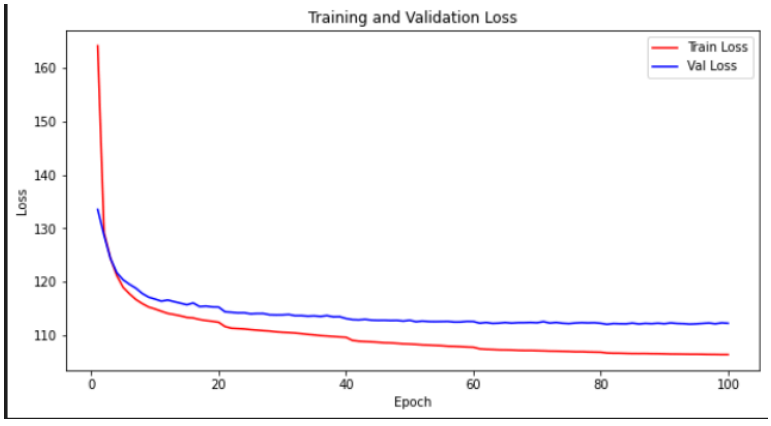
Somesh Agrawal (2024AIB2292)
Shreyash Padeer(2024AIB2800)

Below is a model summary for the input dimension of (1,28,28) for a given image. Encoder -5 represents the end of the encoder model and the start of the decoder model. In total, we have 1,725,204 parameters for training the model.

```
----------------------------------------------------------------
        Layer (type)               Output Shape         Param #
================================================================
            Linear-1                  [-1, 512]         401,920
            Linear-2                  [-1, 256]         131,328
            Linear-3                    [-1, 2]             514
            Linear-4                    [-1, 2]             514
           Encoder-5          [[-1, 2], [-1, 2]]               0
            Linear-6                  [-1, 256]             768
            Linear-7                  [-1, 512]         131,584
            Linear-8                  [-1, 784]         402,192
           Decoder-9                  [-1, 784]               0
================================================================
Total params: 1,068,820
Trainable params: 1,068,820
Non-trainable params: 0
```

Specifications for the above model:
Epochs: 100
Learning rate: 0.0009
Optimizer Used: Adam
Scheduler Used: StepLR, with gamma = 0.5, step_size = 20

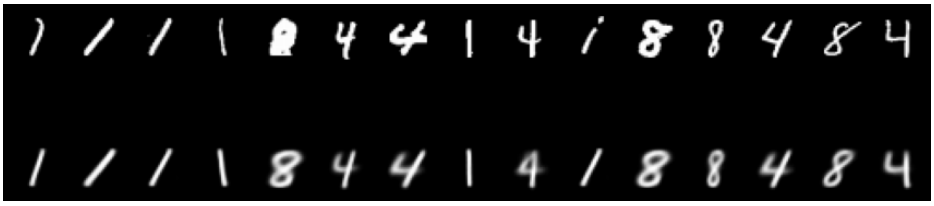
Below is the graph of training loss vs validation Loss over 100 Epochs of the model.



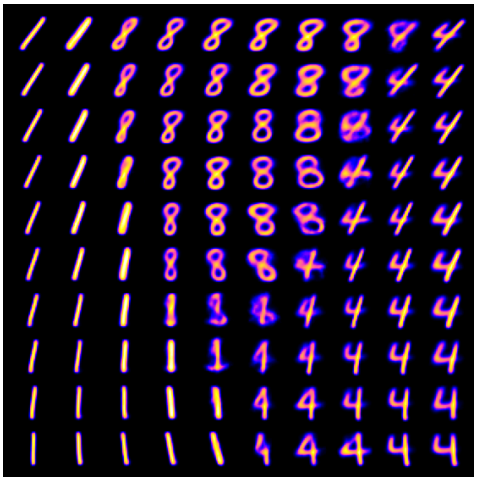
Final Results on the Model Trained

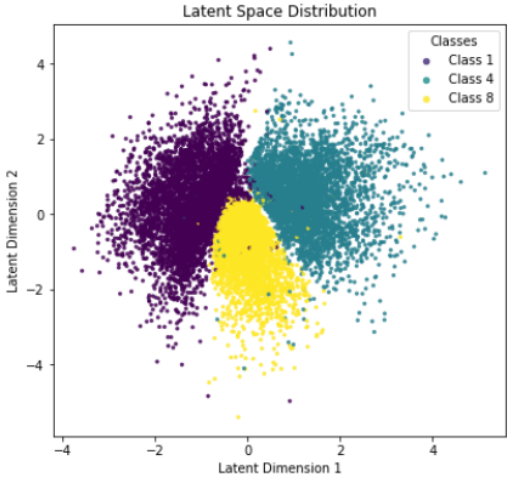
Structural Similar Score: 0.70474	MSE Loss: 0.0266

We split the data to training and validation in a ratio of 90:10. First Row corresponds to the actual image fed to the network and images in the lower row correspond to the reconstructed image.



In this study, a two-dimensional manifold was plotted, representing values ranging from 0.05 to 0.95, serving as the latent vector input for the decoder in generating images of digits. Moving along the horizontal axis (left to right), the X-coordinate increases from 0.05 to 0.95, while moving vertically (bottom to top), the Y-coordinate similarly increases from 0.05 to 0.95. This configuration captures a smooth latent distribution across the plane, as shown in the corresponding figure, which illustrates the distribution of points in the latent space.





A Gaussian Mixture Model (GMM) was subsequently trained on this latent space to cluster data points into three distinct clusters, each corresponding to specific labels (1, 4, and 8). The encoder generates a latent vector for each data point, which is then input to the GMM for assigning a cluster label to the image.

It is observed in the figure that clusters exhibit partial overlap. In cases of such overlapping clusters, the cluster assignment is determined by selecting the cluster with the highest probability, which may lead to occasional misclassifications due to the inherent ambiguity in overlapping regions.

The Following Metrics have reported all the data after running the full end to end model.

	Accuracy(in %)	Precision(x100)	Recall(x100)	F1 Score(x100)
Training Data	96.77	96.68	96.840	96.73
Validation(Split)	96.67	96.62	96.825	97.711
Validation(Given by TA)	100	100	100	100