[Year]

Project 2 – Explainable AI

RAVI TEJA SUNKARA

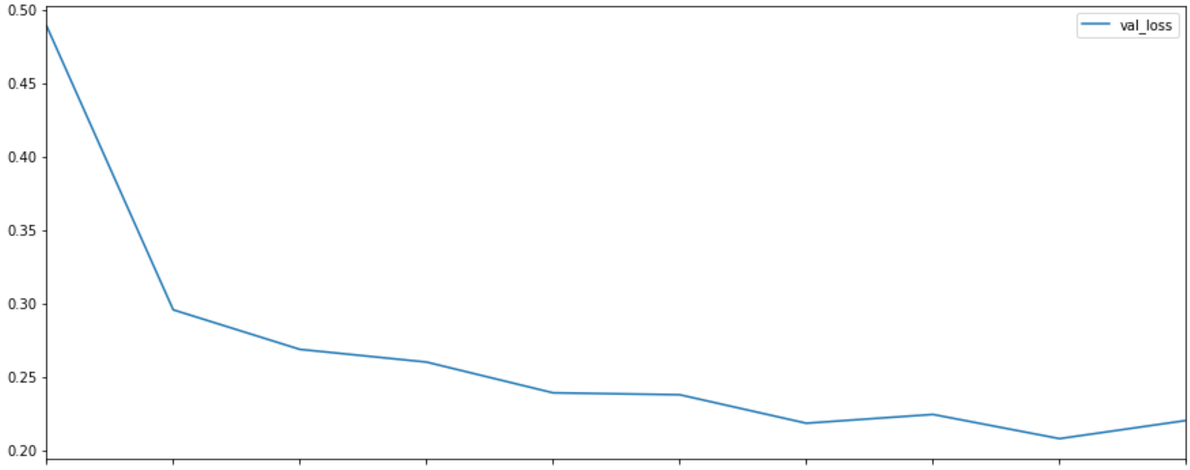
rsunkara@buffalo.edu

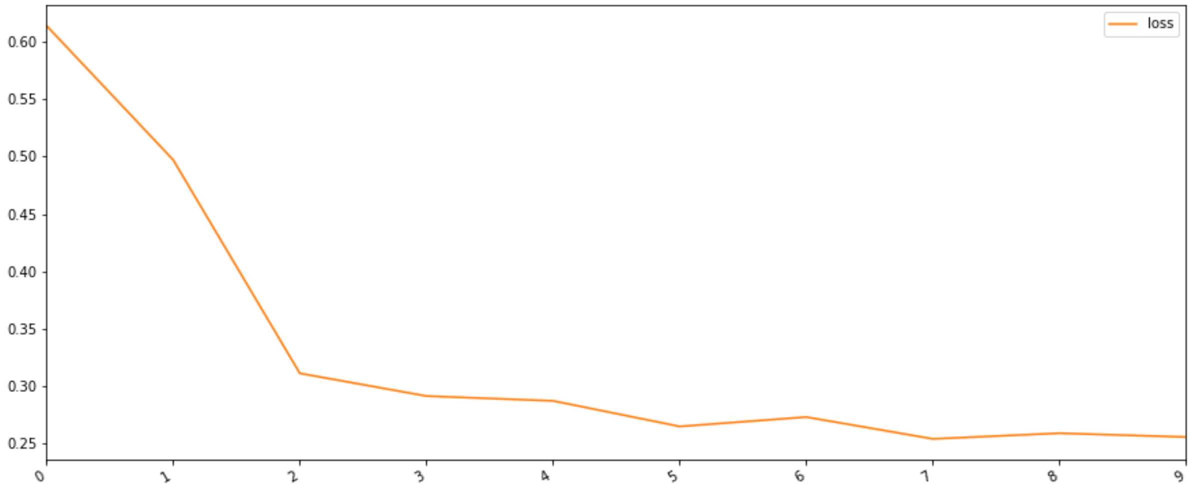
UBIT No. 50292191

TASK 3 – AUTO ENCODER

UNSEEN DATASET

Loss and Validation Loss Plots

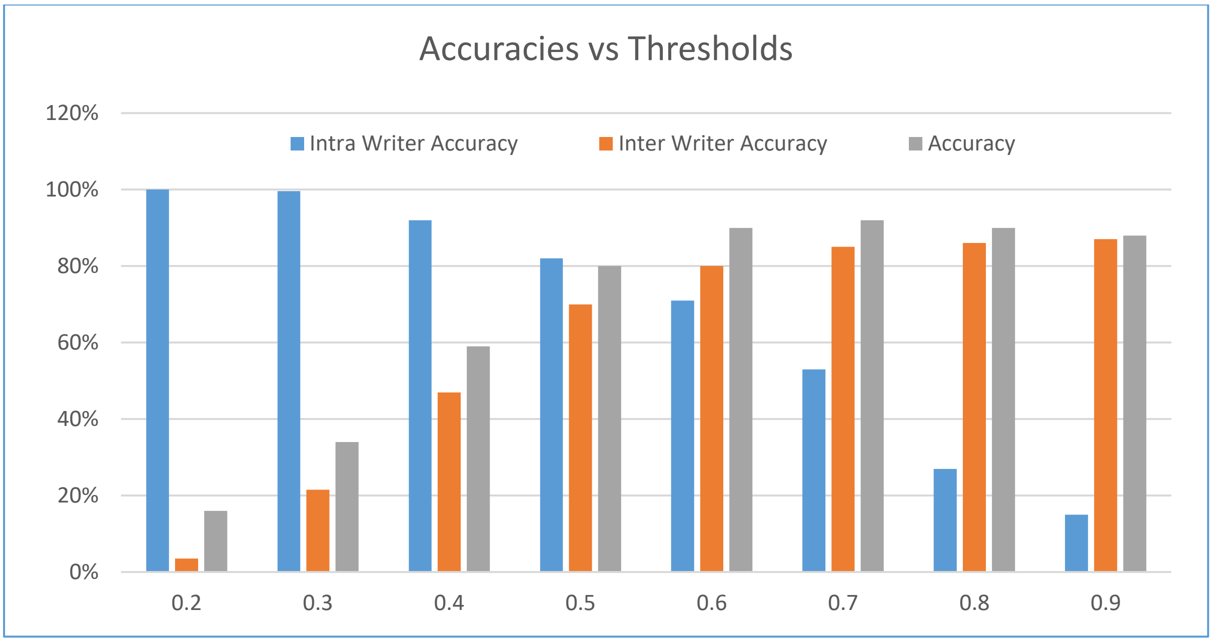




Effect of Thresholds on Accuracy

As the threshold increases, ‘Intra Writer Accuracy’ decreases while the ‘Inter Writer Accuracy’ increases.

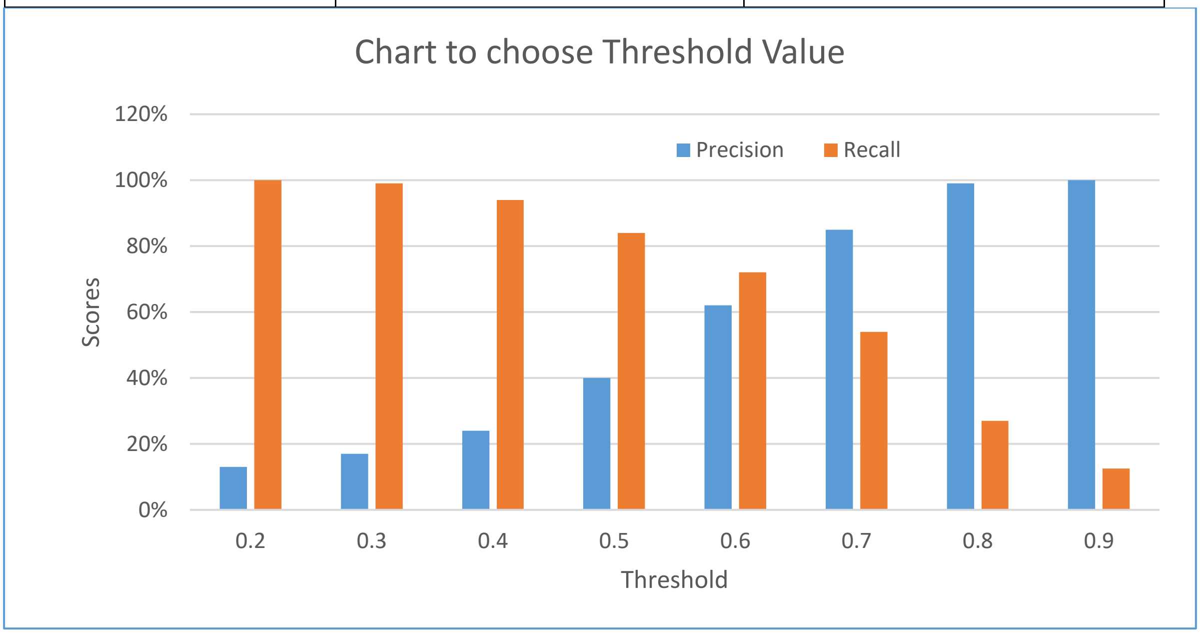
Threshold	Intra Writer Accuracy	Inter Writer Accuracy	Accuracy
0.2	100%	4%	16%
0.3	100%	22%	34%
0.4	92%	47%	59%
0.5	82%	70%	80%
0.6	71%	80%	90%
0.7	53%	85%	92%
0.8	27%	86%	90%
0.9	15%	87%	88%



Choosing Threshold Value:

Threshold is chosen such that ‘precision and recall’ are close enough and the accuracies are also good. I chose threshold as 0.7

Threshold	Precision	Recall
0.2	13%	100%
0.3	17%	99%
0.4	24%	94%
0.5	40%	84%
0.6	62%	72%
0.7	85%	54%
0.8	99%	27%
0.9	100%	13%



Effect of steps_per_epoch

As steps_per_epoch increases, the accuracy is also increasing as model is trained over more batches of samples per epoch.

steps_per_epoch	overall accuracy
1	29%
10	45%
50	57%
100	82%

Effect of number of epochs

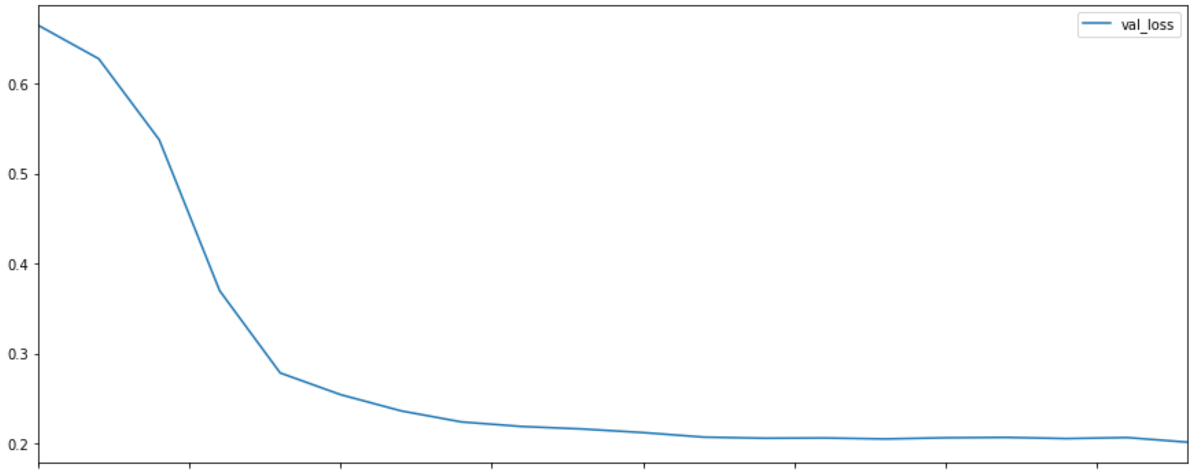
As the number of epochs increases, the accuracy is also increasing because the model is trained over the data those many number of times and it will be better able to learn the features.

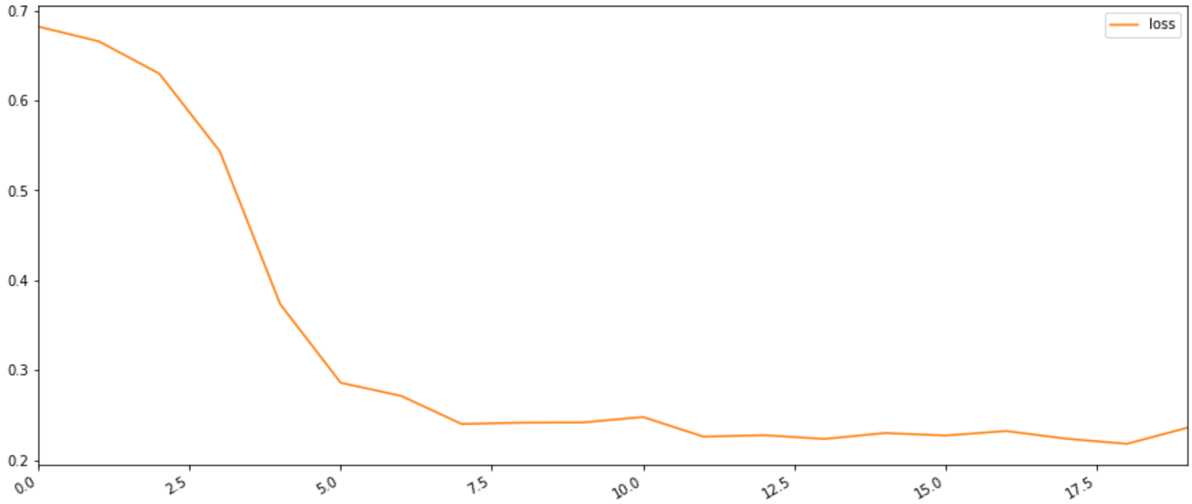
Epochs	overall accuracy
10	34%
50	48%
100	67%
1000	86%

SEEN DATASET

Loss and Validation Plots

For 20 epochs

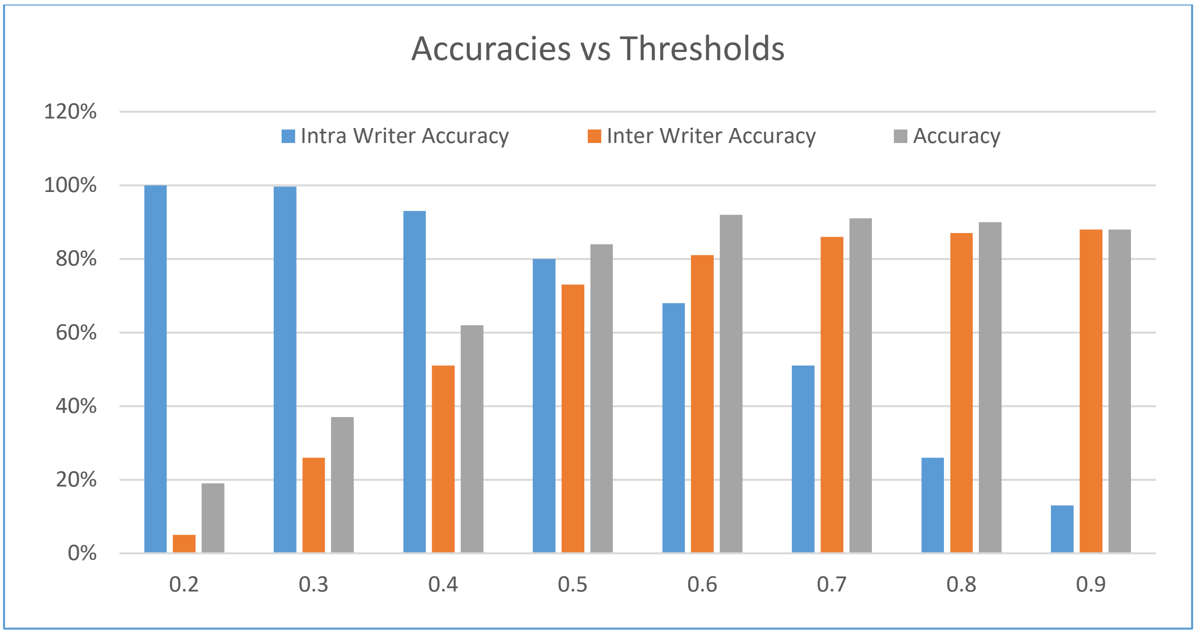




Effect of Thresholds on Accuracy

As the threshold increases, 'Intra Writer Accuracy' decreases while the 'Inter Writer Accuracy' increases.

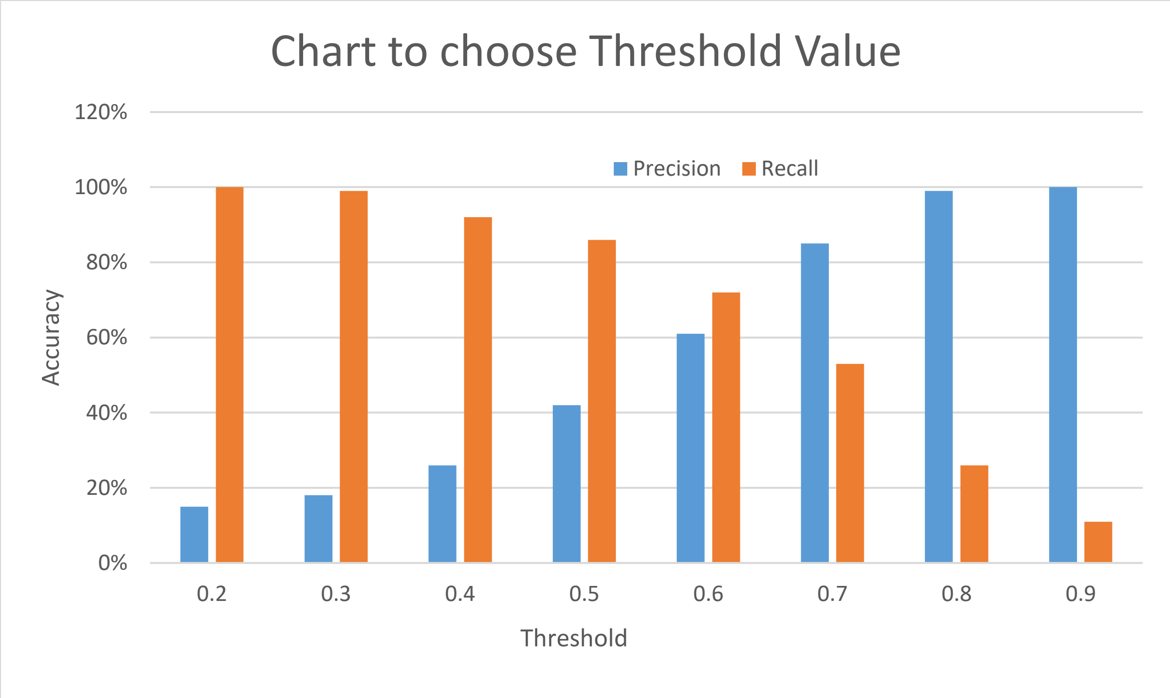
Threshold	Intra Writer Accuracy	Inter Writer Accuracy	Accuracy
0.2	100%	5%	19%
0.3	100%	26%	37%
0.4	93%	51%	62%
0.5	80%	73%	84%
0.6	68%	81%	92%
0.7	51%	86%	91%
0.8	26%	87%	90%
0.9	13%	87%	88%



Choosing Threshold Value

Threshold is chosen such that 'precision and recall' are close enough and the accuracies are also good. I chose threshold as 0.6

Threshold	Precision	Recall
0.2	15%	100%
0.3	18%	99%
0.4	26%	92%
0.5	42%	86%
0.6	61%	72%
0.7	85%	53%
0.8	99%	26%
0.9	100%	11%



Effect of steps_per_epoch

As steps_per_epoch increases, the accuracy is also increasing as model is trained over more batches of samples per epoch.

steps_per_epoch	overall accuracy
1	32%
10	49%
50	60%
100	85%

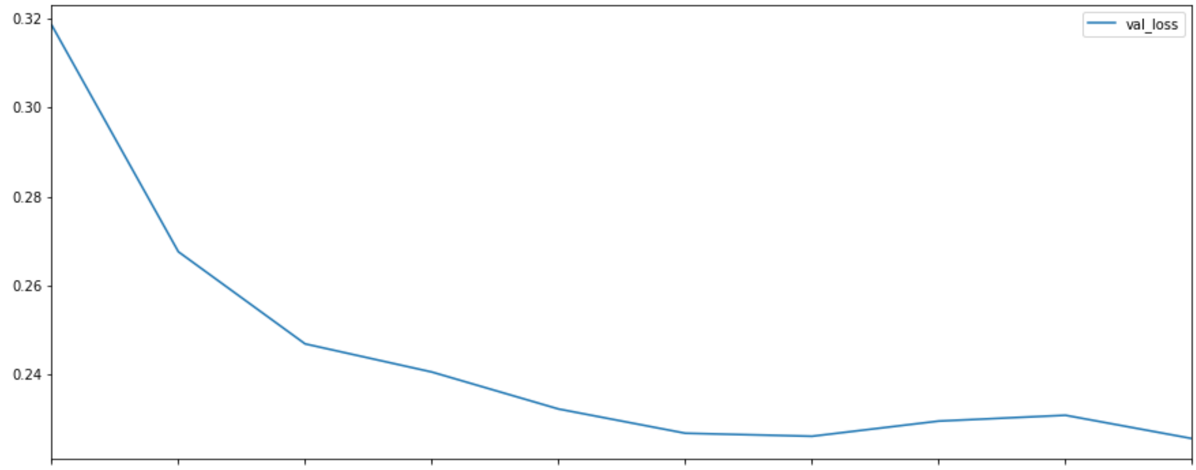
Effect of number of epochs

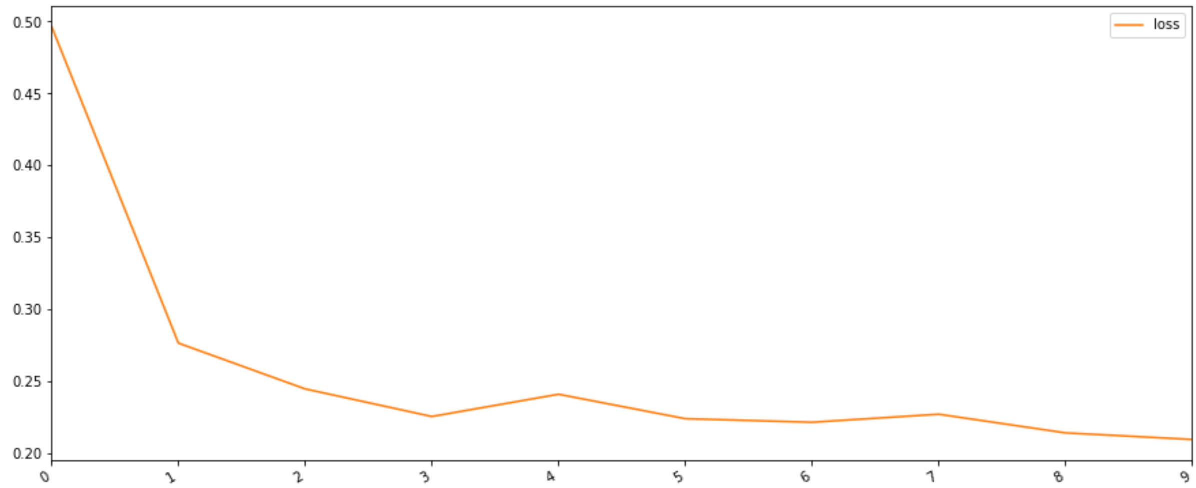
As the number of epochs increases, the accuracy is also increasing because the model is trained over the data those many number of times and it will be better able to learn the features.

Epochs	overall accuracy
10	37%
50	51%
100	68%
1000	87%

SHUFFLE DATASET

Loss and Validation Plots

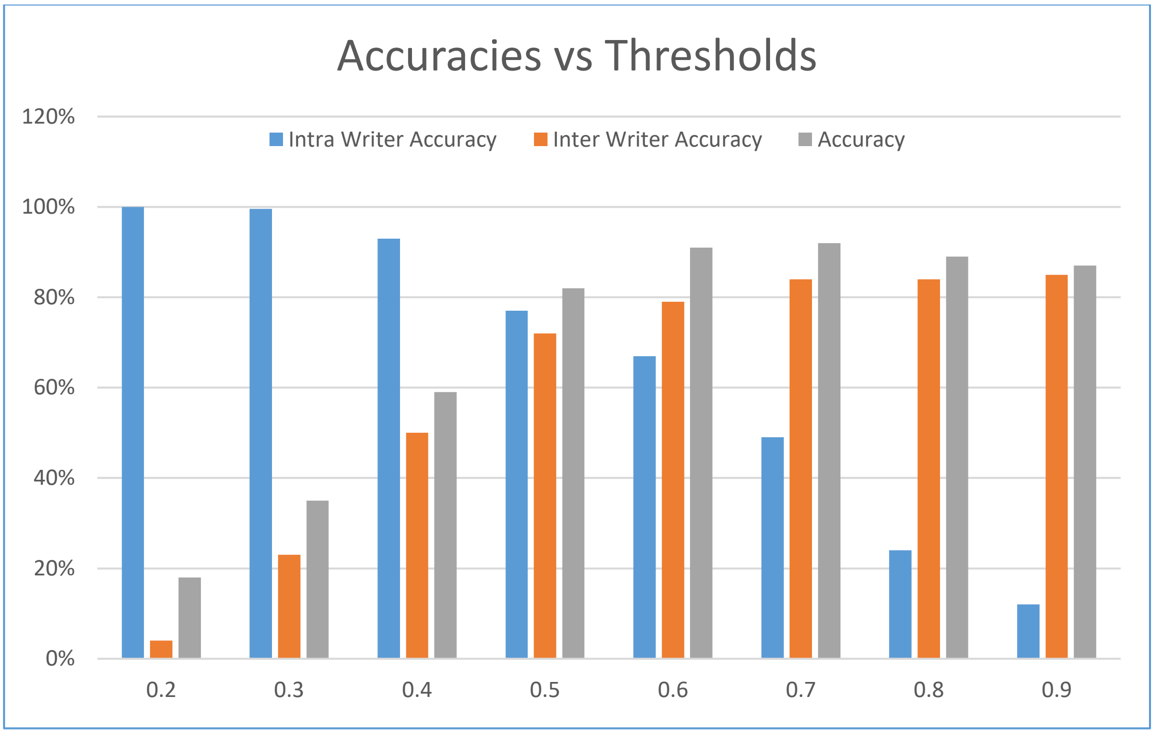




Effect of Thresholds on Accuracy

As the threshold increases, ‘Intra Writer Accuracy’ decreases while the ‘Inter Writer Accuracy’ increases.

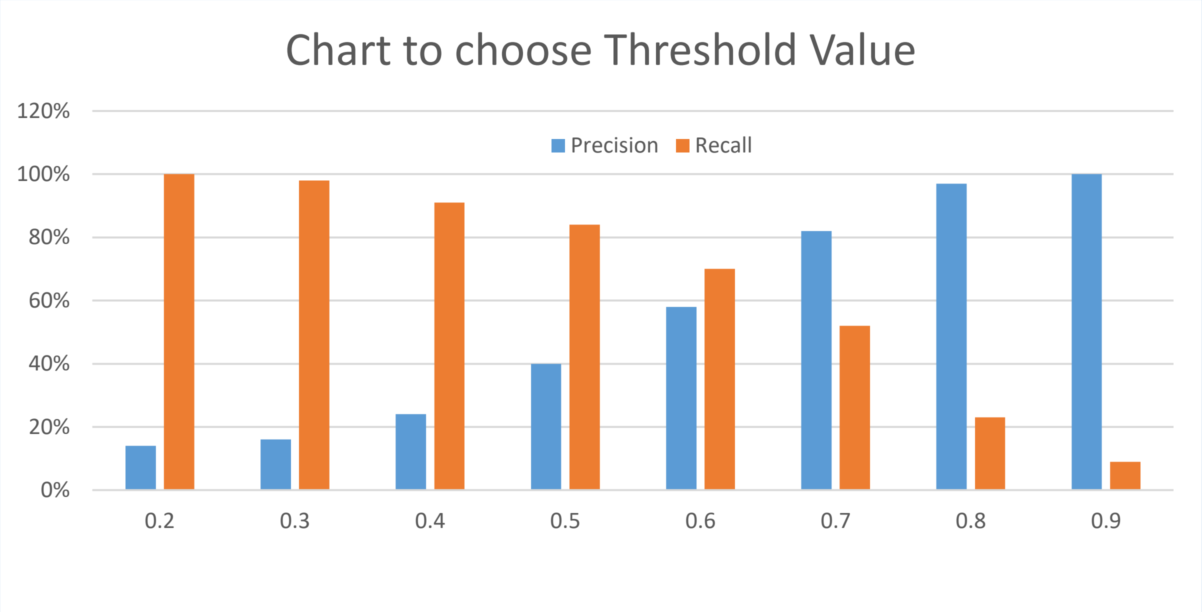
Threshold	Intra Writer Accuracy	Inter Writer Accuracy	Accuracy
0.2	100%	4%	18%
0.3	100%	23%	35%
0.4	93%	50%	59%
0.5	77%	72%	82%
0.6	67%	79%	91%
0.7	49%	84%	92%
0.8	24%	84%	89%
0.9	12%	85%	87%



Choosing a Threshold Value

Threshold is chosen such that 'precision and recall' are close enough and the accuracies are good. I chose threshold as 0.6

Threshold	Precision	Recall
0.2	14%	100%
0.3	16%	98%
0.4	24%	91%
0.5	40%	84%
0.6	58%	70%
0.7	82%	52%
0.8	97%	23%
0.9	100%	9%



Effect of steps_per_epoch

As steps_per_epoch increases, the accuracy is also increasing as model is trained over more batches of samples per epoch.

steps_per_epoch	overall accuracy
1	27%
10	44%
50	58%
100	81%

Effect of number of epochs

As the number of epochs increases, the accuracy is also increasing because the model is trained over the data those many number of times and it will be better able to learn the features.

Epochs	overall accuracy
10	34%
50	46%
100	62%
1000	84%

TASK 2 – BAYESIAN INFERENCE

UNSEEN DATASET

K2 Scores

The K2 score is calculated on the three models using the unseen training dataset as can be seen in Figure 2. After calculating the K2 score, the model with lowest K2 score is selected to get the accuracy for the unseen validation dataset.

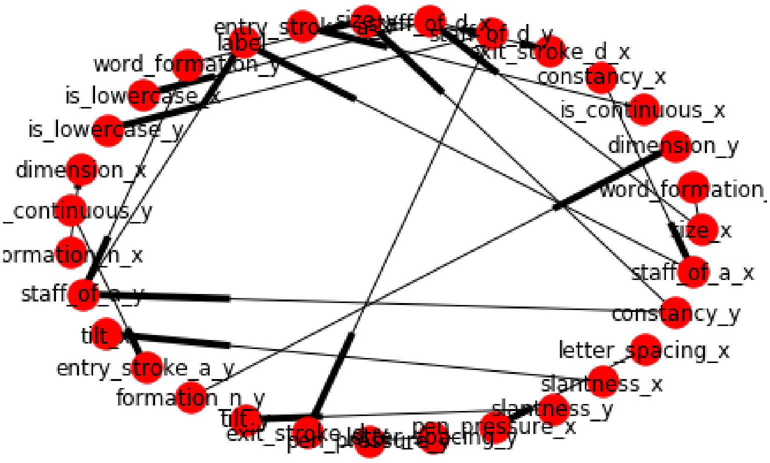


Figure: Graph of Best Bayesian Model for unseen dataset

Model	K2 Score
Model_1	-2891077.86
Model_2	-2888104.46
Model_3	-2885682.79

Figure: Different models along with their K2 scores on unseen dataset.

Time to Train

Computation time for fit using different models.

Model	Time to Train
Model_1	4.3765626
Model_2	4.2
Model_3	3.94

Time to Infer

Computation time for inference using the best Bayesian model which is model 3 in our case.

Model	Time to Infer
Model_3	203.31

Best Model

Model	Model3
Accuracy	73.45%
Number of Edges	24
Time to Train	3.94
Time to Infer	203.31

SEEN DATASET

K2 Scores

The K2 score is calculated on the three models using the unseen training dataset as can be seen in Figure 2. After calculating the K2 score, the model with lowest K2 score is selected to get the accuracy for the seen validation dataset.

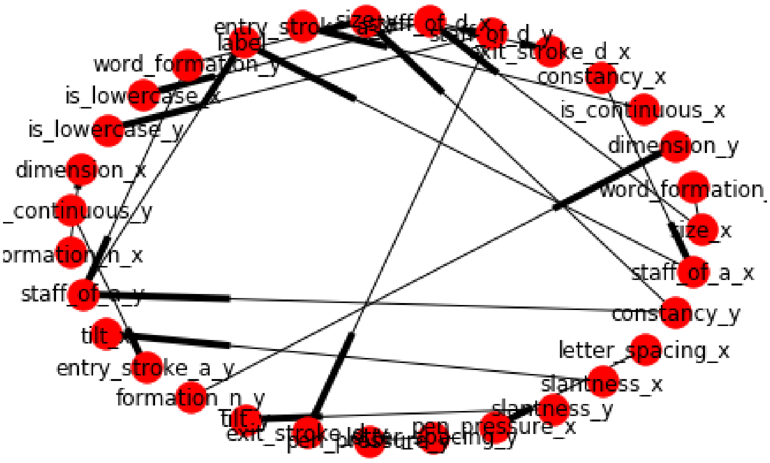


Figure: Graph of Best Bayesian Model for seen dataset

Model	K2 Score
Model_1	-2533282.8860581904
Model_2	-2528411.997269278
Model_3	-2524527.4315670864

Figure: Different models along with their K2 scores on seen dataset.

Time to Train

Computation time for fit using different models.

Model	Time to Train
Model_1	4.3765626
Model_2	4.2
Model_3	3.94

Time to Infer

Computation time for inference using the best Bayesian model which is model 3 in our case.

Model	Time to Infer
Model_3	45.025

Best Model

Model	Model3
Accuracy	79.88%
Number of Edges	24
Time to Train	3.94
Time to Infer	45.025

SHUFFLED DATASET

K2 Scores

The K2 score is calculated on the three models using the unseen training dataset as can be seen in Figure 2. After calculating the K2 score, the model with lowest K2 score is selected to get the accuracy for the shuffled validation dataset.

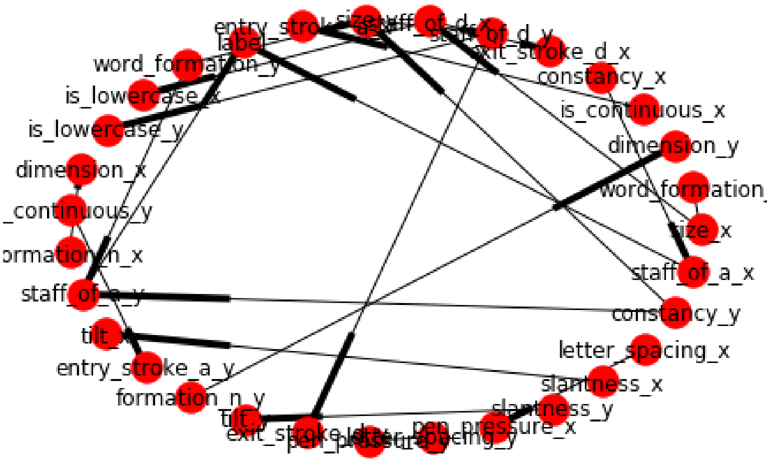


Figure: Graph of Best Bayesian Model for shuffled dataset

Model	K2 Score
Model_1	-2268686.60
Model_2	-2263798.64
Model_3	-2260734.50

Figure: Different models along with their K2 scores on seen dataset.

Time to Train

Computation time for fit using different models.

Model	Time to Train
Model_1	4.3765626
Model_2	4.2
Model_3	3.94

Time to Infer

Computation time for inference using the best Bayesian model which is model 3 in our case.

Model	Time to Infer
Model_3	267.558932

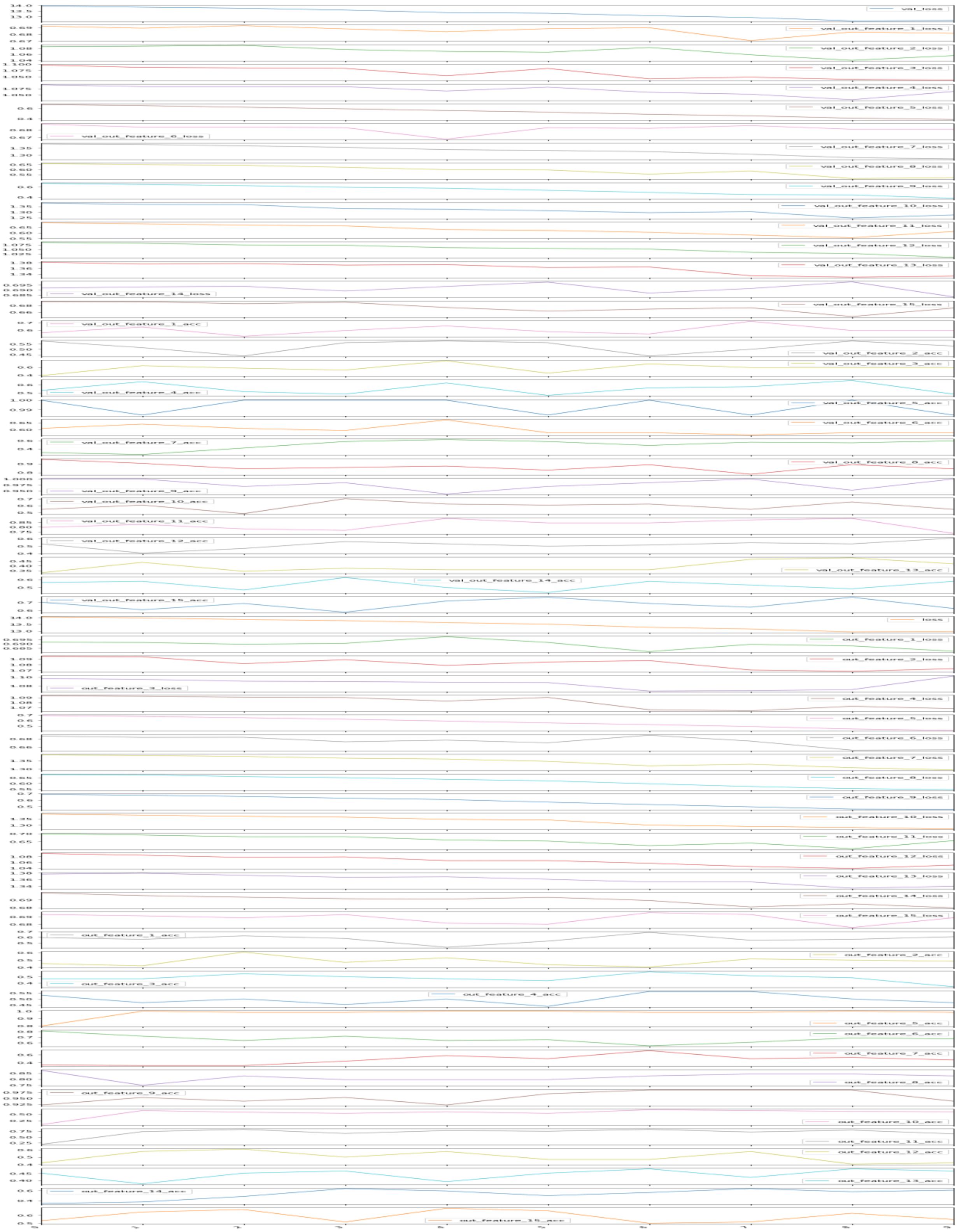
Best Model

Model	Model3
Accuracy	76.73%
Number of Edges	24
Time to Train	3.94
Time to Infer	267.558932

TASK 4 – EXPLAINABLE AI

UNSEEN DATASET

Loss and Accuracy Plots



Accuracies of val_out_features

These values are for only 10 epochs and that is the reason these are low.

0.59375
0.5625
0.53125
0.484375
1
0.59375
0.578125
0.859375
0.984375
0.703125
0.765625
0.5625
0.375
0.625
0.578125

total_acc: 0.653125

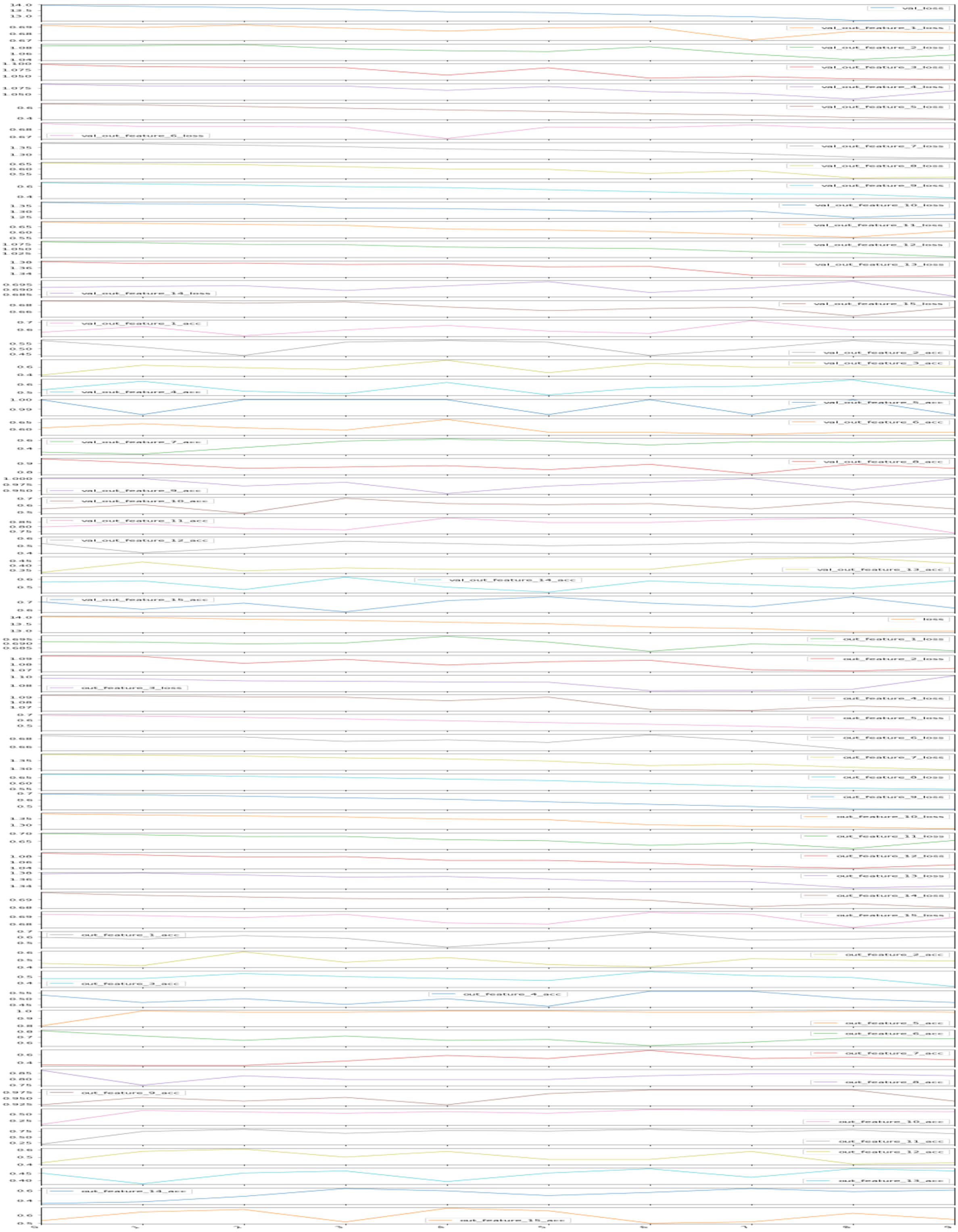
Explainability using Cosine Similarity

Threshold = 0.7
Accuracy = 92.45%
Precision = 83%
Recall = 57%

We were able to get an accuracy score of nearly 95% with the multi task learned model.

SEEN DATASET

Loss and Accuracy plots



Accuracies of val_out_features

0.79
0.76
0.73
0.68

1.00
0.79
0.78
0.96
0.99
0.90
0.97
0.76
0.58
0.83
0.78
0.820458 (total_acc)

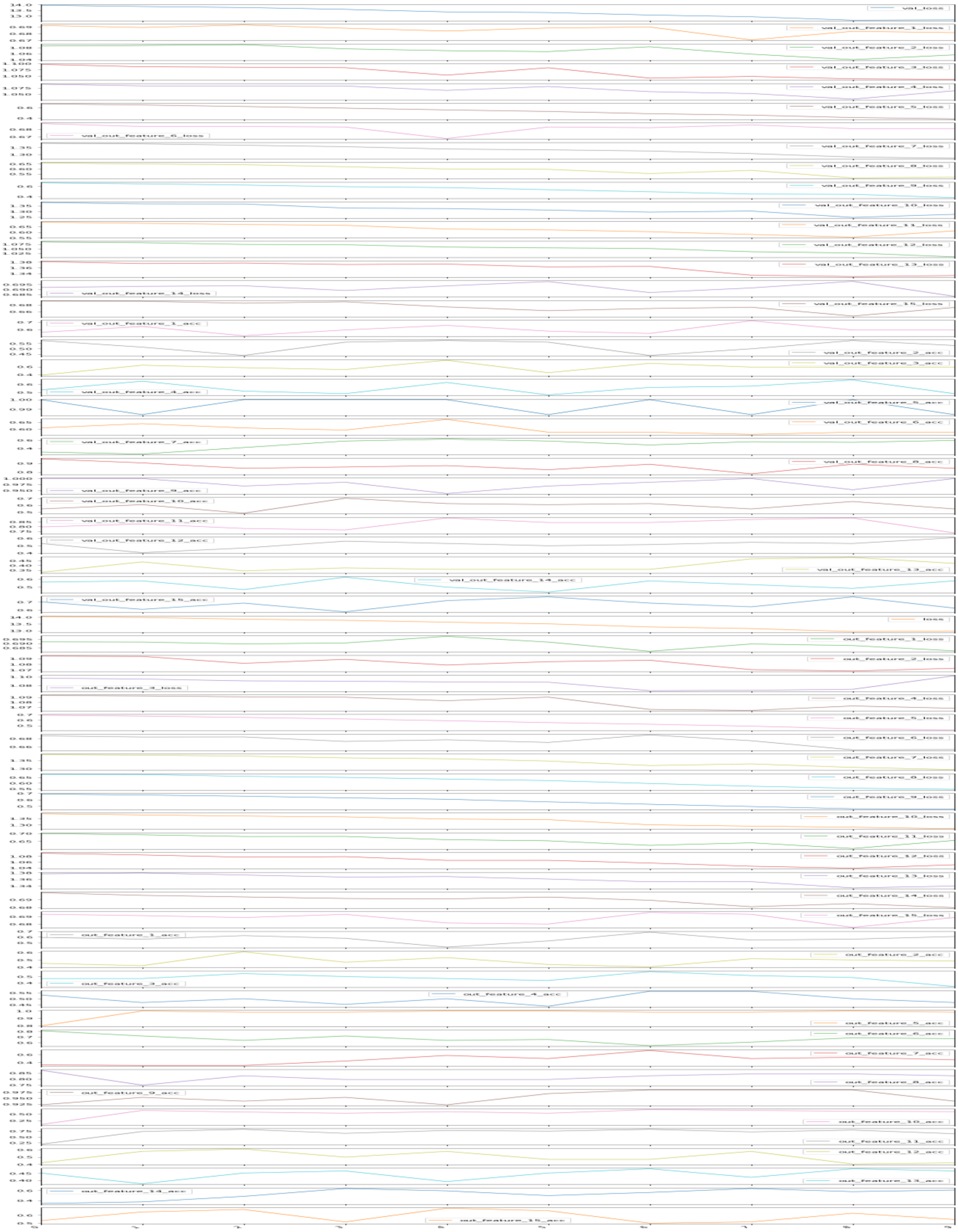
Explainability using Cosine Similarity

Threshold = 0.7
Accuracy = 94.72%
Precision = 79%
Recall = 54%

We were able to get an accuracy score of nearly 95% with the multi task learned model.

SHUFFLED DATASET

Loss and Accuracy Plots



Accuaries of val_out_features

0.73
0.72

0.69
0.62
0.96
0.72
0.73
0.86
0.69
0.90
0.97
0.86
0.68
0.83
0.78
0.783125

The last row is the total_acc

Explainability using Cosine Similarity

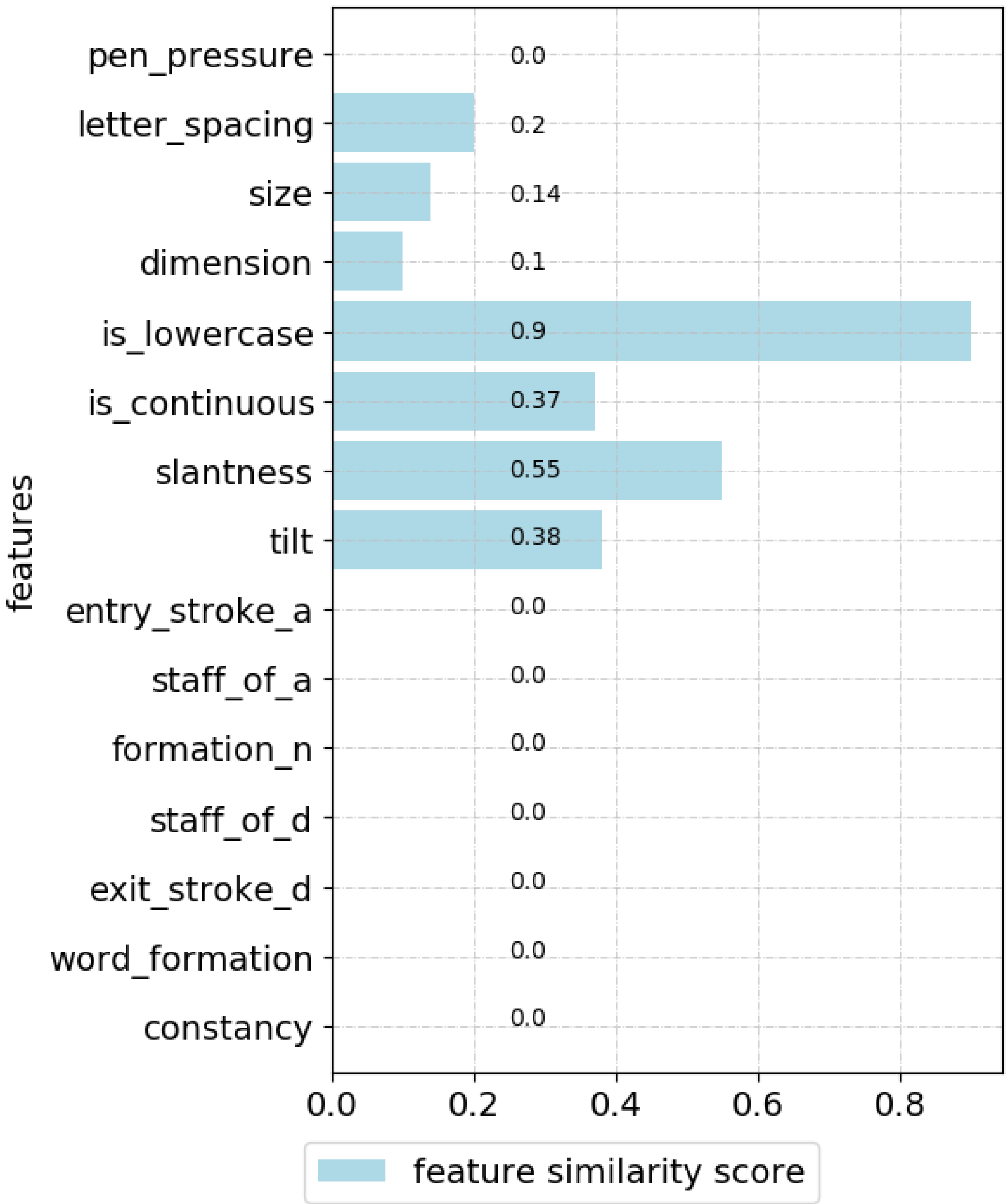
Threshold = 0.6
Accuracy = 93.72%
Precision = 81%
Recall = 61%

We were able to get an accuracy score of nearly 94% with the multi task learned model.

CAVEATS

- Siamese network also has been implemented for Task 3 but due to less accuracy and time constraints the details aren't included in this. Please find the accuracies and inference times in the jupyter notebooks.
- Initially it has been informed that implementation of Bayesian Model from the features obtained using Multi-Task learning was not compulsory. For that reason that hasn't been implemented but as the features have already been extracted these could be passed to the Bayesian model that we have created by adding the CPDs on our own and extracting similarity scores like below.

The below similairy score has been obtained for seen dataset.



features
pen_pressure
letter_spacing
size
dimension
is_lowercase
is_continuous
slantness
tilt
entry_stroke_a
staff_of_a
formation_n
staff_of_d
exit_stroke_d
word_formation
constancy
0.0
0.2
0.14
0.1
0.9
0.37
0.55
0.38
0.0
0.0
0.0
0.0
0.0
0.0
0.0
0.0
0.2
0.4
0.6
0.8
feature similarity score