

In this project, the transfer learning of three visual recognition models were compared – VGG-11, ResNet-18, and DenseNet121.

Preprocessing:

In the preprocessing stage, the images were imported in batches of fifty and enriched with randomized resizing, cropping, and flipping; the enrichment was conducted to increase the difficulty of the learning task, with the goal of preventing overfitting. Two train-validation-test splits were performed: 40-30-30 (respectively) and 50-25-25 (respectively). All models were run on both types of data splits. Results will be presented per model and per data split.

The models:

The modifications made to the pretrained models were similar across all models. The final fully connected layer in the pretrained models were replaced with a new fully connected layer, containing the new task's number of training classes and the trained weights of the new task. All models were trained with fifty epochs.

The following table present the accuracy and loss scores for all the models and data split types. The results for each model and data split are then presented individually, alongside visualizations of their accuracy and loss curves.

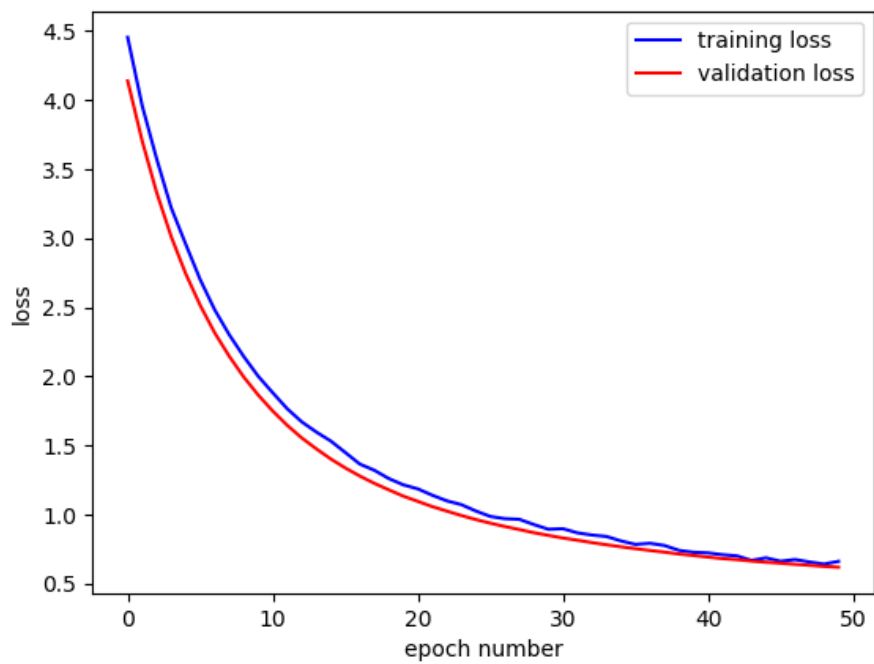
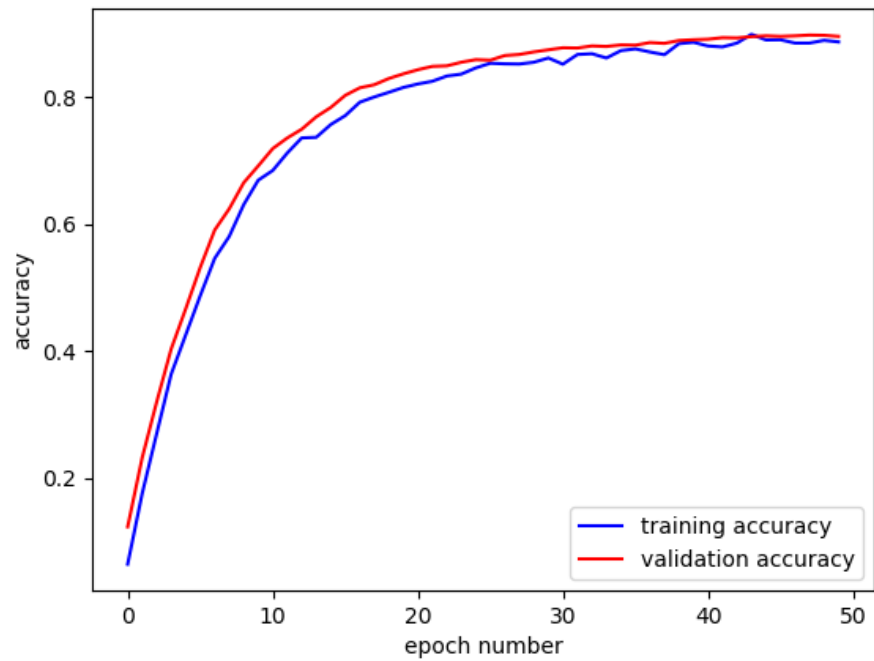
		Accuracy			Loss		
		Train	Validation	Test	Train	Validation	Test
40-30-30	ResNet-18	0.886718	0.895401	0.891331	0.662	0.619297	0.609655
	VGG-11	0.902901	0.89418	0.894587	0.512027	0.481379	0.485419
	DenseNet121	0.926718	0.9186	0.921042	0.503499	0.50154	0.502231
50-25-25	ResNet-18	0.897899	0.90816	0.894043	0.577332204	0.517990635	0.536989
	VGG-11	0.887885	0.909135	0.902832	0.529961	0.42677	0.438972
	DenseNet	0.912066	0.943332	0.928711	0.482592	0.403073	0.432185

As the results show, the highest performing model on the accuracy metric is DenseNet121, with an accuracy score over 92% in both data splits, much higher than the competing models. As for the effect of data split, the average test loss score in the 50-25-25 split is 0.469382, while the average test loss score in the 40-30-30 split is 0.532435. The difference in accuracy scores between data splits is vanishingly - the average test accuracy score in the 50-25-25 split is 0.9085287, while the average test accuracy score in the 40-30-30 split is 0.90232. The best performer by far is DenseNet121 with the 50-25-25 data split. Unfortunately, there are not enough data points to examine the statistical significance of these findings.

ResNet-18

Train-validation-test split: 40-30-30

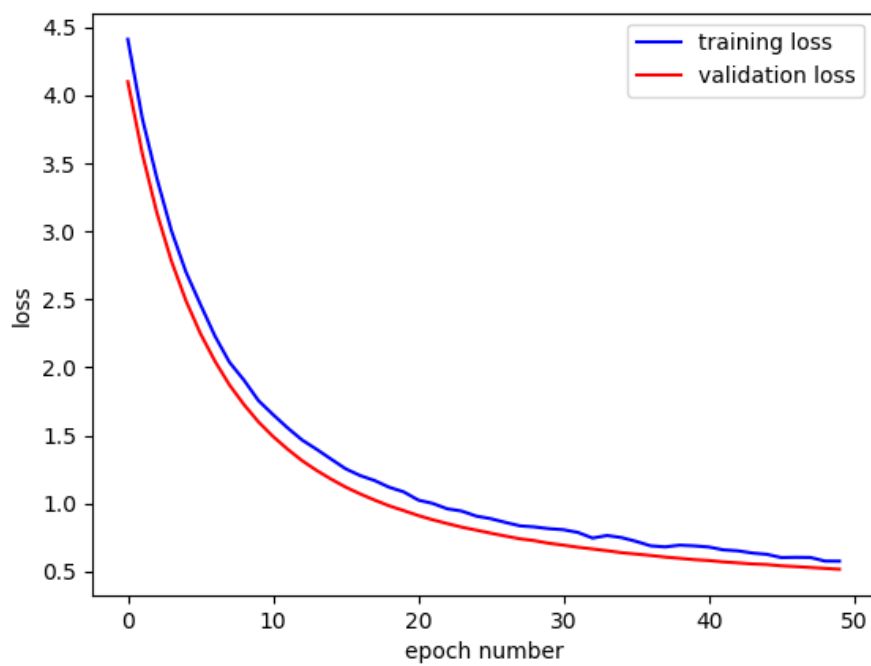
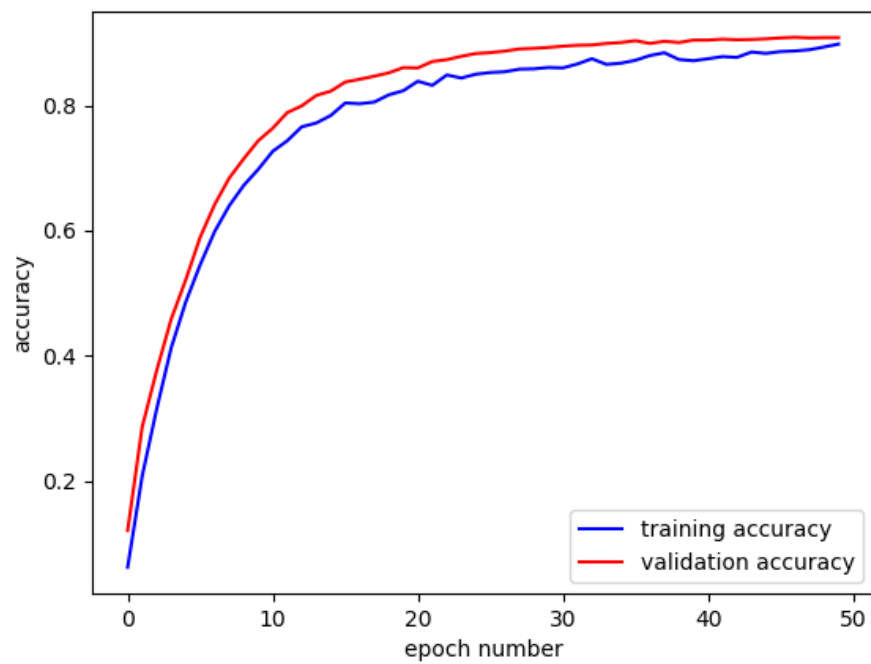
	Accuracy	Loss
Train	0.886718	0.662
Validation	0.895401	0.619297
Test	0.891331	0.609655



ResNet-18

Train-validation-test split: 50-25-25

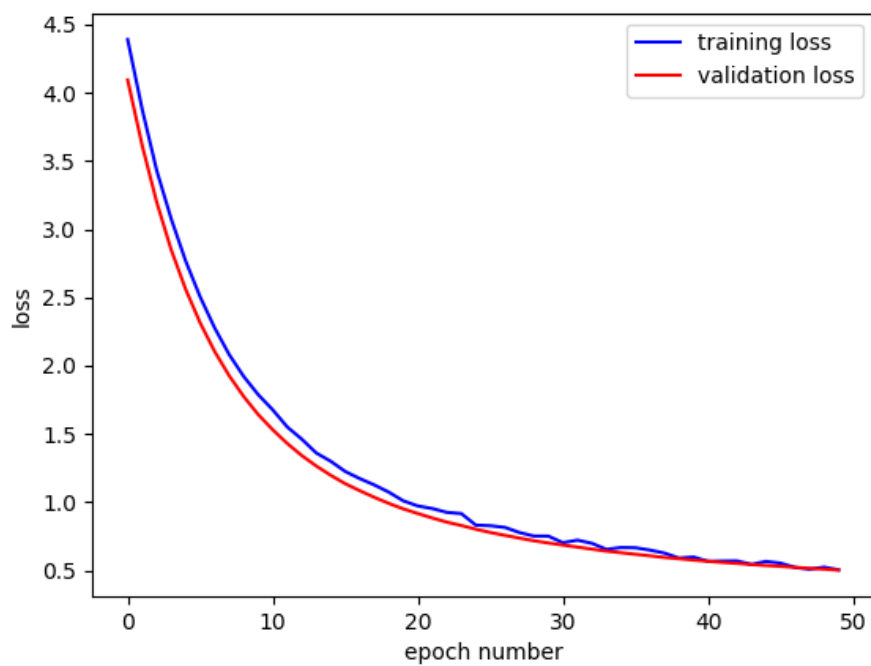
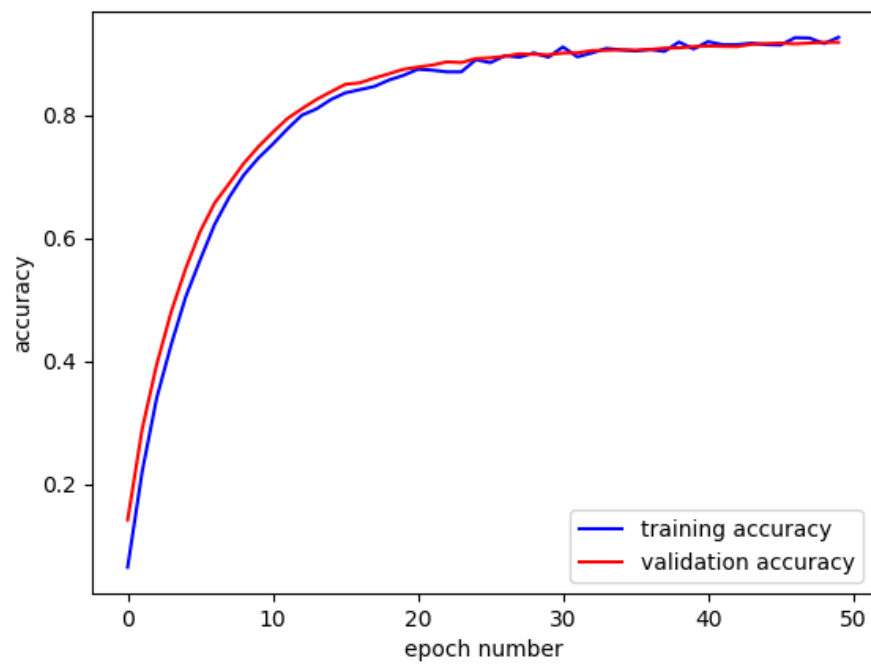
	Accuracy	Loss
Train	0.897899365	0.577332204
Validation	0.90815828	0.517990635
Test	0.894043	0.536989



DenseNet121

Train-validation-test split: 40-30-30

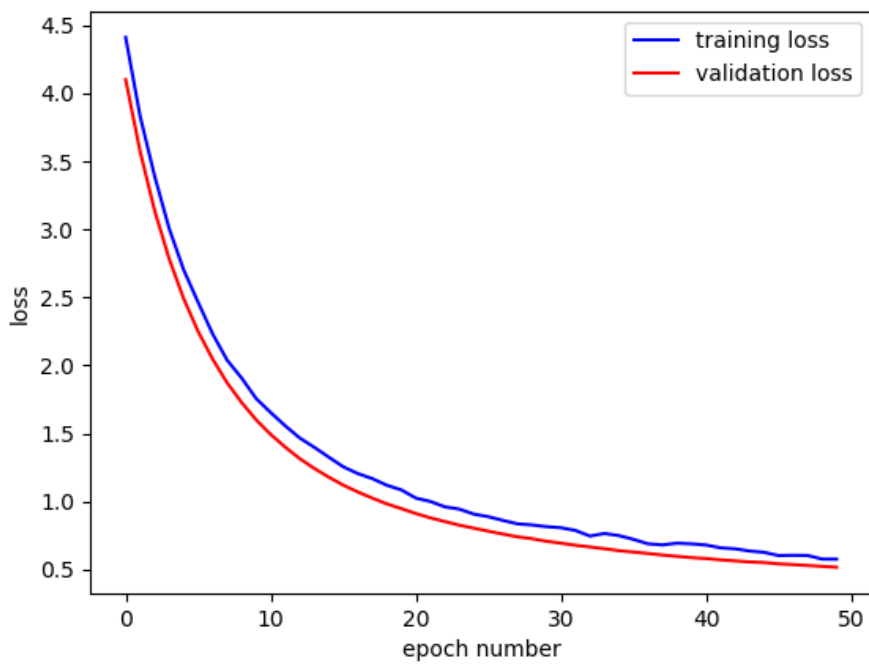
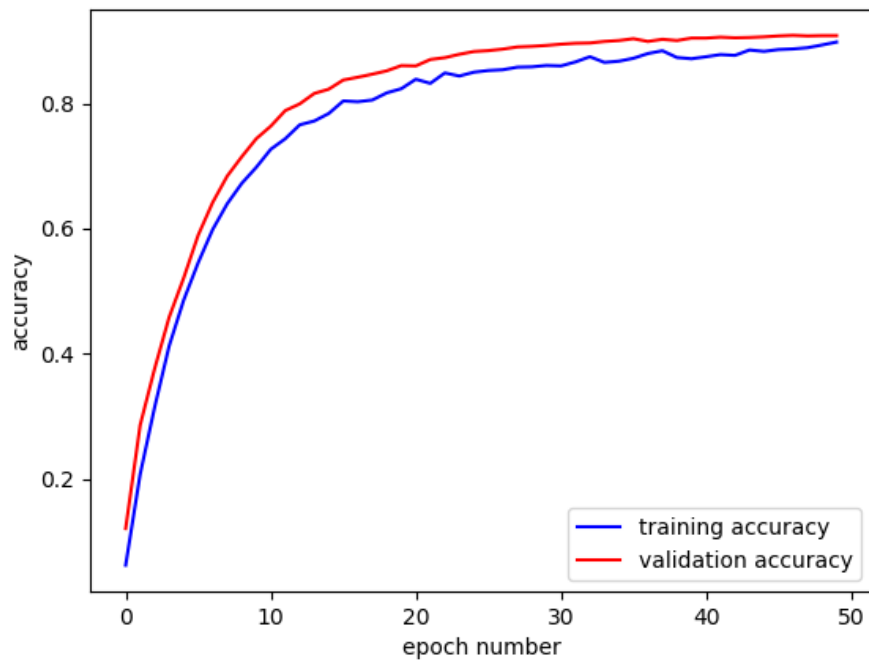
	Accuracy	Loss
Train	0.926718	0.503499
Validation	0.9186	0.50154
Test	0.921042	0.502231



DenseNet121

Train-validation-test split: 50-25-25

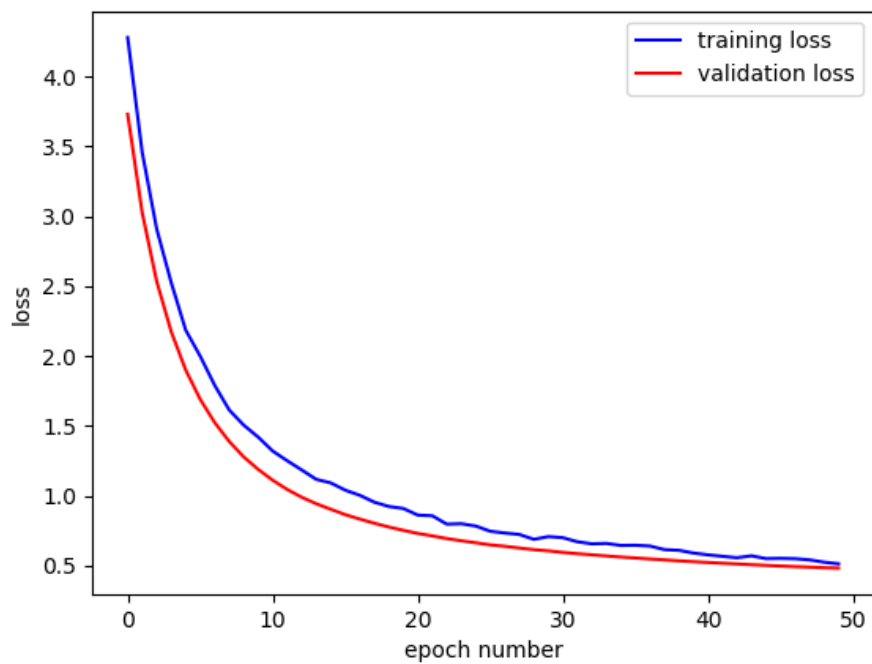
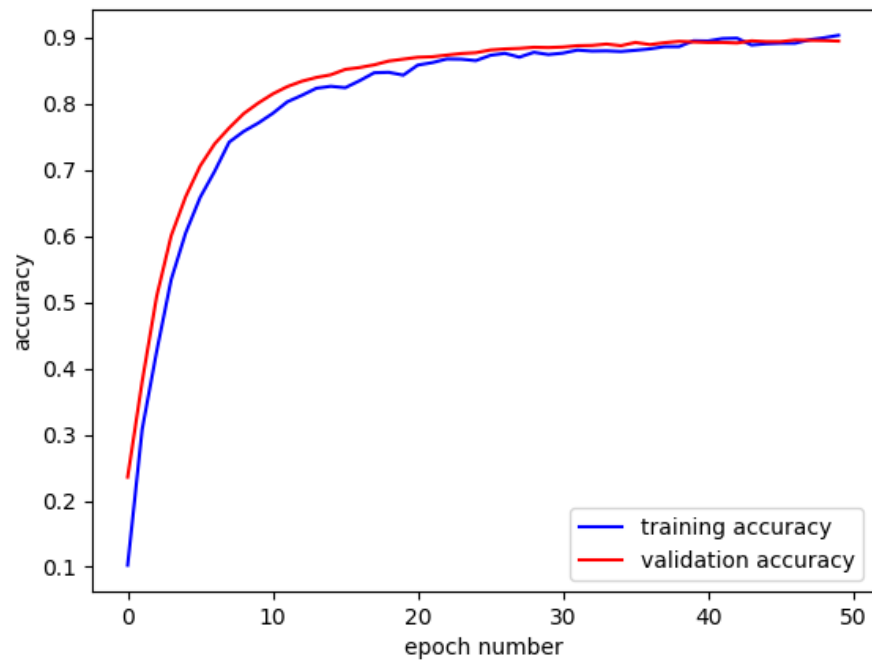
	Accuracy	Loss
Train	0.912066	0.482591578
Validation	0.943332	0.40307297
Test	0.928711	0.432185



VGG-11

Train-validation-test split: 40-30-30

	Accuracy	Loss
Train	0.902901	0.512027
Validation	0.89418	0.481379
Test	0.894587	0.485419



VGG-11

Train-validation-test split: 50-25-25

	Accuracy	Loss
Train	0.887885	0.529961
Validation	0.909135	0.42677
Test	0.902832	0.438972

