Model-1:

- The training performance of this model is very high, with accuracy reaching up to 99.26%, similar to the previous model.

- However, the validation performance in this set of images shows a discrepancy, with a notable loss increase during training, suggesting potential overfitting.

- The test performance (95.26%) is consistent with high confidence predictions, but when evaluating on a different dataset, the accuracy drops to 77.53%, which is a substantial decrease and suggests that the model might not generalize as well to new data compared to the training set.

- \*\*model1 vs model-3\*\*:

- The previous model had perfect ROC curves with AUC of 1.00 for both training and validation, indicating potentially overfit models.

- The accuracy on an additional test set for the previous model was 91.08%, which was lower than the training but higher than this current model's 77.53% on a different dataset.

In summary, while both models demonstrate high training accuracies, they show signs of overfitting as indicated by the decrease in performance on different datasets. The current model's architecture, with a significant number of parameters, may contribute to its ability to fit the training data well but also makes it more prone to overfitting, as seen by the less consistent performance on the validation set and additional dataset. The previous model seems to have had a better balance, with a somewhat more consistent but still imperfect generalization to new data.

Model-2

Model-3

The second image is also a ROC curve with an AUC of 1.00 for the training set. Just like the first, this indicates perfect classification performance on the training data.

The third image is a snippet of code output showing the training process of a model over three epochs. The model has achieved high accuracy on both the training set (up to 99.36%) and validation set (up to 98.97%), with decreasing loss over time, which is a sign of good fitting to the data.

Comparing these outputs:

- The ROC curves indicate that the model has perfect performance on both training and validation sets. However, such perfect scores (AUC = 1.00) may suggest overfitting, as it is unusual for a model to perform without any errors unless the problem is very simple or the dataset is not diverse enough.

- The training log shows a consistent improvement in performance, with accuracy increasing and loss decreasing, which is generally a good sign. However, the high accuracy seen here (over 99%) may not always translate to new, unseen data.

- The predictions on the test set show a very high confidence level (predictions around 0.9935), which aligns with the high performance indicated by the ROC curves.

- The evaluation on what seems to be another validation or test set shows slightly reduced performance (91.08% accuracy), which could indicate the model's actual performance on unseen data and suggests that the perfect ROC curves might not fully represent the model's generalization capabilities.

Overall, these images suggest that the model performs exceptionally well on the training and validation datasets but has a slightly reduced accuracy when applied to a separate test set, which is a more realistic indicator of its performance on unseen data. The perfect ROC curves suggest that either the model is exceptionally good or there might be a concern of overfitting. The evaluation accuracy of 91.08% seems more realistic for real-world performance, although it's still quite high.

Model-2:

Comparing the results from training and testing, and the performance of this model with the previous models discussed:

- The training and validation accuracy of this model is high, with the validation accuracy being slightly higher initially but then decreasing, which might suggest overfitting as the model trains.

- The ROC curve with an AUC of 0.89 is not as perfect as the AUC of 1.00 seen in the previous models, but it might be a more realistic representation of the model's performance on unseen data.

- The test predictions show very low probabilities, which contrasts with the high confidence seen in the previous model's predictions. This might indicate that the model is more conservative in predicting the positive class or that the test data is significantly different from the training data.

- The accuracy on the different dataset (82.59%) is lower than the training accuracy but is in the same range as the previous model evaluated on its different dataset (77.53%). This consistency in performance drop suggests that both models are likely overfitting to their respective training data and not generalizing as well as desired.

Overall, "sequential\_1" seems to be a smaller model in terms of parameters compared to the previous models, which had around 6,901,001 parameters. Despite this, it shows signs of overfitting, much like the previous models, albeit to a lesser extent given the more realistic ROC AUC value. The lower performance on the different dataset compared to the training set suggests that there may be room for improvement in the model's ability to generalize.

Amongst all these three BILSTM models, the first model has the best validation accuracy which is above 0.99.

When these model was tested on the external dataset DAIGT, the model achieved 0.8432 accuracy, with this we can infer that the model is not generalizing well and not giving good results certainly, so to avoid this problem of overfitting we tried to add the regularizer ,dropout layer and early stopping and observe if the model has performed better or not. The model achieved the validation accuracy of 0.952 , and training accuracy of 0.783 by the end of the epoch 3. The model was then run over the external ( unseen ) data ( DAIGT) to evaluate its performance.

The AUC score is 0.89, which is quite high, indicating that the model has good discriminative ability between the positive and negative classes.The model's performance on DAIGT dataset captured a loss of 0.6784 and an accuracy of 0.825, which is lower than the training and validation accuracies, but still not too bad, and is definitely better than the model-1’s performance which was not at all good in capturing the unseen data with the accuracy of approx 0.7752.

Lastly, the model with the added custom attention layer, has the validation accuracy of 0.987 and the maximum accuracy on the testing data with 0.9107 and AUC score of . This tells that this model fits the best amongst these three models to the data.

**Table #**

*Table of Evaluation of all models on the testing dataset*

| **Model** | **Accuracy on test data** | **AUC** |
| --- | --- | --- |
| RNN | 0.8065 | 0.9036 |
| Bi-LSTM | 0.8432 | 0.9643 |
| Bi-LSTM with attention layer | 0.8203 | 0.9696 |
| BERT |  |  |