Summary

Classical Model Training and Validation Accuracy-loss logs

Dataset Name	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
FB-Hindi	87.03	7.3457	81.80	10.0226
WA-Hindi	77.68	6.6767	71.76	10.0618
TW-Hindi	80.26	6.1887	71.32	10.2023
FB-Telugu	81.72	5.2883	70.64	7.8993
WA-Telugu	73.52	6.9347	63.99	10.2721
TW-Telugu	73.72	6.9461	64.89	8.2022
FB-Bengali	89.80	12.9458	80.42	21.3304
WA-Bengali	85.91	4.7105	70.33	10.9235
TW-Bengali	81.21	10.8032	70.62	15.3960

Hybrid Model Training and Validation Accuracy-loss logs

Dataset Name	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
FB-Hindi	85.17	8.4169	77.77	12.5667
WA-Hindi	80.06	6.7736	69.55	11.1335
TW-Hindi	79.39	6.9773	69.65	11.2484
FB-Telugu	77.64	7.2082	63.19	10.9479
WA-Telugu	79.39	6.9773	69.65	11.2280
TW-Telugu	75.13	7.5832	58.93	10.4739
FB-Bengali	86.70	13.9220	74.25	26.5977
WA-Bengali	86.48	5.8872	65.83	11.4204
TW-Bengali	76.56	11.7028	63.19	17.2765

Classical Model Testing Accuracy-Loss logs

Dataset Name	Testing Accuracy	Testing Loss
FB-Hindi	80.29	9.9528
WA-Hindi	69.58	9.8855
TW-Hindi	69.09	10.0289
FB-Telugu	70.75	7.2481
WA-Telugu	62.68	8.8873
TW-Telugu	62.72	8.6370
FB-Bengali	78.11	20.4141
WA-Bengali	72.27	11.0362
TW-Bengali	6999	17.7104

Hybrid Model Testing Accuracy-Loss logs

Dataset Name	Testing Accuracy	Testing Loss
FB-Hindi	78.13	12.2256
WA-Hindi	67.03	10.9464
TW-Hindi	68.06	10.9776
FB-Telugu	63.60	9.2218
WA-Telugu	54.76	10.4635
TW-Telugu	55.85	11.0743
FB-Bengali	75.09	25.2992
WA-Bengali	67.65	13.1142
TW-Bengali	5830	21.6946

Dataset Statistics

Dataset Name	Total Samples	Sentence Min Length	Sentence Max Length	Sentence 95^{th} perc. Length
FB-Hindi	1069	1	245	62
WA-Hindi	1001	1	33	24
TW-Hindi	1001	1	33	24
FB-Telugu	987	1	87	28
WA-Telugu	748	1	28	21
TW-Telugu	1164	1	30	22
FB-Bengali	231	1	128	90
WA-Bengali	346	1	62	26
TW-Bengali	209	1	33	28

Since the best performance was observed on the Facebook Hindi-English code-mixed dataset, I've included the detailed classification report for the same, for each tag in the test set:

```
=== Hybrid (Quantum-Classical) Evaluation ===
Total tokens evaluated: 3521

Classification report (per-tag precision / recall / f1):

              precision    recall  f1-score   support

          CC       0.59      0.47      0.52       118
          DT       0.92      0.87      0.89       238
         G_J       0.62      0.62      0.62       199
         G_N       0.79      0.84      0.82       755
       G_PRP       0.79      0.88      0.83       336
       G_PRT       0.51      0.51      0.51       142
         G_R       0.64      0.62      0.63       188
       G_SYM       0.67      0.32      0.43        31
         G_V       0.87      0.82      0.85       697
         G_X       0.96      0.96      0.96       478
         PSP       0.80      0.83      0.81       339

    accuracy                           0.80      3521
   macro avg       0.74      0.70      0.72      3521
weighted avg       0.80      0.80      0.80      3521

Per-tag counts and accuracy:
Tag                    Support Precision   Recall       F1   Accuracy
------------------------------------------------------------------------
CC                         118     0.591    0.466    0.521      0.466
DT                         238     0.916    0.866    0.890      0.866
G_J                        199     0.620    0.623    0.622      0.623
G_N                        755     0.792    0.841    0.816      0.841
G_PRP                      336     0.789    0.878    0.831      0.878
G_PRT                      142     0.514    0.507    0.511      0.507
G_R                        188     0.639    0.622    0.631      0.622
G_SYM                       31     0.667    0.323    0.435      0.323
G_V                        697     0.871    0.824    0.847      0.824
G_X                        478     0.960    0.958    0.959      0.958
PSP                        339     0.796    0.829    0.812      0.829
------------------------------------------------------------------------
```

```
=== Hybrid (Quantum-Classical) Evaluation ===
Total tokens evaluated: 3521

Classification report (per-tag precision / recall / f1):

              precision    recall  f1-score   support

          CC       0.47      0.12      0.19       118
          DT       0.95      0.87      0.91       238
         G_J       0.55      0.54      0.54       199
         G_N       0.79      0.84      0.81       755
       G_PRP       0.86      0.86      0.86       336
       G_PRT       0.50      0.37      0.42       142
         G_R       0.53      0.60      0.56       188
       G_SYM       0.00      0.00      0.00        31
         G_V       0.79      0.86      0.82       697
         G_X       0.96      0.94      0.95       478
         PSP       0.72      0.85      0.78       339

    accuracy                           0.78      3521
   macro avg       0.65      0.62      0.62      3521
weighted avg       0.77      0.78      0.77      3521

Per-tag counts and accuracy:
Tag                    Support Precision   Recall       F1   Accuracy
----------------------------------------------------------------------
CC                         118     0.467    0.119    0.189      0.119
DT                         238     0.949    0.866    0.905      0.866
G_J                        199     0.545    0.543    0.544      0.543
G_N                        755     0.790    0.837    0.813      0.837
G_PRP                      336     0.858    0.863    0.861      0.863
G_PRT                      142     0.500    0.366    0.423      0.366
G_R                        188     0.533    0.601    0.565      0.601
G_SYM                       31     0.000    0.000    0.000      0.000
G_V                        697     0.792    0.858    0.824      0.858
G_X                        478     0.964    0.941    0.952      0.941
PSP                        339     0.720    0.850    0.779      0.850
----------------------------------------------------------------------
```