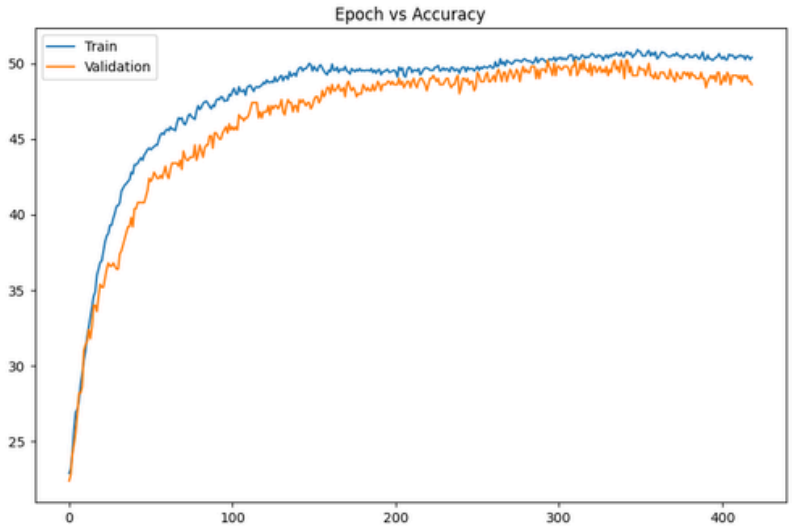
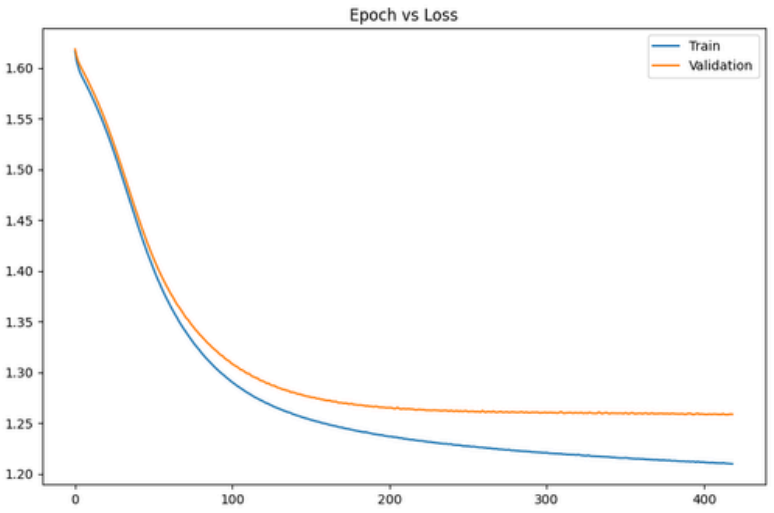
Task2: Comparison of normalization methods for classification on Image dataset 2

Hyper parameters: Seed - 42 Learning rate - 0.0003 Threshold- 1e-6

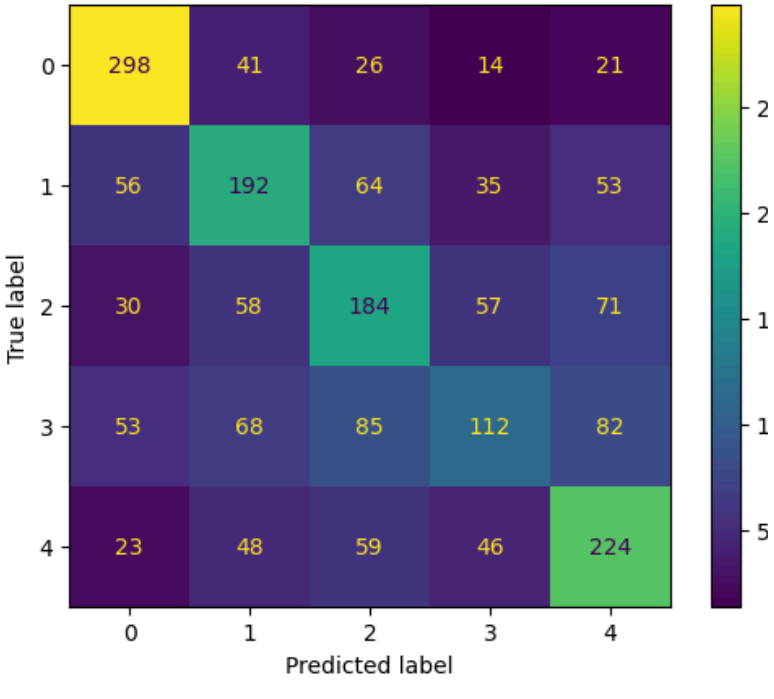
No Normalization

- The activations are used as they are, without any modification.
- Can lead to unstable training, slower convergence, and issues like vanishing or exploding gradients.
- Epochs to converge -419
- Training accuracy - 50.4%
- Validation accuracy - 48.6000%

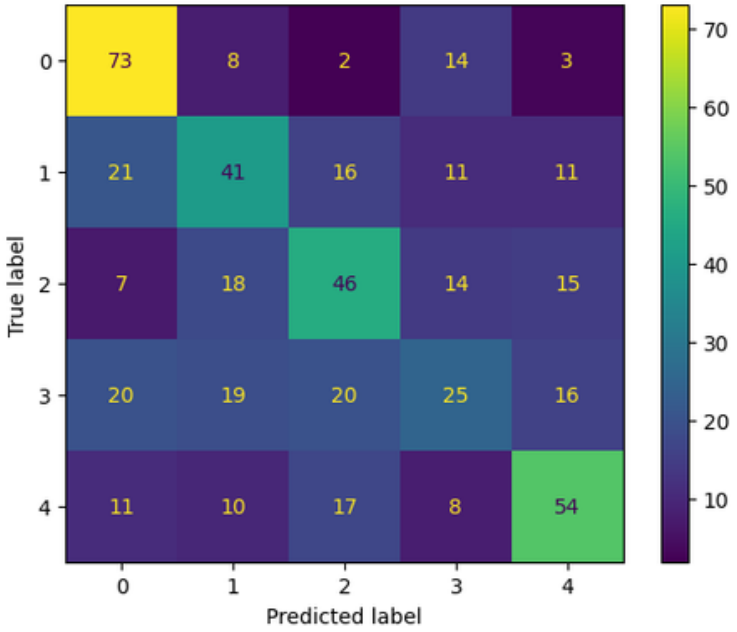




Confusion matrix for train data

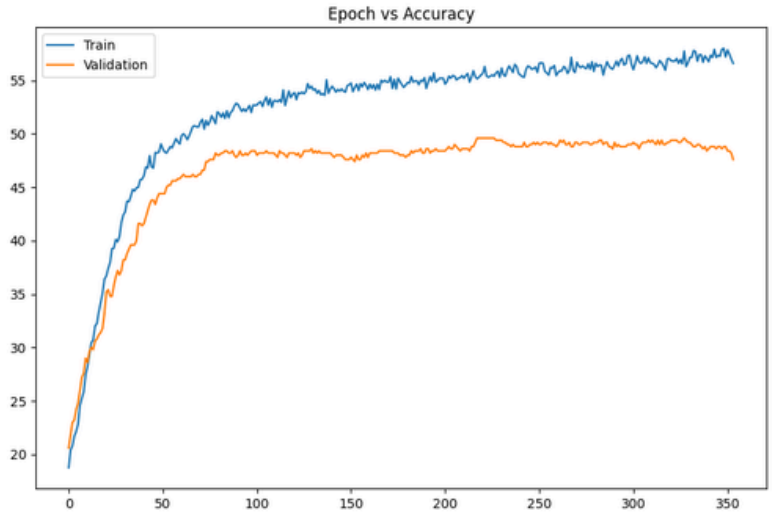


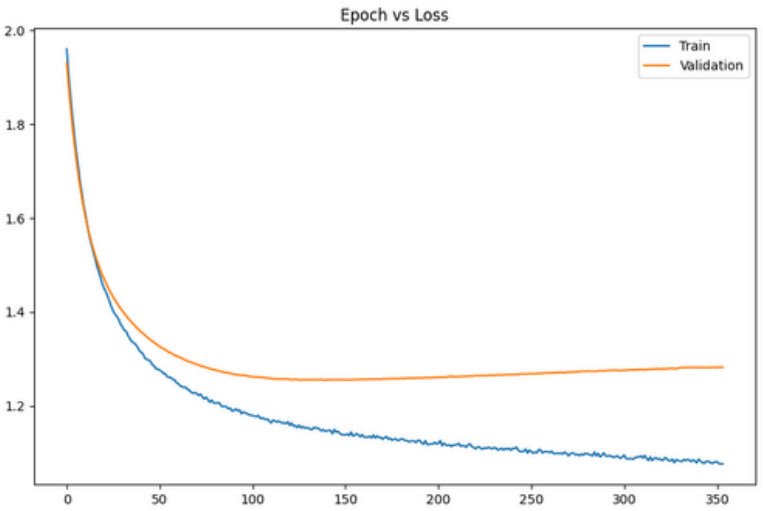
Confusion matrix for test data



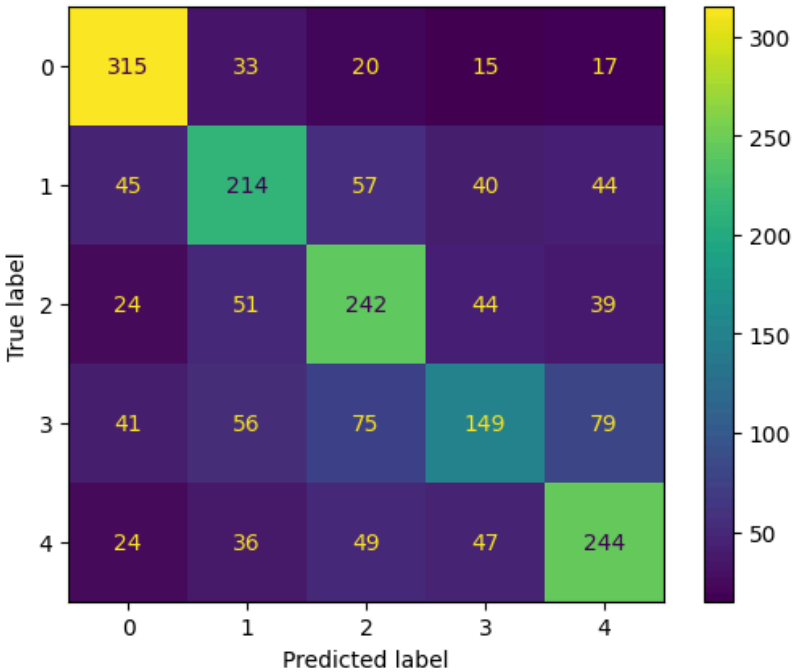
Batch Normalization

- Normalization is applied across the batch dimension and stabilizes training.
- Epochs to converge -354
- Training accuracy - 57%
- Validation accuracy - 48.2%

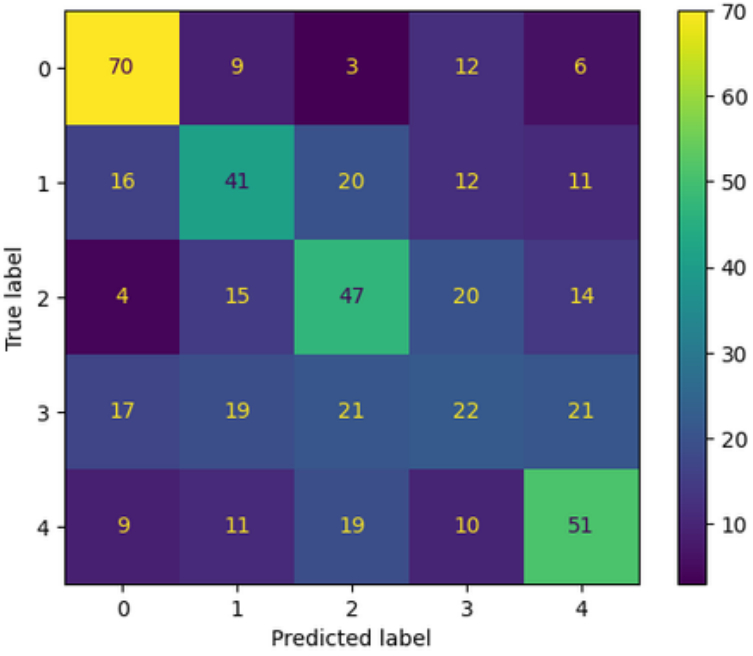




Confusion matrix for train data

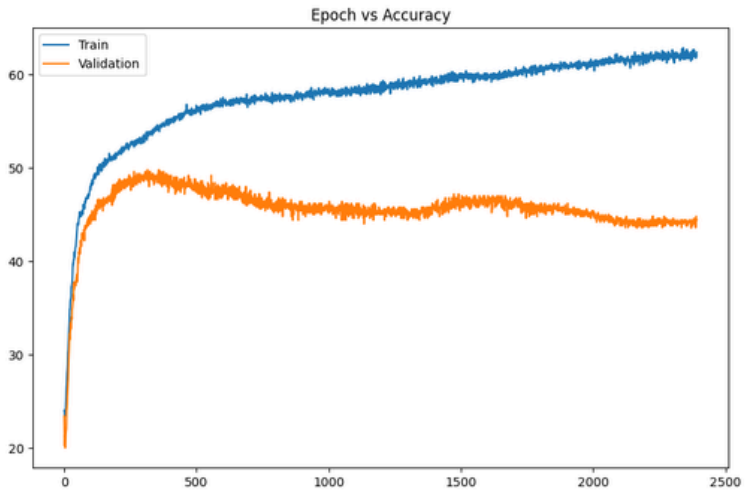


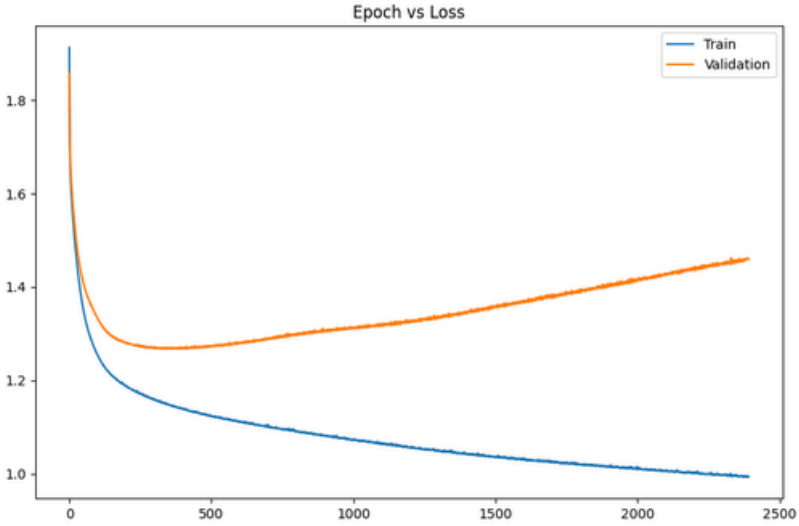
Confusion matrix for test data



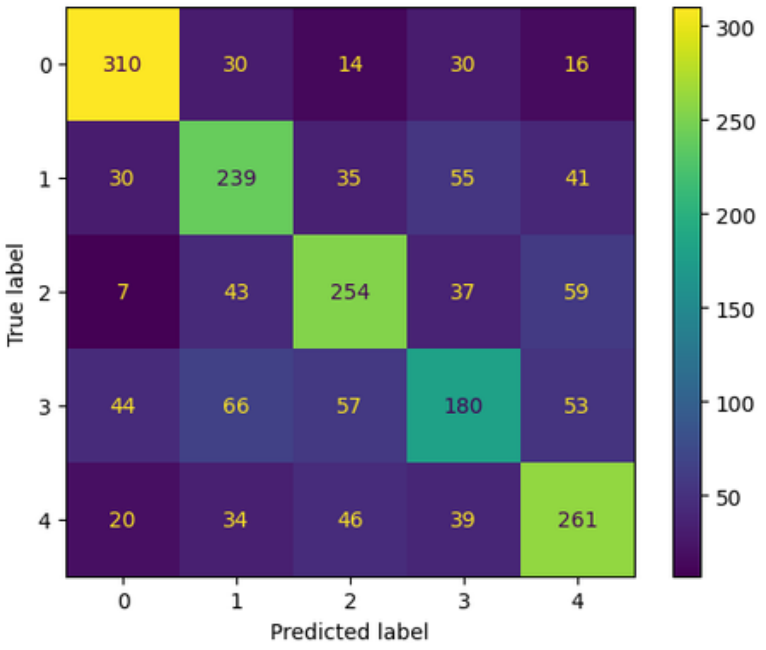
Layer Normalization

- Normalization is applied across the feature dimension for each sample making it more efficient but slow.
- Epochs to converge -2392
- Training accuracy - 62.4%
- Validation accuracy - 44%

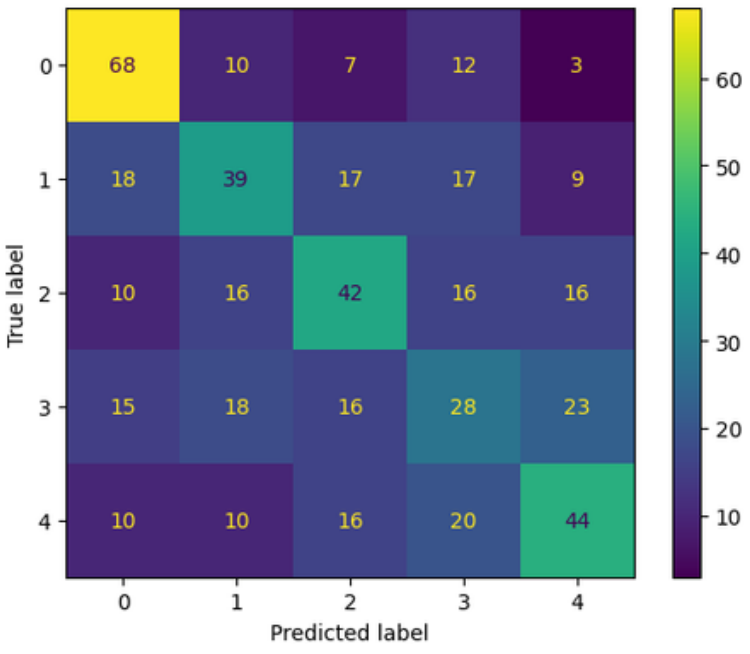




Confusion matrix for train data



Confusion matrix for test data



OBSERVATIONS

- By looking at the confusion matrix for the three methods we can see that the layer normalized model performs better on training data but it other two models outperforms it in test data
- The number of epochs it took for Layer normalization is very high compared to other methods.
- High training accuracy and low validation accuracy in Layer Normalization indicates overfitting to some extent. The validation dropping accuracy with epochs from the graph indicates the same.
- Batch Normalization and No normalization has similar validation accuracies but batch normalized model converged quick.
- Looking at the learning curve we can see that the normalized model converges to a lower loss value as compared to other models also, the expected rate of convergence is much faster for normalized. But there is not much change in validation loss which indicates that the model is overfitting, to overcome this we need to prune our model or increase training data.