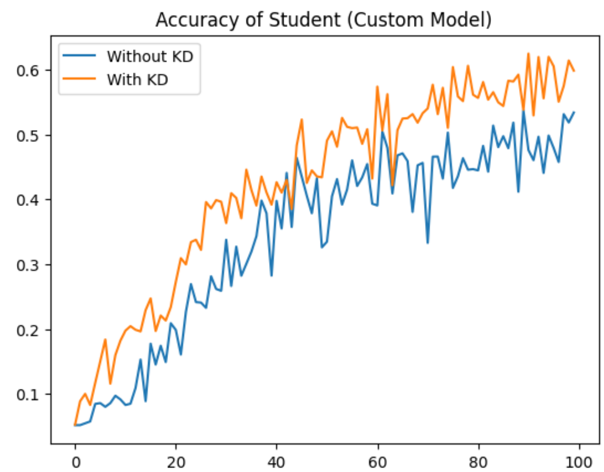

Figure 9. Comparison of knowledge distillation performance for Student1 with and without using teacher network guidance.

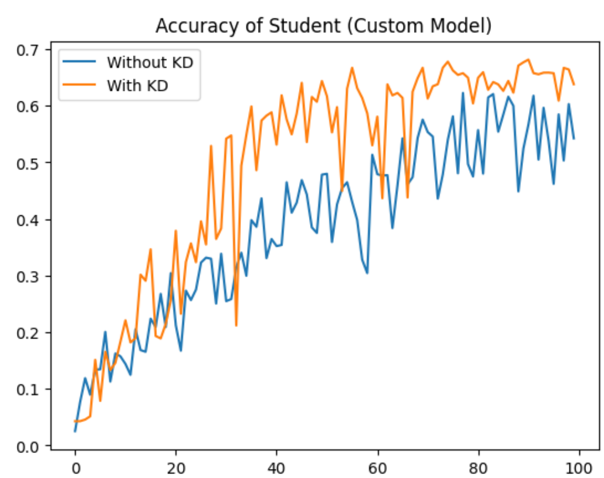

Figure 10. Comparison of knowledge distillation performance for Student4 with and without using teacher network guidance.