INLP ASSIGNMENT-4:ELMO

Iswar Mahapatro

2023201047

Introduction:

Unlike traditional word embeddings like Word2Vec or GloVe, which generate a single word embedding for each word in the vocabulary, ELMo produces unique embeddings for each word instance, making the embeddings dynamic based on the surrounding text.The dataset used here is News Classification dataset.And I am using the preprocessed_train.csv.

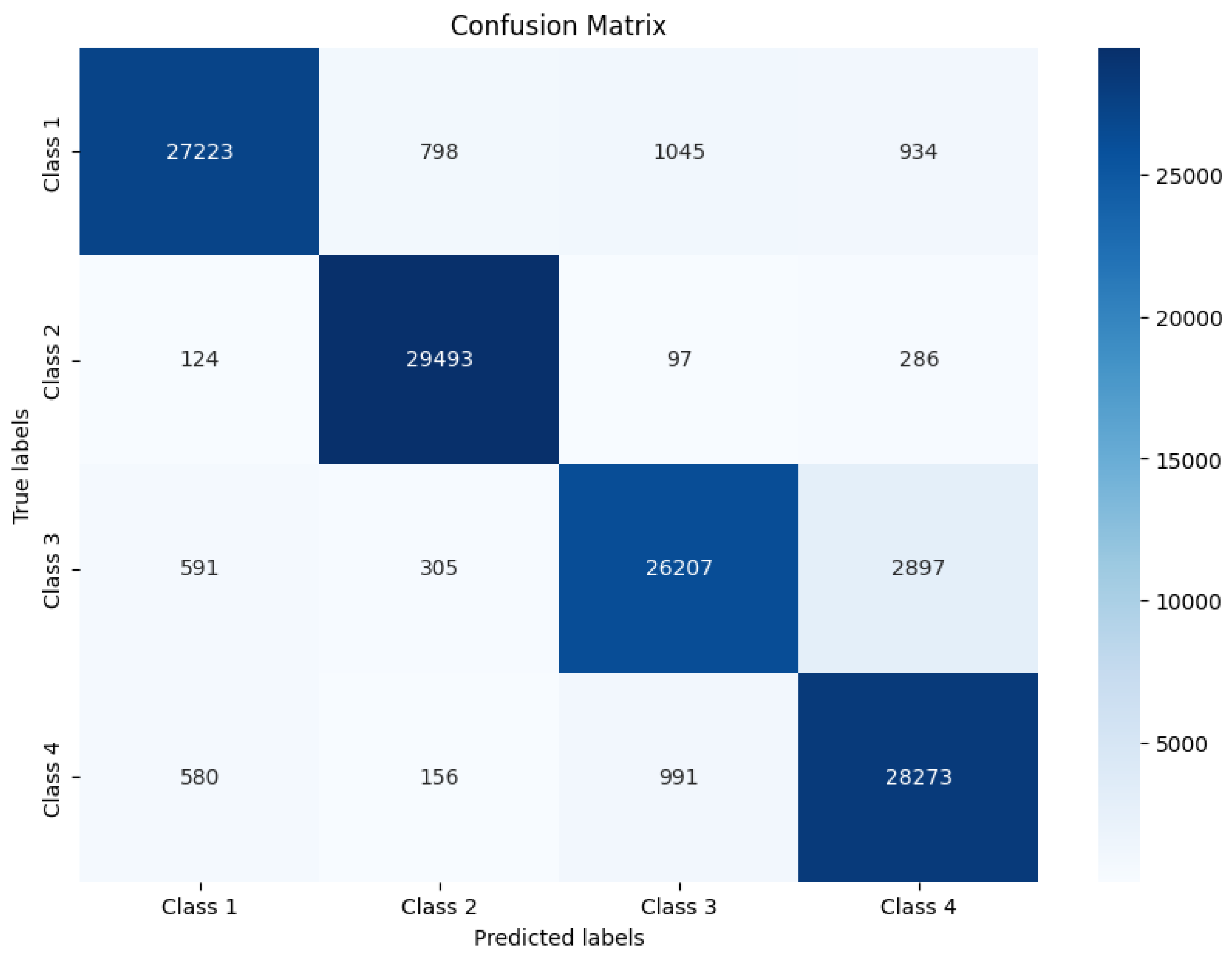
ELMO

1. Trainable $Lambda$:

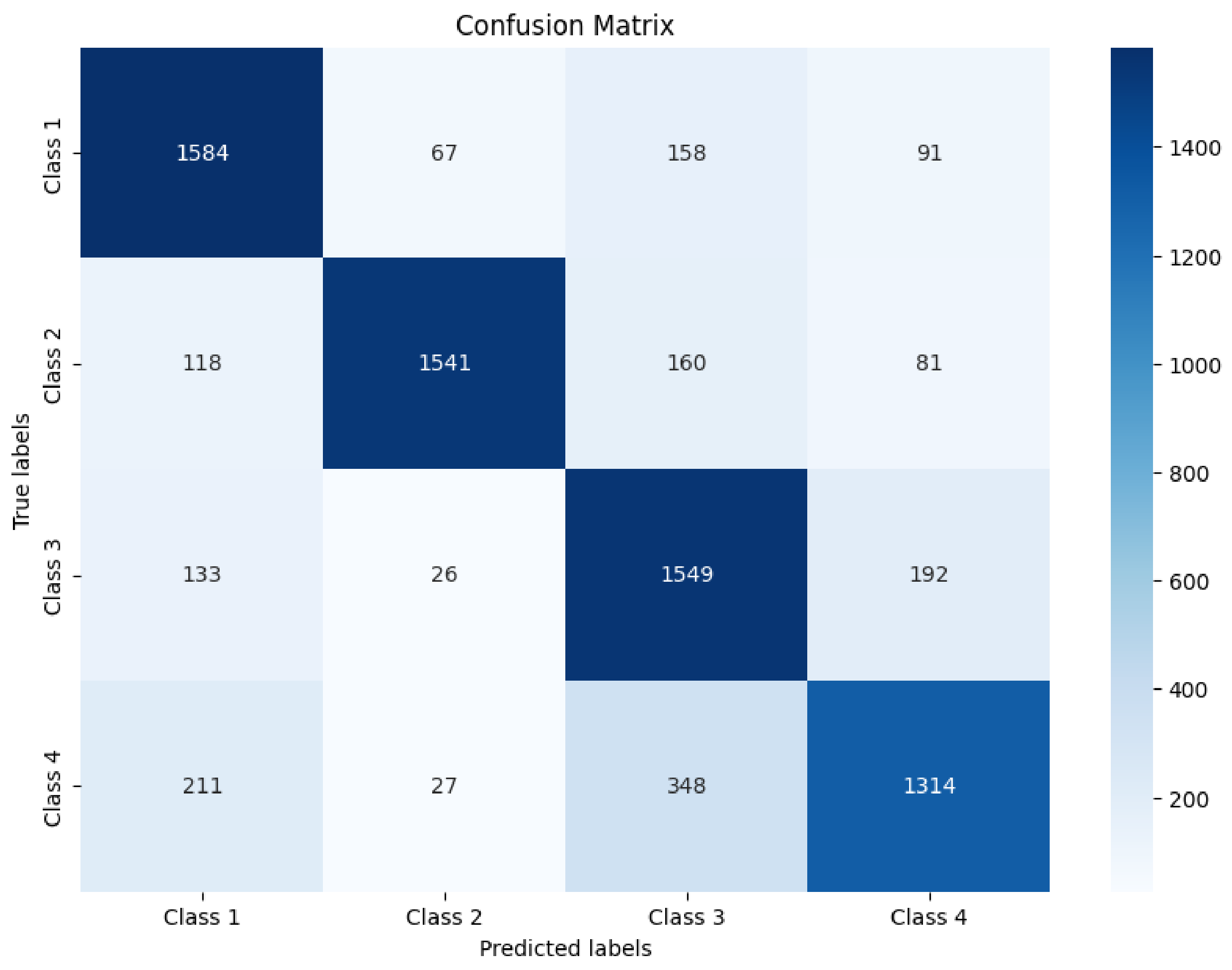
Performance metrics:

	Accuracy	Precision	Recall	F1 Score
Test	0.8005263157894736	0.8099540865900984	0.8005263157894738	0.7991750795839305
Train	0.9266166666666666	0.9277871658215255	0.9266166666666668	0.9265052470063843

Confusion Matrix(Train):



Confusion Matrix(Test):



Lambda values:

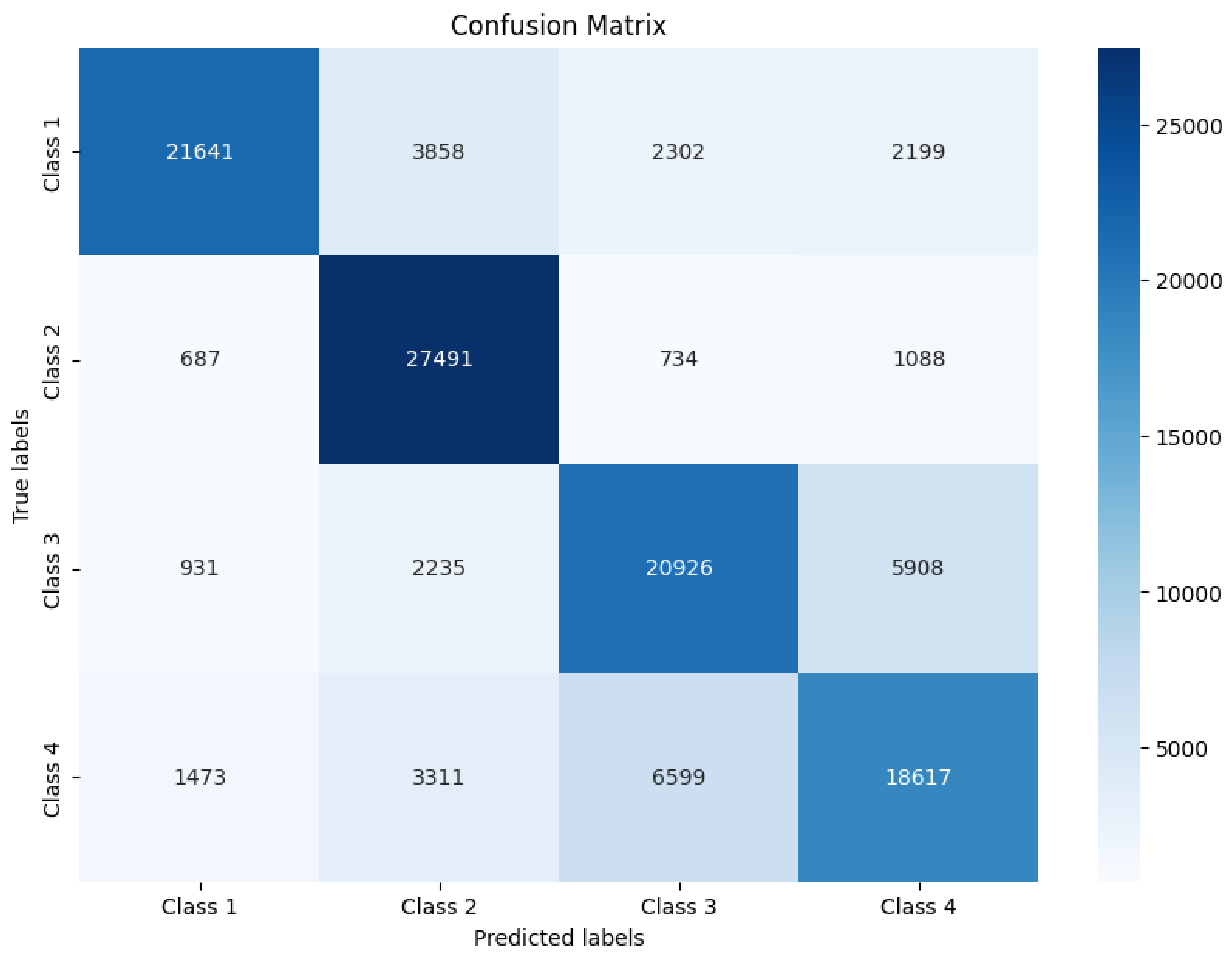
[0.0176,0.1093,0.8731]

2. Frozen $Lambda$:

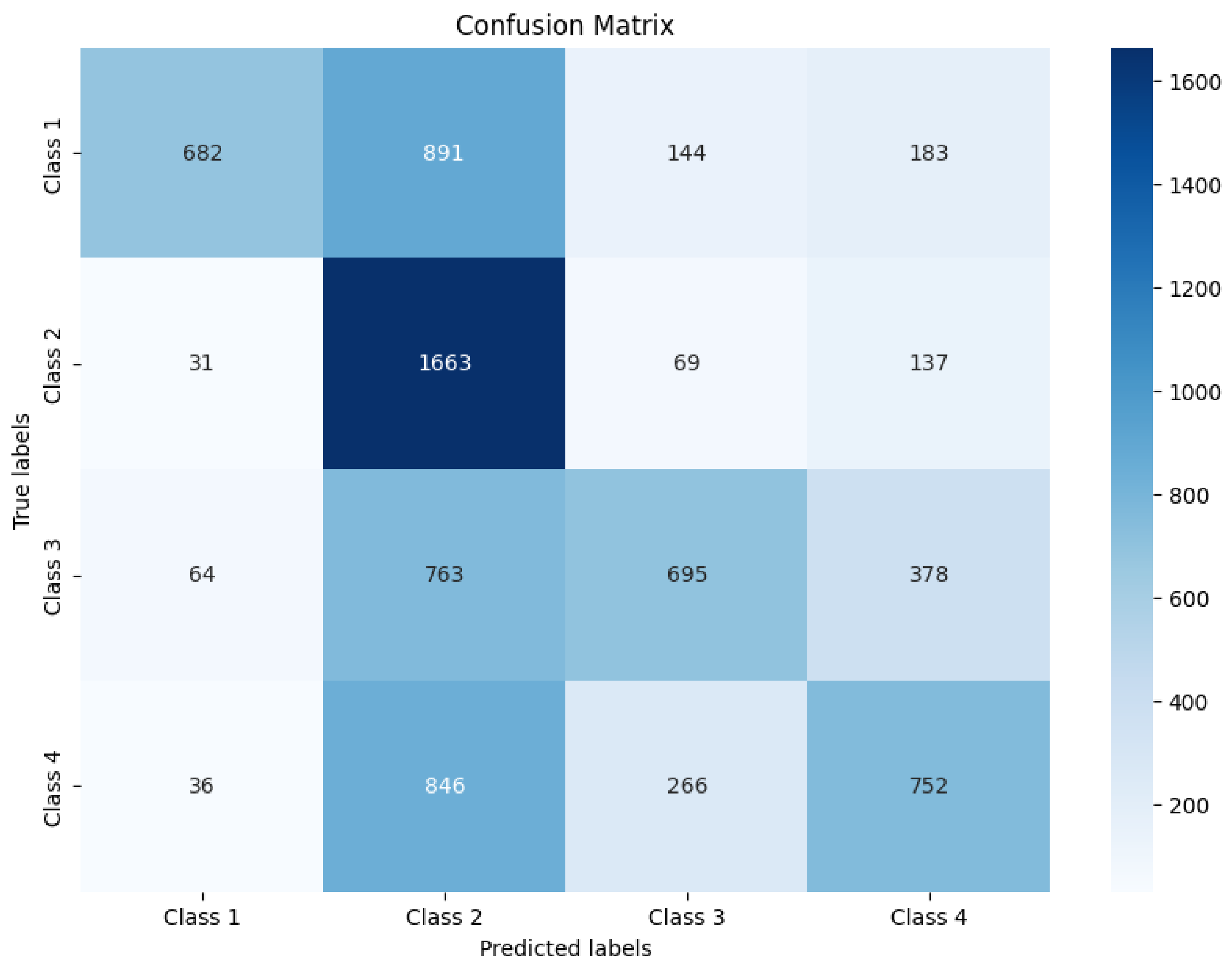
Performance metrices:

	Accuracy	Precision	Recall	F1 Score
Test	0.7578947368421053	0.7560864504291884	0.7578947368421052	0.7589173720263432
Train	0.9066166666666666	0.9077871658215255	0.9066166666666668	0.9065052470063843

Confusion Matrix(Train):



Confusion Matrix(Test):



Hyperparameter Observation:

- The trainable λ setting consistently outperforms the frozen λ setting across all metrics. This indicates that allowing the model to learn the optimal values for λ during training leads to better generalization and performance on unseen data.
- The trainable λ setting likely performs better because it adapts the weighting of different layers of the ELMo model based on the specific task and data, allowing for more flexibility and fine-tuning. In contrast, the frozen λ setting keeps the weights fixed, which may limit the model's ability to capture task-specific nuances in the data.
- Therefore, based on the provided metrics and analysis, the hyperparameter setting with trainable λ performs the best for this task.

Comparision(Word2Vec vs ELMO vs SVD):

Word2Vec:

	Accuracy	Precision	Recall	F1 Score
Test	0.8978	0.8962	0.8978	0.8971
Train	0.9702	0.9702	0.9702	0.9701

SVD:

	Accuracy	Precision	Recall	F1 Score
Test	0.8632	0.8632	0.8632	0.8630
Train	0.9420	0.9424	0.9420	0.9420

ELMO:

	Accuracy	Precision	Recall	F1 Score
Test	0.8005263157894736	0.8099540865900984	0.8005263157894738	0.7991750795839305
Train	0.9266166666666666	0.9277871658215255	0.9266166666666668	0.9265052470063843

Analysis:

- Word2Vec (Skipgram with Negative Sampling) performs moderately well on the test set but shows decent performance on the train set.
- SVD demonstrates good performance on the train set but suffers from a drop in performance on the test set.
- ELMo outperforms both Word2Vec and SVD on both train and test sets, achieving the highest accuracy and F1 score.

Conclusion:

- ELMo consistently outperforms Word2Vec and SVD in terms of accuracy, F1 score, recall, and precision.
- ELMo's contextual embeddings capture semantic information effectively, making it the most effective word vectorization method for the
given task.

	Accuracy	Precision	Recall	F1 Score
Test	0.8005263157894736	0.8099540865900984	0.8005263157894738	0.7991750795839305
Train	0.9266166666666666	0.9277871658215255	0.9266166666666668	0.9265052470063843

	Accuracy	Precision	Recall	F1 Score