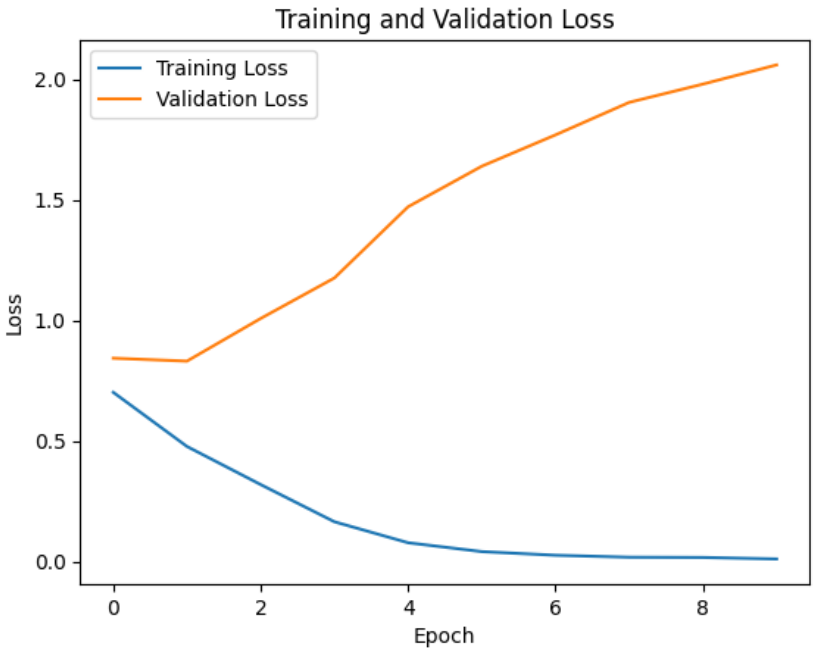
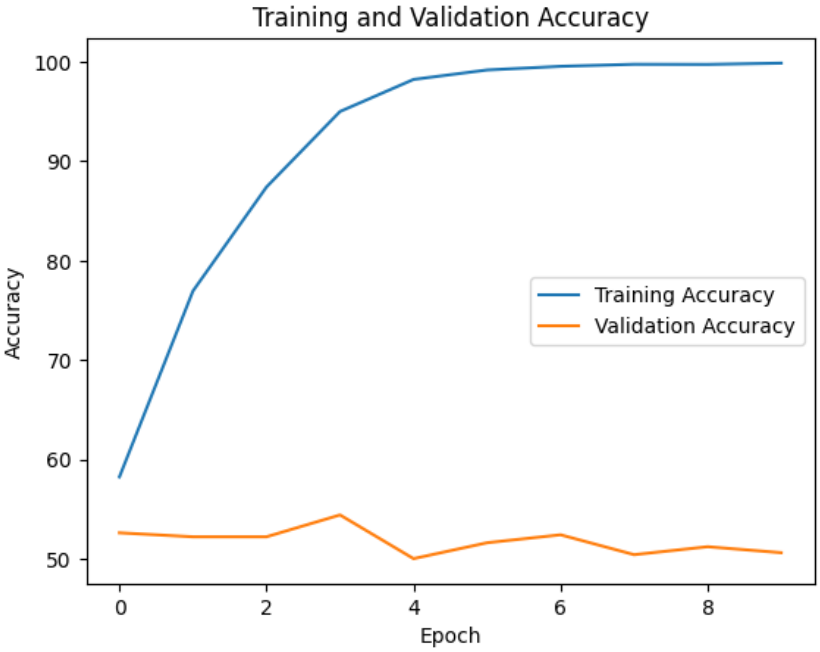
DL A4 Report

Q1 - Param Shah (MT22048)
Q2 - Tarang Viroja (MT22081)
Q3 - Manvendra Kumar Nema (MT22038)

Q1)



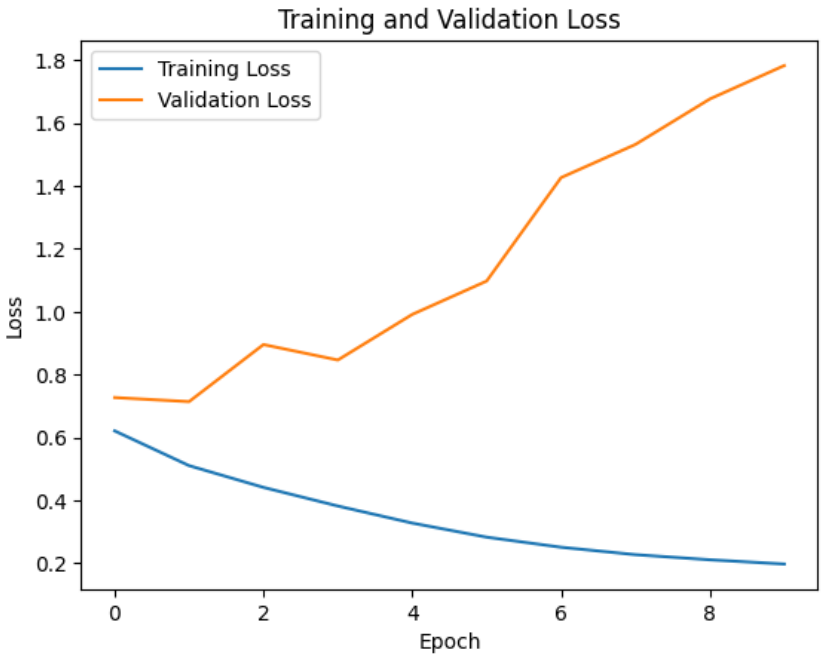
Train loss decreases, but validation loss increases, implying the model is overfitting.



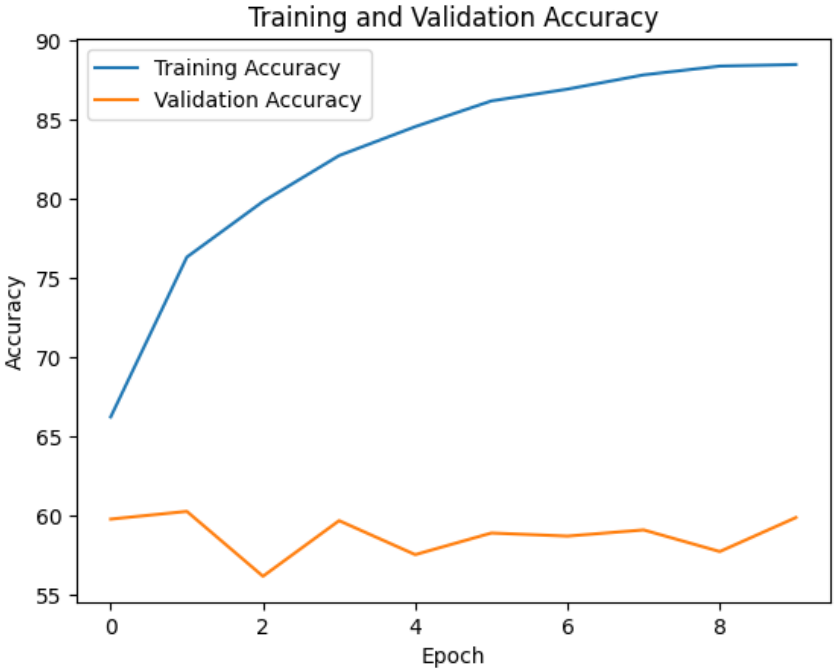
Validation loss remains almost constant, which confirms overfitting. The model is unable to generalise the relationship on the validation dataset.

Test Accuracy: 54%
Overall Precision: 0.54
Overall Recall: 0.53
Overall F1 Score: 0.52
Class 0 - Precision: 0.53, Recall: 0.72, F1 Score: 0.61
Class 1 - Precision: 0.54, Recall: 0.35, F1 Score: 0.42

Q2)



Validation loss increases, implying that the model is overfitting.



Validation accuracy isn't changing a lot which confirms overfitting. The model is unable to generalize the relationship over the validation set. Text performs Better than visual data.

Test Accuracy: 55%

Overall Precision: 0.5687

Overall Recall: 0.5554

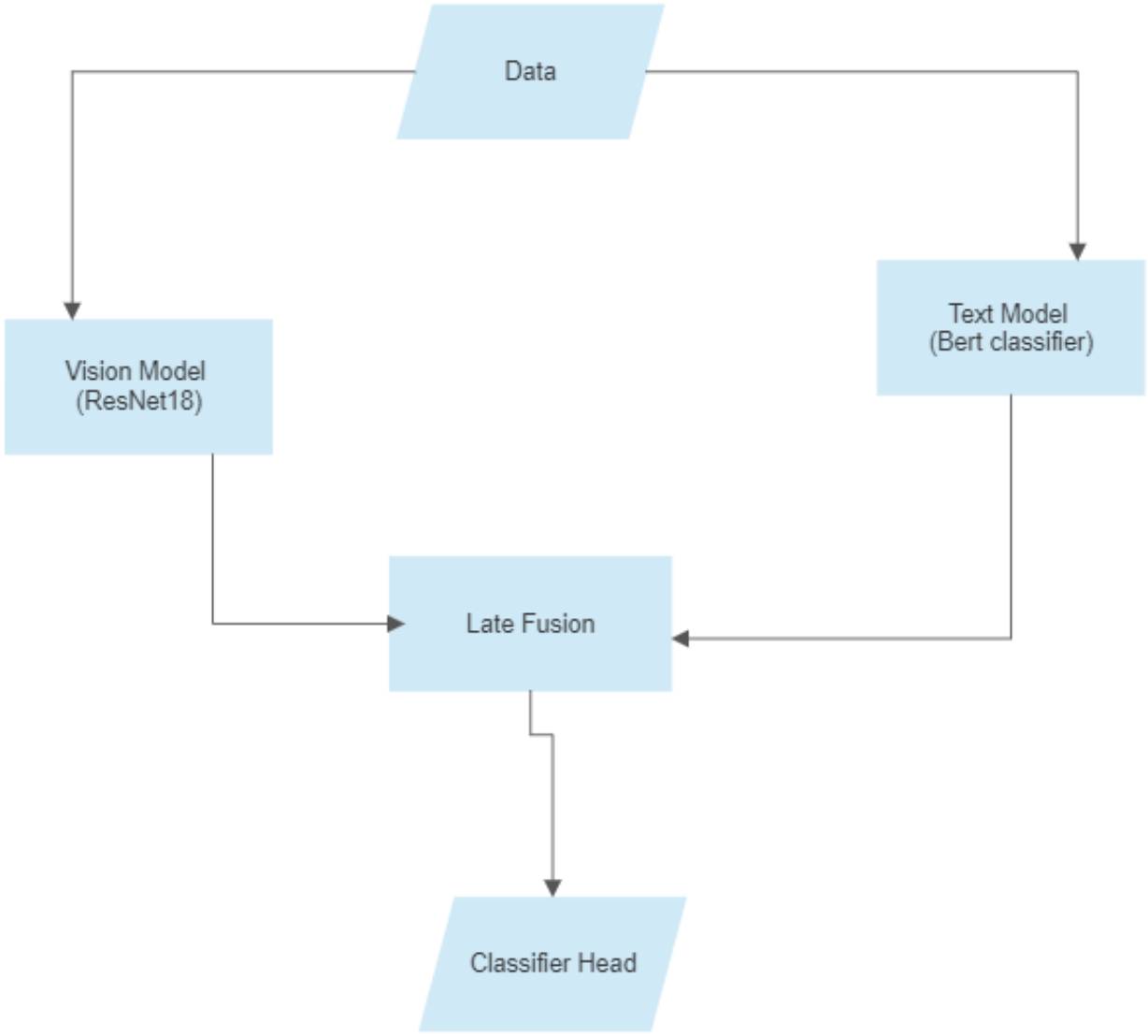
Overall F1 Score: 0.5342

Class 0 - Precision: 0.5444, Recall: 0.7747, F1 Score: 0.6395

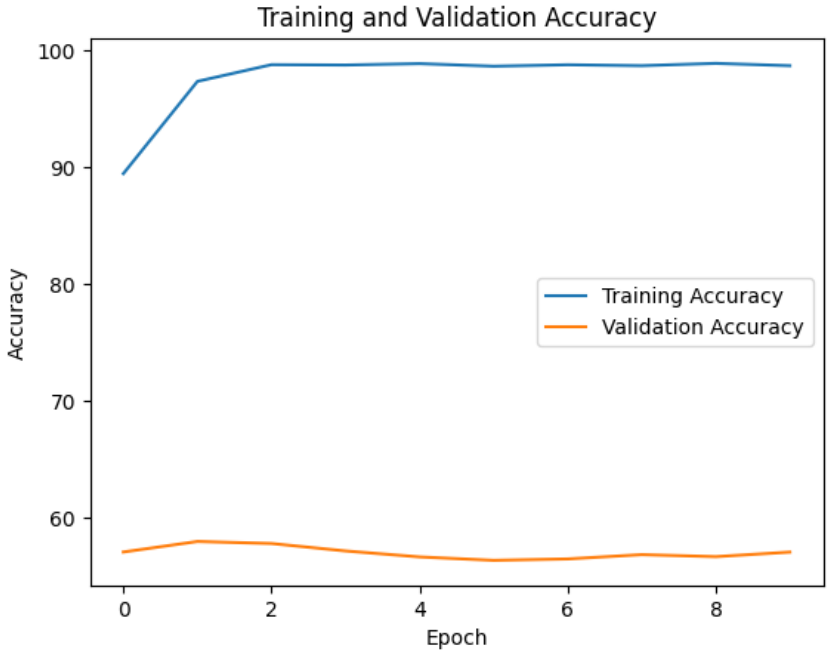
Class 1 - Precision: 0.5929, Recall: 0.3360, F1 Score: 0.4289

Q3)

Model-Explain: We have used a pre-trained vision model and text model from Q1 and Q2, i.e. ResNet18 and Bert classifier, and then have merged them using late fusion by one linear Layer followed by a classification head.



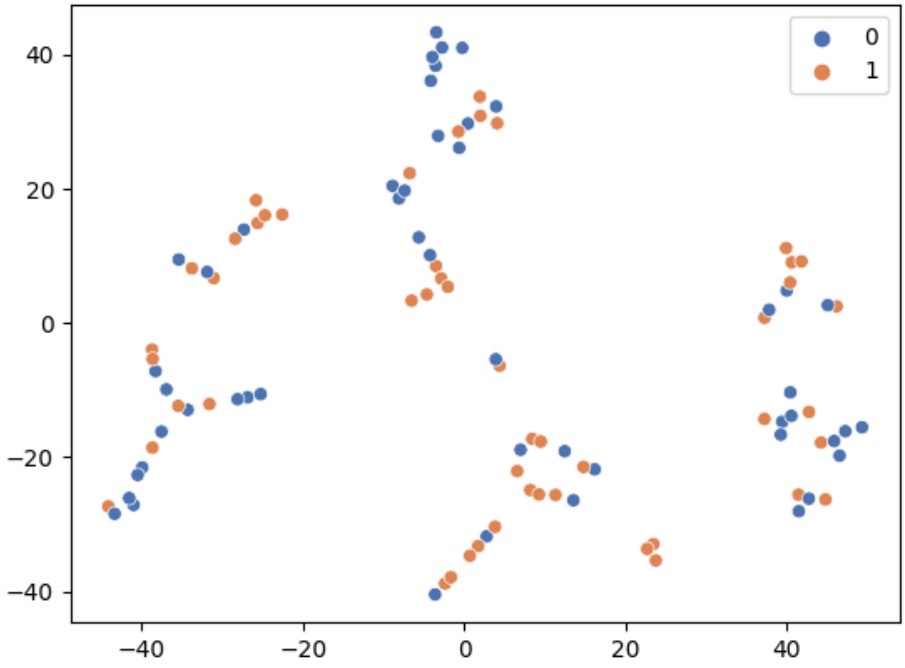
Train loss decreases, but validation loss increases, implying the model is overfitting.



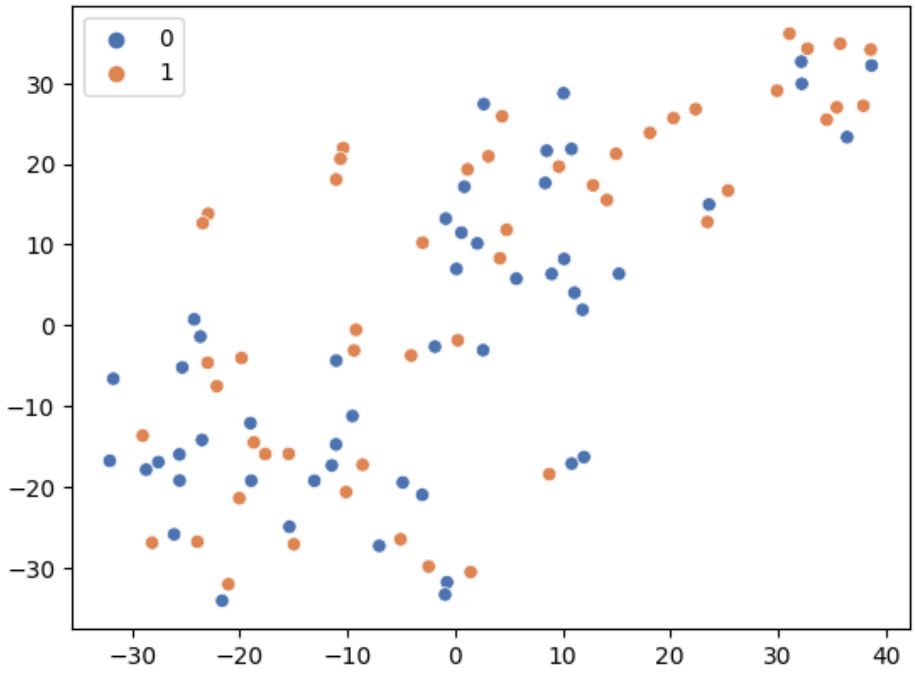
Validation loss remains almost constant, which confirms overfitting. The model is unable to generalise the relationship on the validation dataset.

Test Accuracy :- 55.6 %
Overall Precision: 0.5729
Overall Recall: 0.5529
Overall F1 Score: 0.5216
Class 0 - Precision: 0.5407, Recall: 0.8142, F1 Score: 0.6498
Class 1 - Precision: 0.6050, Recall: 0.2915, F1 Score: 0.3934

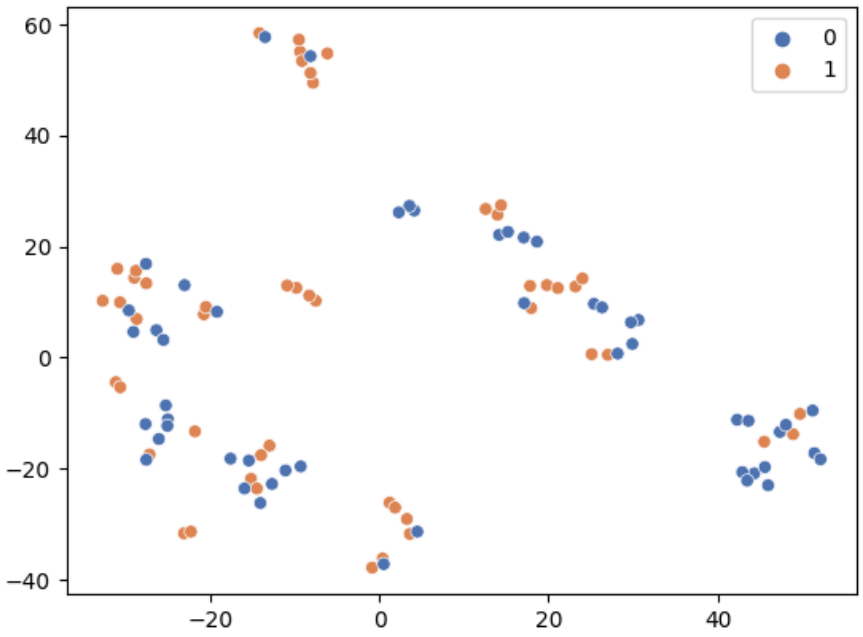
TSNE-Q1



TSNE-Q2



TSNE-Q3



Uni-Text Vs Multi-Model

- In Multi-Model, there is a point that is **slightly** more spreadable than **Text**-model.
- This means that multi-model has a better understanding of the data for classification.

Uni-Image Vs Multi-Model

- In Multi-Model, there are points that are **much** more spreadable than in the **Image**-model.
- This means that multi-model has a better understanding of the data for classification.

In order to understand meme knowledge of the whole context sarcasm, which is often not possible just from image and text, which is why models are unable to generalize.