

FINAL REPORT
DESIGN CREDITS PROJECT
KNOWLEDGE DISTILLATION FOR RADIOLOGY DATASET

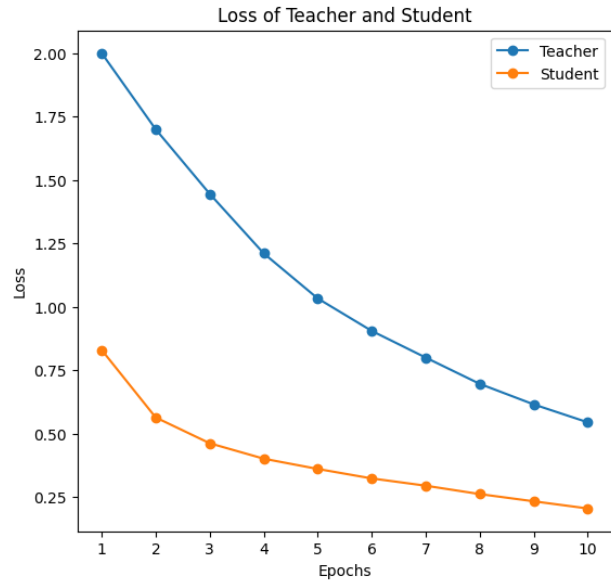
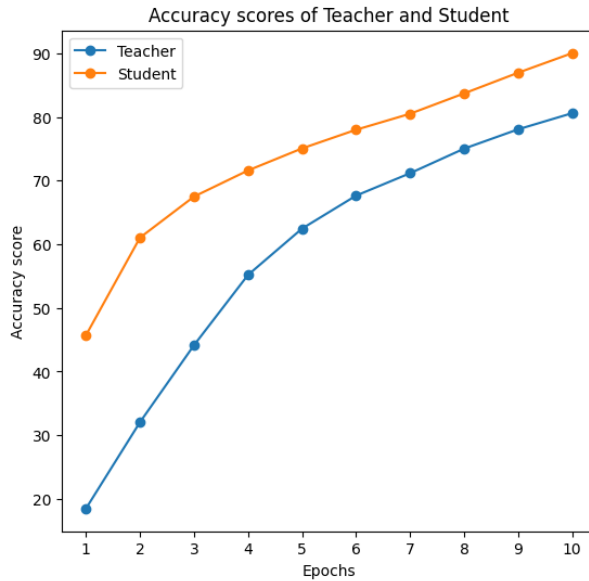
Following report contains all the results right from starting of the project

IMPLEMENTATION OF STUDENT TEACHER MODEL:

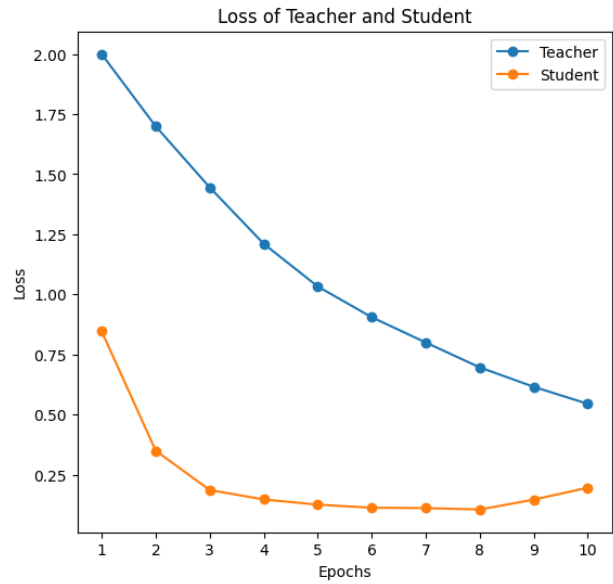
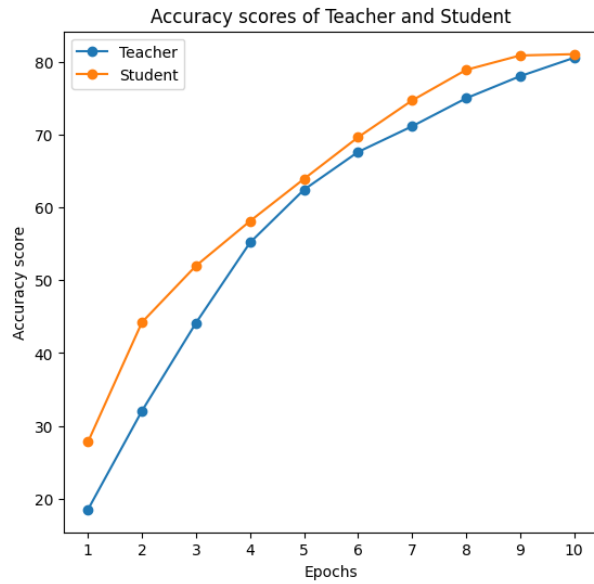
On CIFAR-10 dataset:

- Model Details:
Teacher Network Layers = 10 conv layers and 1 pooling layer
Student Network = 3 conv layers + 1 pool layer
Fully-Connected layers (in both) = 1 FC layer with 256 nodes.
- Accuracy on testing set: 73.42%
- Comparison between Teacher and Student:

For training set



For validation set:

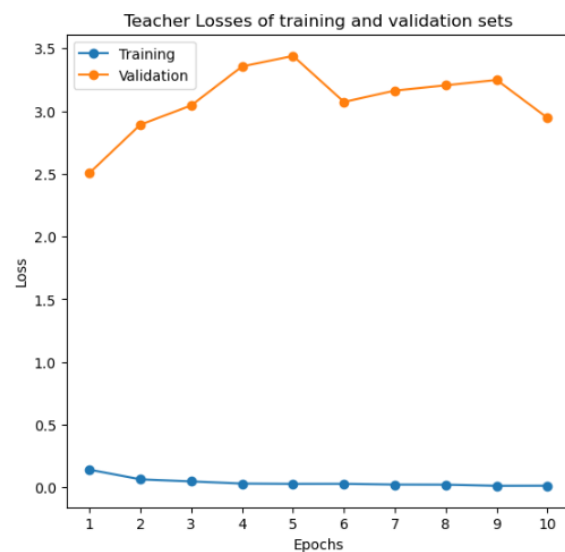
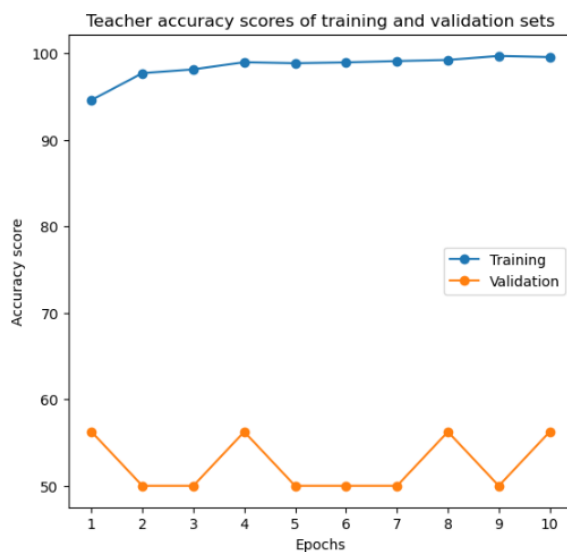


On Pneumonia Chest X-Rays dataset:

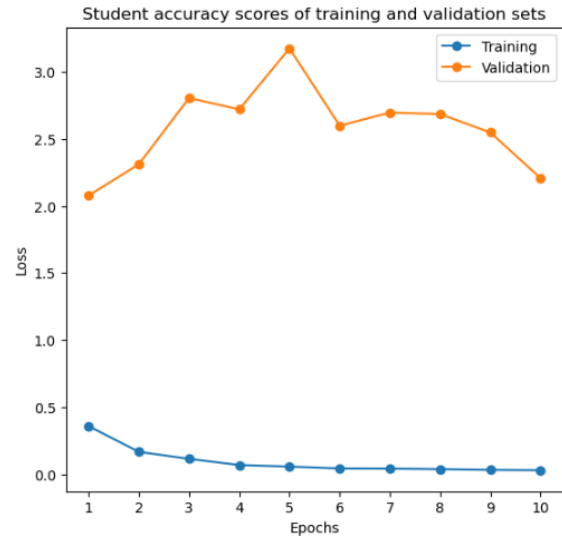
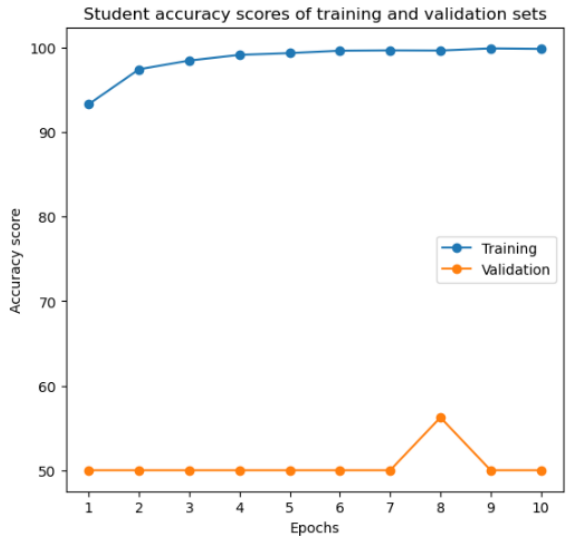
1) Model 1 Details:

Directly using the Densent121 and MobilenetV2 models for teacher and student with pre trained as True and one extra Fully connected layer

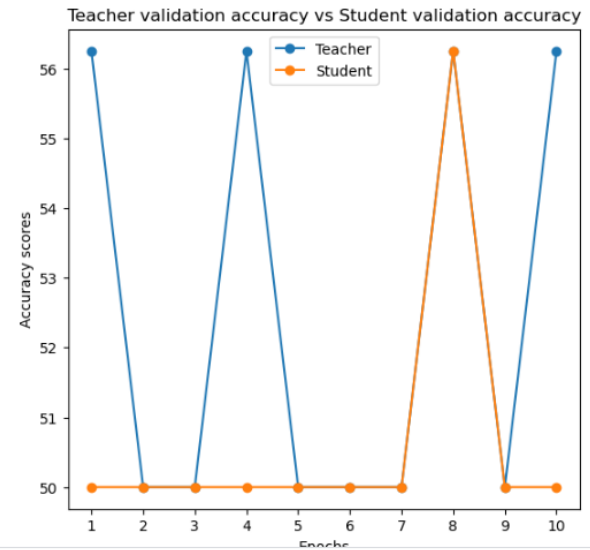
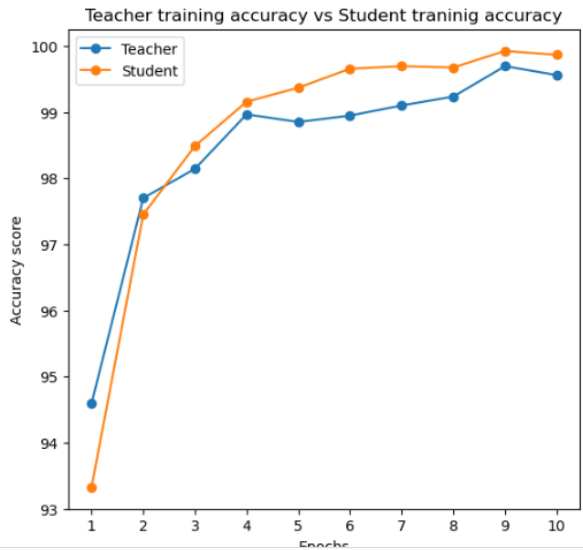
- Teacher accuracies and losses for testing and validation sets.



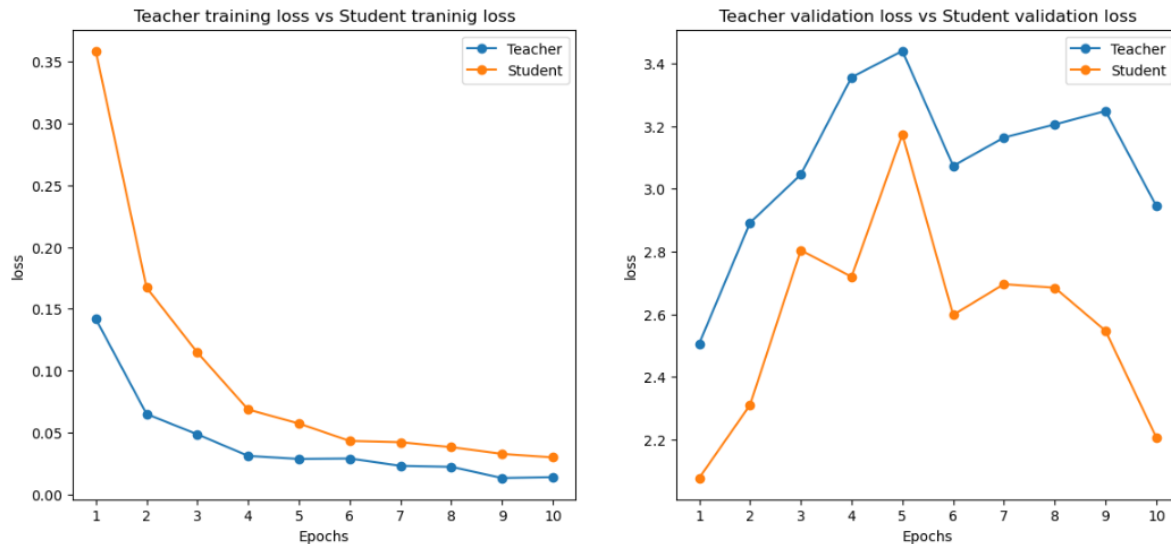
- Student accuracies and losses for testing and validation sets.



- Comparison between Teacher and Student models (accuracies):



- Comparison between Teacher and Student models (losses):



- **Conclusions:**
 - Because the models are pretrained, the layers are already frozen and hence could not learn properly
 - Yet, the performance of Student is better in terms of accuracies and losses.
 - More layers must be added and pretrained must be made FALSE so that the models learn the dataset.

2) **Model 2 Details:**

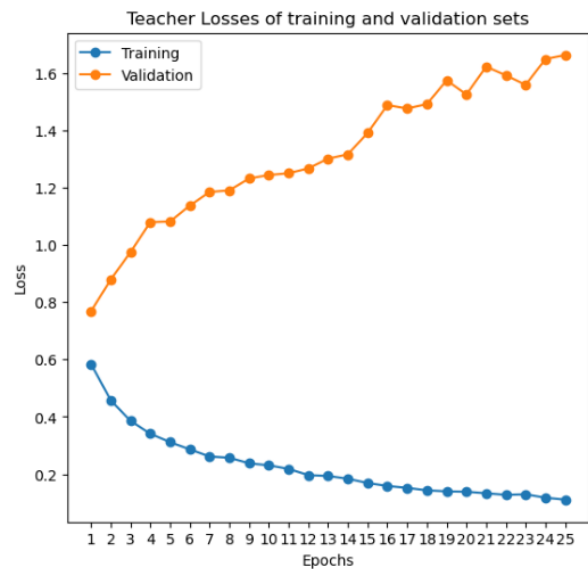
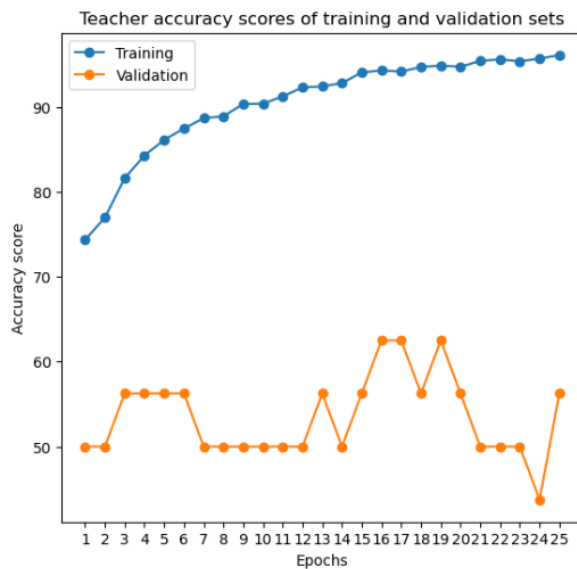
The layers in the TeacherModel are:

- DenseNet121: pre-trained convolutional layers with pretrained = False
- Dropout: with probability 0.5 to avoid overfitting
- Conv2d: with 512 filters of size 3x3 and padding of 1
- ReLU: Rectified Linear Unit activation function
- MaxPool2d: with kernel size of 7x7 and stride of 1
- Linear: fully connected layer with 512 input features and 64 output features
- Linear: fully connected layer with 64 input features and 2 output features

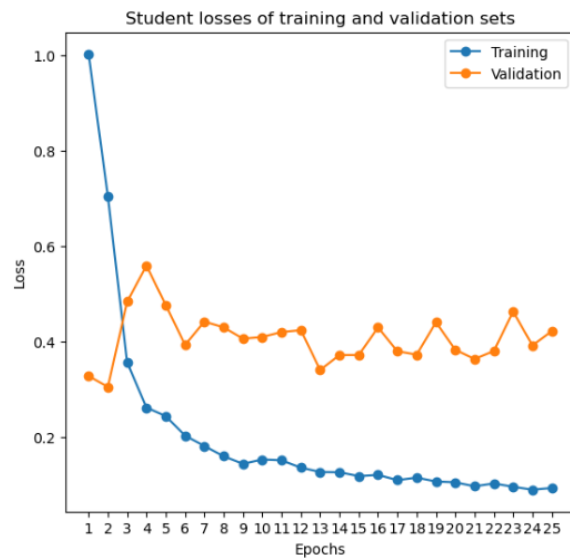
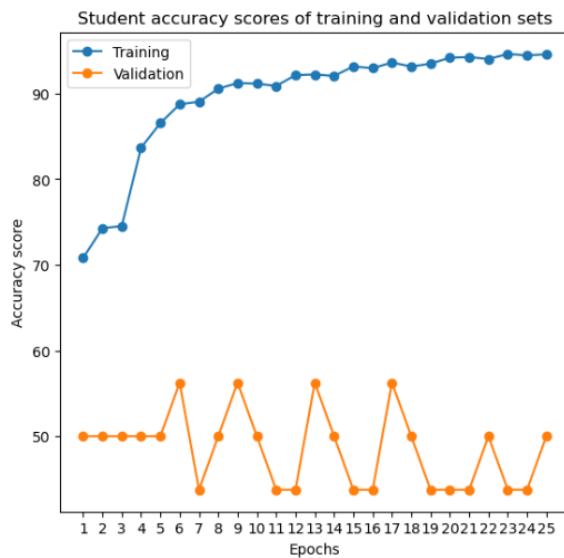
The layers in the StudentModel are:

- mobilenet: A MobileNetV2 feature extractor with no weights pretrained.
- adaptive_pool: An adaptive average pooling layer that reduces the spatial dimensions of the features to a size of (1, 1).
- fc1: A fully connected layer with 1280 input features and 512 output features.
- fc2: A fully connected layer with 512 input features and 2 output features (assuming it is a binary classification problem).
- ReLU activation function is applied after fc1 layer.

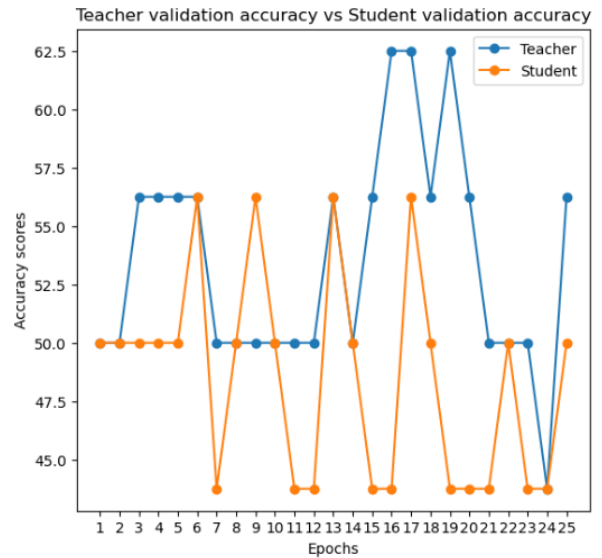
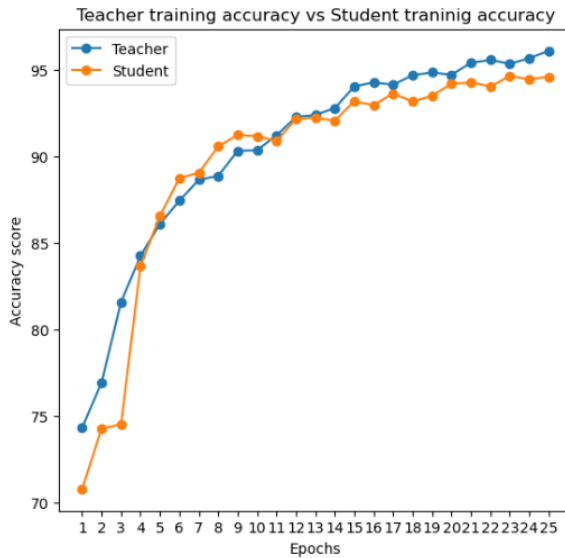
- SGD optimizer used with learning rate 0.0001 and momentum = 0.8
- The model was run on 25 epochs
- Teacher accuracies and losses for training and validation datasets:



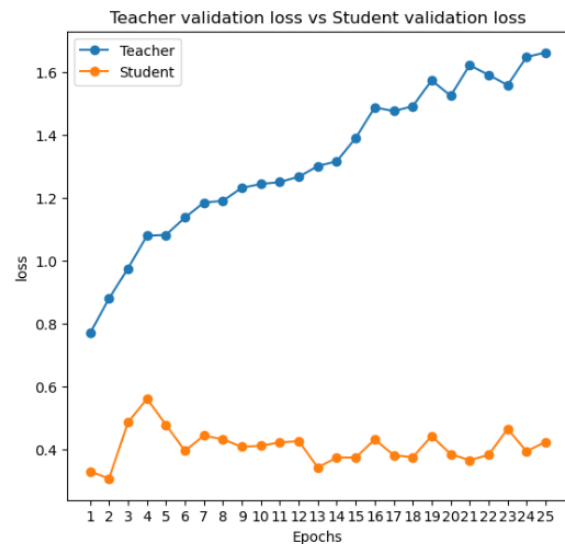
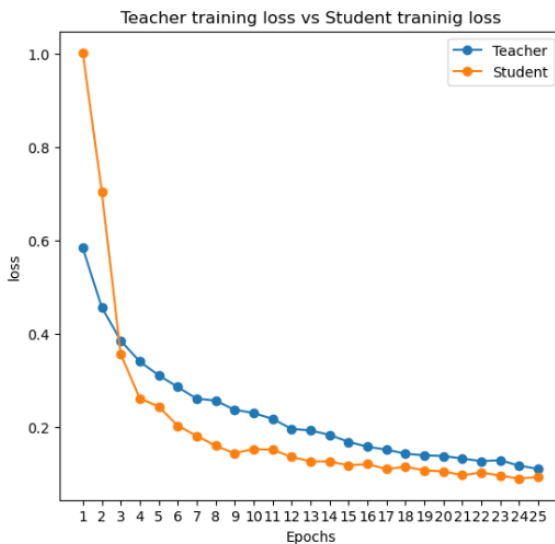
- Student accuracies and losses for training and validation datasets:



- Comparison between Teacher and Student for accuracies



- Comparison between Teacher and Student for losses:

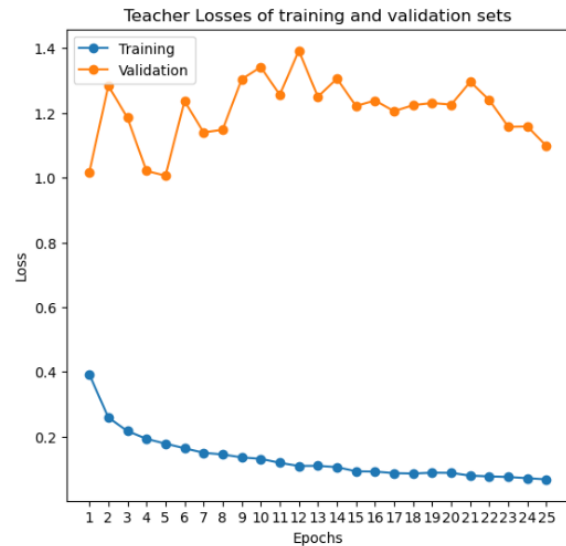
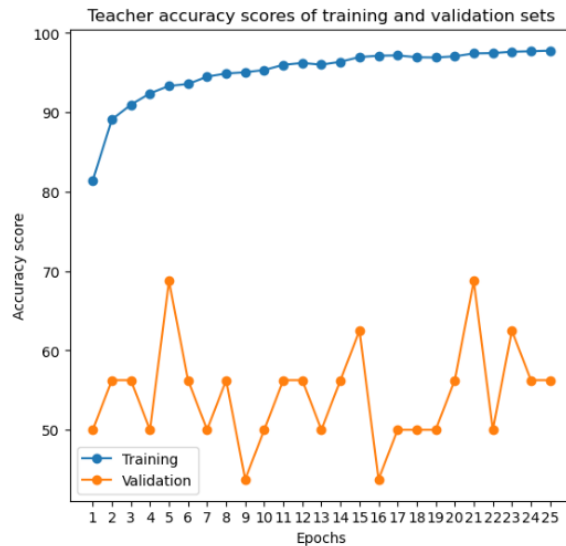


- Conclusions:
 - Even though the Student model is performing better than the teacher model in terms of accuracy and loss (that is accuracies of student model is more than teacher model and the losses of student model is less than that of teacher model), the models are overfitting over the dataset (that is validation losses are more than training losses for both models)
 - So, model needs to be optimized
 - Lr regularization must be done, dropout layer, weight decay must be added

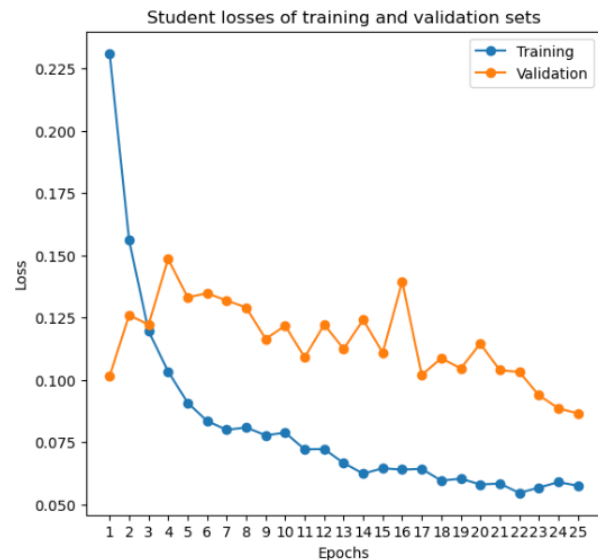
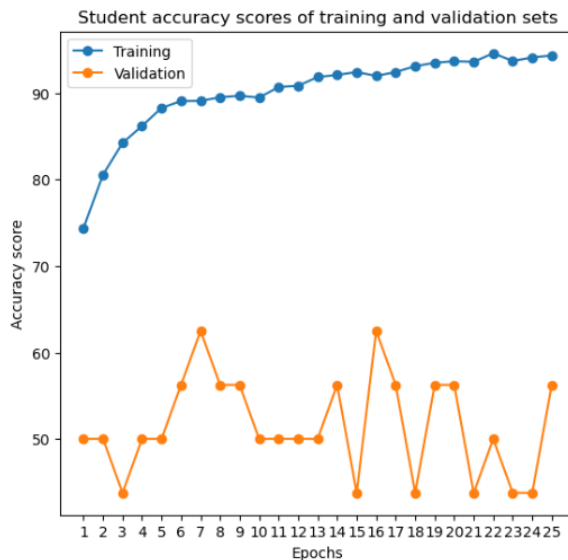
3) Final Model details:

To the original models, **Lr regularization done , Dropout layer added, Weight decay added, Data Augmentation**

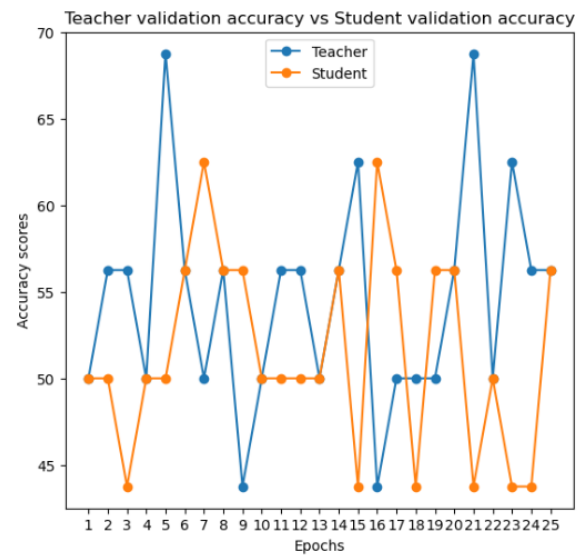
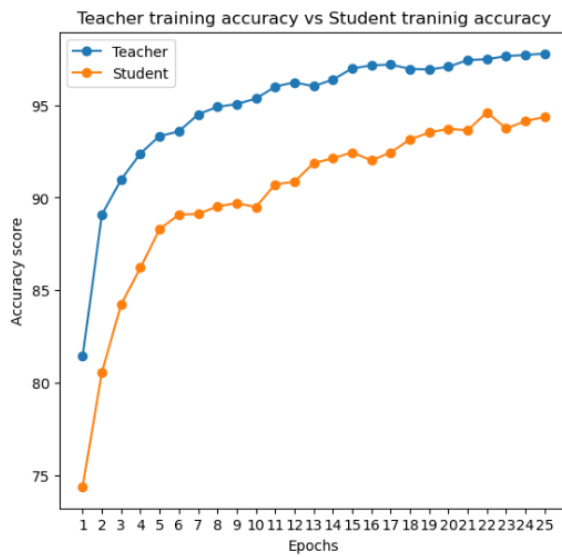
- The model was run on 25 epochs
- Teacher accuracies and losses for training and validation datasets:



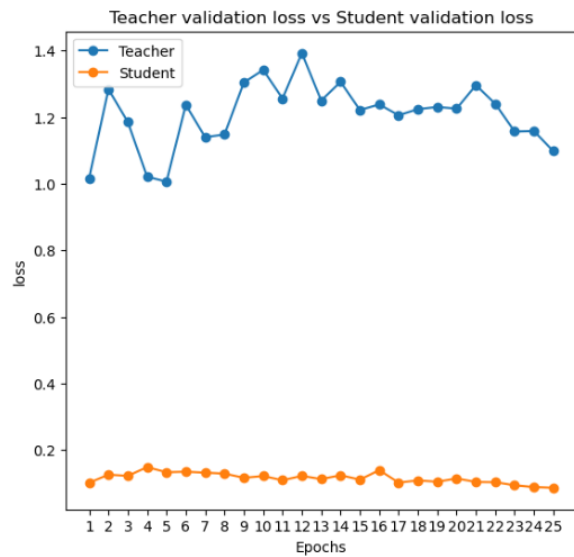
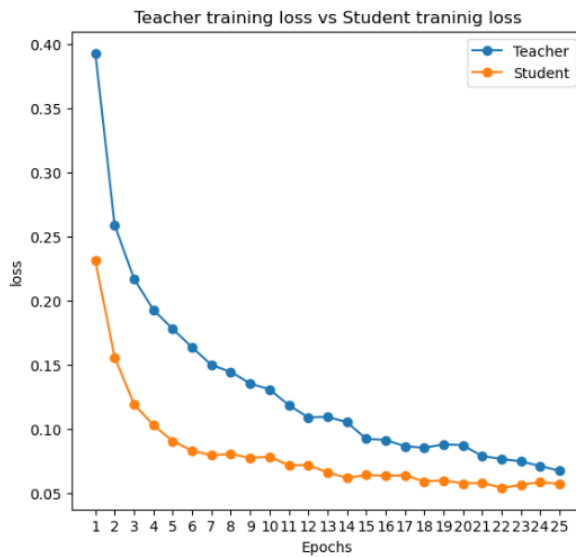
- Student accuracies and losses for training and validation datasets:



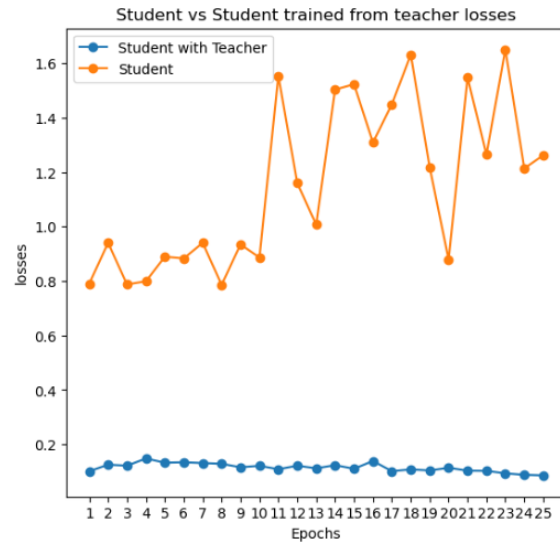
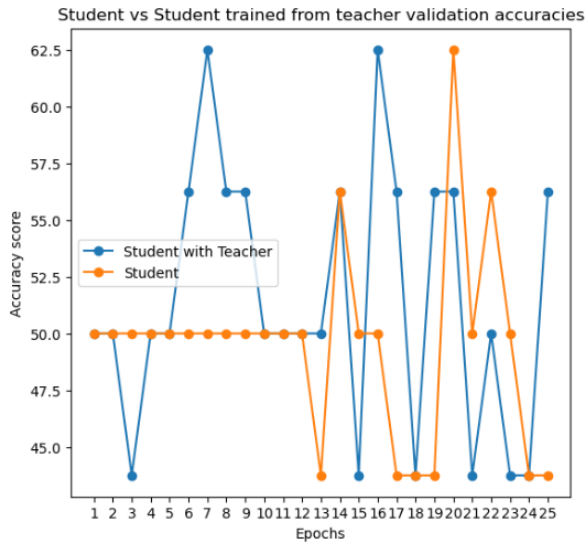
- Comparison between Teacher and Student for accuracies



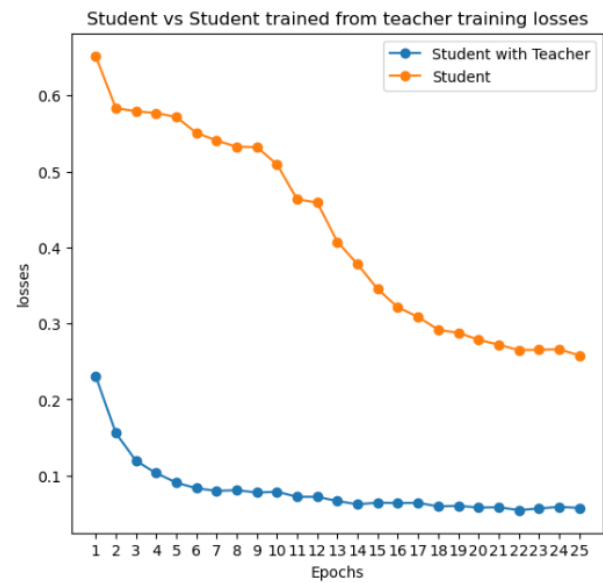
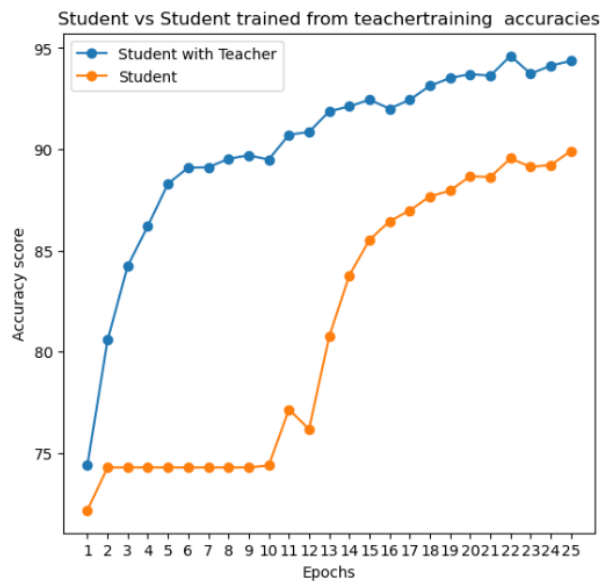
- Comparison between Teacher and Student for losses:



- Teacher accuracy is 80.12% and Student accuracy is 79.8% on test data
- Comparison between Only Student and Student with Teacher:
 - Losses



○ Accuracies



Therefore Student model performed better when trained from a Teacher model than when trained as a standalone model (Hence the concept of Knowledge Distillation is successful)

