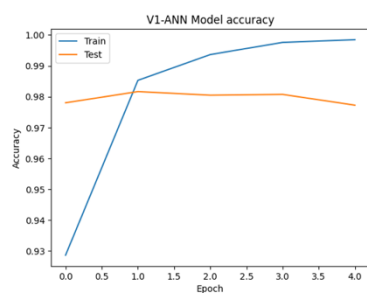# 4   Interpretation and Validation

4.1   Model performance

The evaluation of the models involved multiple metrics to assess their performance comprehensively. The primary metrics used include accuracy, precision, recall, and the confusion matrix. Here's a summary of the results obtained from different models:
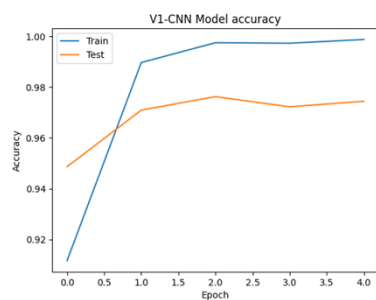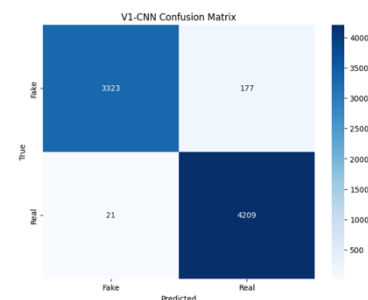
4.1.1   V1: 5 Epochs CNN & ANN

- ANN Model:

	precision	recall	f1-score	support
Fake	0.98	0.97	0.97	3500
Real	0.98	0.98	0.98	4230
accuracy			0.98	7730
macro avg	0.98	0.98	0.98	7730
weighted avg	0.98	0.98	0.98	7730
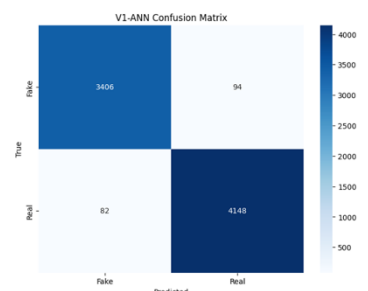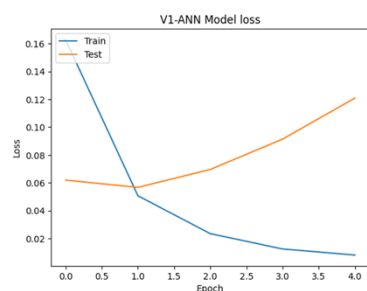
- CNN Model:

	precision	recall	f1-score	support
Fake	0.99	0.95	0.97	3500
Real	0.96	1.00	0.98	4230
accuracy			0.97	7730
macro avg	0.98	0.97	0.97	7730
weighted avg	0.98	0.97	0.97	7730
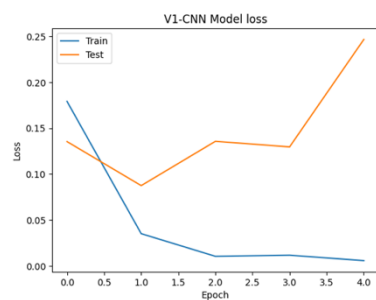
4.1.2 V2: 20 Epochs CNN & ANN

- ## ANN Model:
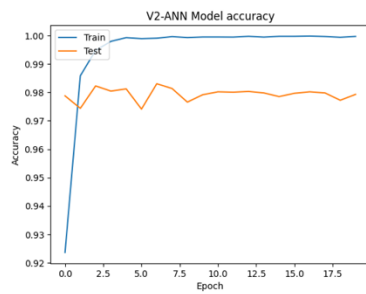
	precision	recall	f1-score	support
Fake	0.98	0.97	0.98	3500
Real	0.98	0.99	0.98	4230
accuracy			0.98	7730
macro avg	0.98	0.98	0.98	7730
weighted avg	0.98	0.98	0.98	7730
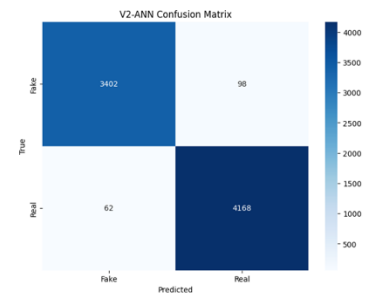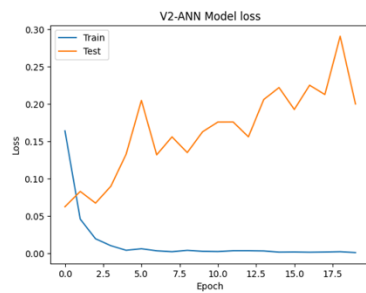
- ## CNN Model:
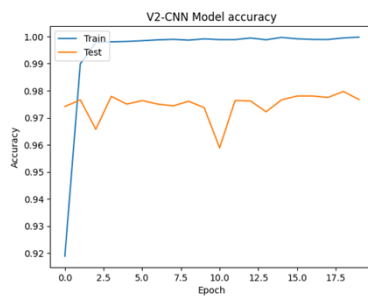
	precision	recall	f1-score	support
Fake	0.99	0.96	0.97	3500
Real	0.96	0.99	0.98	4230
accuracy			0.98	7730
macro avg	0.98	0.97	0.98	7730
weighted avg	0.98	0.98	0.98	7730

4.1.3 V3: ANN with 12 Epochs, Regularizers, Batch Normalization, and Early Stopping

V3-ANN Model accuracy

```
              precision    recall  f1-score   support

        Fake       0.96      0.97      0.97      3500
        Real       0.98      0.97      0.97      4230

    accuracy                           0.97      7730
   macro avg       0.97      0.97      0.97      7730
weighted avg       0.97      0.97      0.97      7730
```
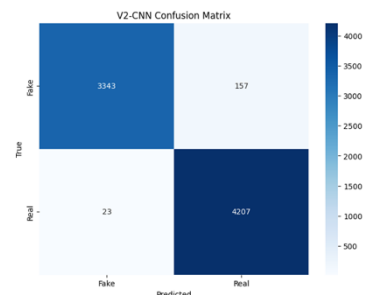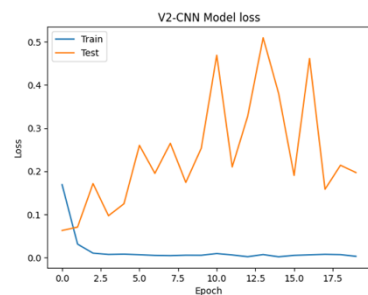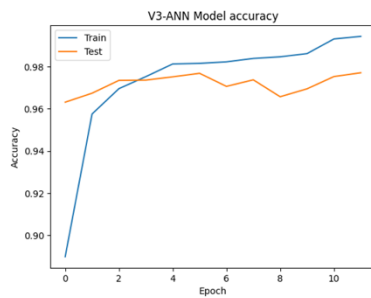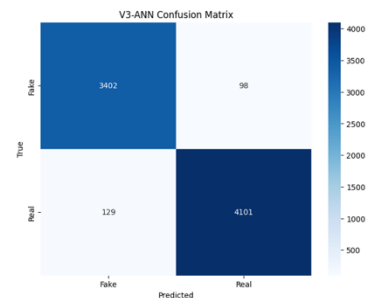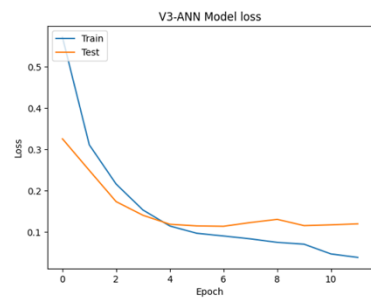
V3-ANN Model loss

V3-ANN Confusion Matrix

4.1.4 Best performing model:
- V3 ANN: Exhibits the best performance in terms of reducing overfitting. While its accuracy is slightly lower than V2 CNN (0.97 vs. 0.98), the regularization and early stopping techniques used in V3 ANN help it generalize better to unseen data and reduce overfitting significantly, making it the best model for practical applications.
- V2 Models: High accuracy (0.98) but also increasing loss on test data with increasing number of epochs, some signs of overfitting, indicating it may not generalize as well as V3 ANN.
- V1 Models: Both CNN and ANN exhibited more significant overfitting, with higher training metrics but lower validation performance, making them less suitable for practical use.

4.1.5 Impact of specific choices:
- Epochs: Increasing the epochs from 5 to 20 (V1 to V2) showed marked improvement in performance metrics, indicating the importance of sufficient training duration.
- Regularization and Batch Normalization: Implementing these techniques in V3 significantly reduced overfitting, evidenced by lower loss and high accuracy.
- Early Stopping: This technique in V3 helped in avoiding over-training, maintaining model generalization capability.

4.2 Model validation

To validate the models effectiveness, I compared their performance against a The V3 ANN model is the best performer overall, due to its effective reduction of overfitting and balanced performance metrics. While the V2 CNN model has a slightly higher accuracy, it shows signs of overfitting, making it less reliable for generalization to new data. The use of regularization and early stopping in V3 ANN has minimized overfitting, providing a robust model with good generalization capabilities.

To further validate the V3 model, I tested the model prediction on three new datasets that have not benn used for training (Generated & Scraped Dataset, DS-02 and DS-03).

V3-ANN MODEL Confusion Matrix (GEN- / SCR. DATA)

	precision	recall	f1-score	support
Fake	0.67	1.00	0.80	100
Real	1.00	0.50	0.67	100
accuracy			0.75	200
macro avg	0.83	0.75	0.73	200
weighted avg	0.83	0.75	0.73	200

V3-ANN Model Confusion Matrix (DS-02)

	precision	recall	f1-score	support
Fake	0.20	0.25	0.22	34975
Real	0.06	0.05	0.06	36430
accuracy			0.15	71405
macro avg	0.13	0.15	0.14	71405
weighted avg	0.13	0.15	0.14	71405

V3-ANN Model Confusion Matrix (DS-03)

	precision	recall	f1-score	support
Fake	0.57	0.88	0.69	3164
Real	0.74	0.35	0.48	3171
accuracy			0.61	6335
macro avg	0.66	0.61	0.58	6335
weighted avg	0.66	0.61	0.58	6335

The model's performance on unseen data was generally suboptimal. While the predictions for the generated and scraped dataset, as well as DS-03, were satisfactory, the predictions for DS-02 were less promising. However, when new articles were provided to the model via the frontend, the model's predictions were quite accurate.

The poor performance on DS-02 can be attributed to the differences between the datasets (refer to "2-data-collection-generation.pdf"). Overall, I am still quite pleased with the outcomes and various predictions produced by the ANN model V3.

4.3 Conclusion

The V3 ANN model is the best performer overall, due to its effective reduction of overfitting and balanced performance metrics. While the V2 CNN model has a slightly higher accuracy (0.98), it shows signs of overfitting, making it less reliable for generalization to new data. The use of regularization and early stopping in V3 ANN has minimized overfitting, providing a robust model with good generalization capabilities.

The validation on different datasets using confusion matrices and comprehensive performance metrics confirms the robustness and accuracy of the V3 ANN model. This model ensures not only a good fit to the training data but also good generalization to unseen data, making it the best choice for practical applications.