ASSIGNMENT-4
ELMO Report

MANGLESH PATIDAR
2023201059

ELMo (Embeddings from Language Models)

ELMo is a contextual word embedding technique that leverages stacked Bidirectional LSTMs to capture both syntactic and semantic information from text data. By considering word context within sentences, ELMo generates word embeddings that vary based on the word's meaning in different contexts. This contextual sensitivity allows ELMo embeddings to capture nuances and ambiguities in language effectively.

EMLO Pretraining Architecture:

Word2vec 300 dim

two LSTM ’s in the forward direction and backword direction

Number of stacks: 2

Input ,output, hidden dimension of embedding layer, lstm : 300

output Dimension: Vocab size

Hyperparameter

Epoch=10

Optimizer=Adam

Learning rate=0.001
Embedding dimension= 150
Optimizer = Adam
Loss = CrossEntropyLoss
Batch Size = 32

Classification Model

1. **Frozen Weights**

2 Trainable Weights

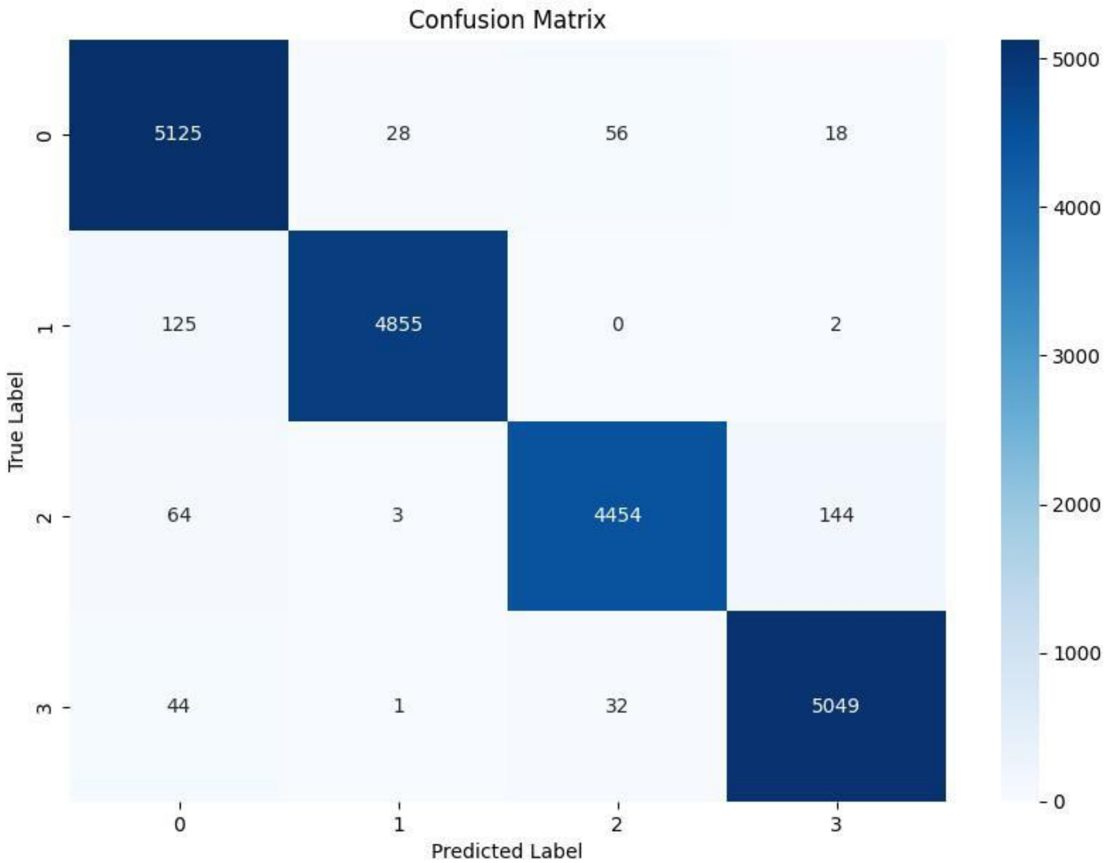
3. Learnable Function

Frozen Weights

weights = [0.33, 0.33, 0.33]

Train Metrics

Loss: 0.0759, Accuracy: 0.9817
Accuracy: 0.9817
Precision: 0.9817333478191024
Recall: 0.9817
F1 Score: 0.9816908247356889



Test Metrics

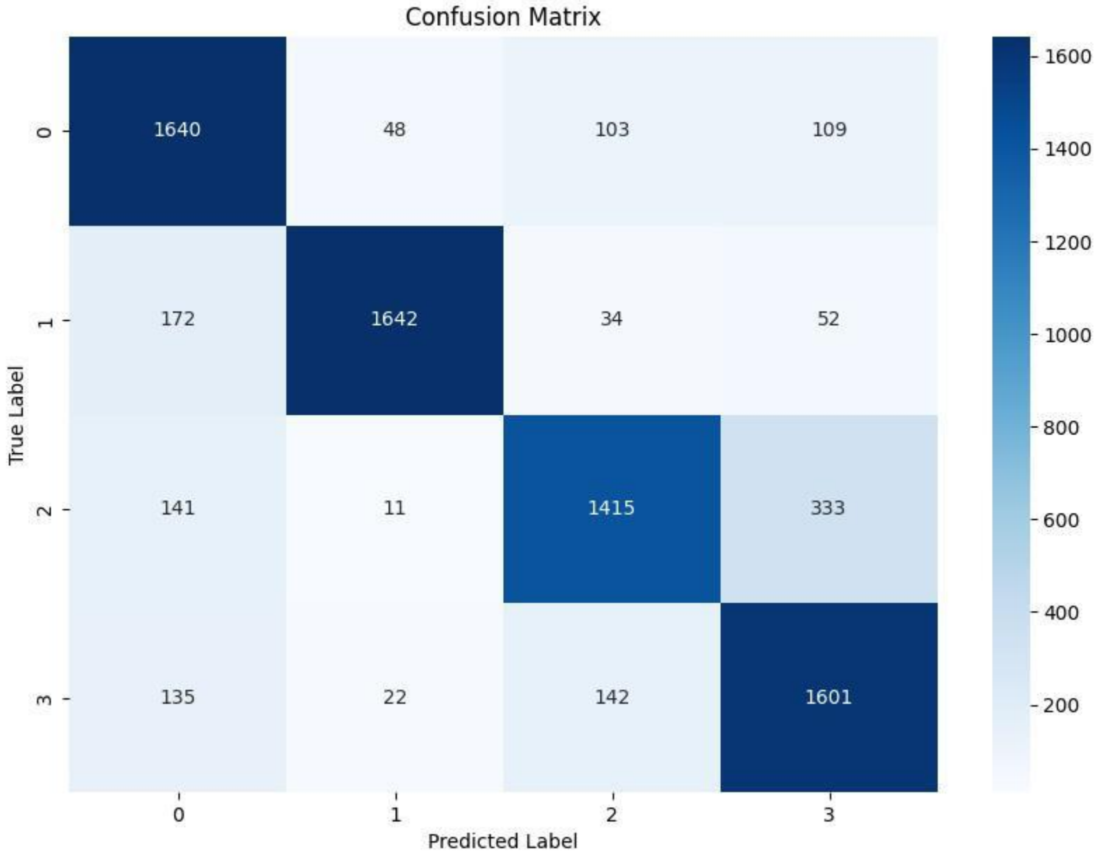
Loss: 0.5277, Accuracy: 0.8421

Accuracy: 0.8421052631578947

Precision: 0.8439886948681167

Recall: 0.8421052631578947

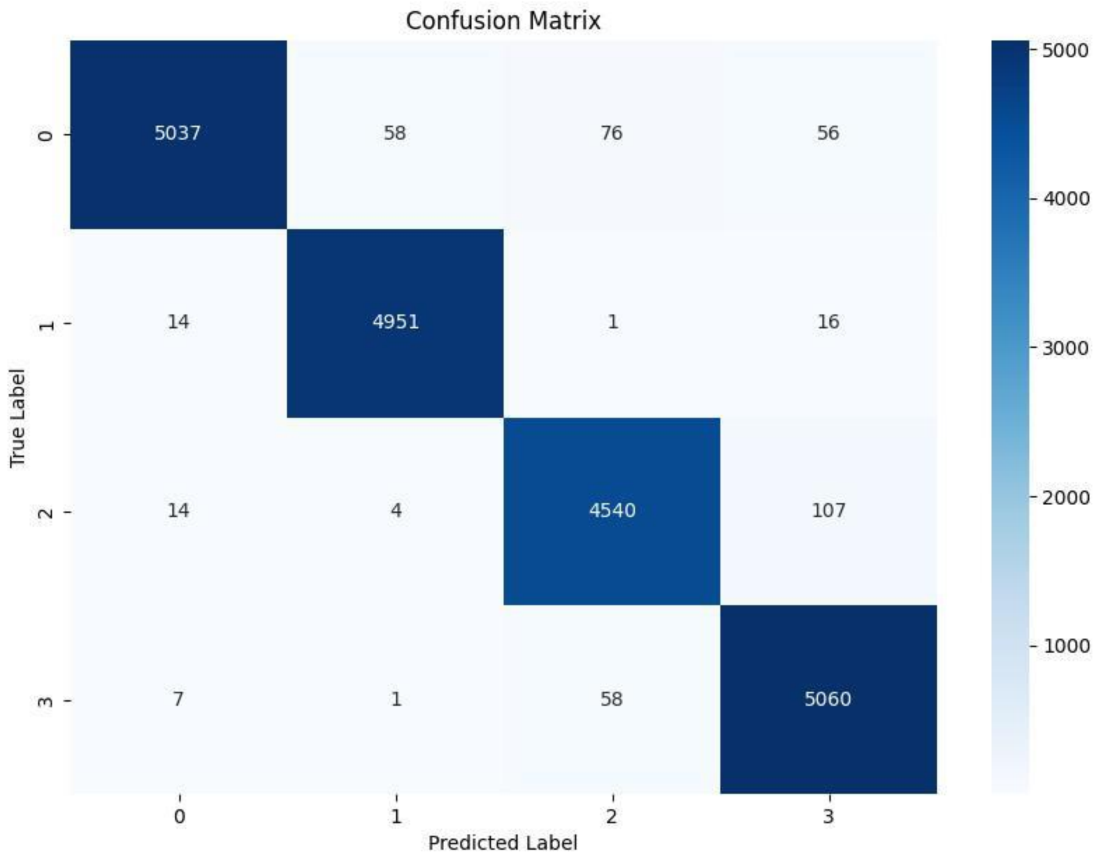
F1 Score: 0.8421575257512016



Trainable Weights

weights = [0.2465, 0.3400, 0.5102]

Train Metrics
Loss: 0.0748, Accuracy: 0.9806
Accuracy: 0.98055
Precision: 0.9806313767858823
Recall: 0.98055
F1 Score: 0.9805672255565027



Test Metrics

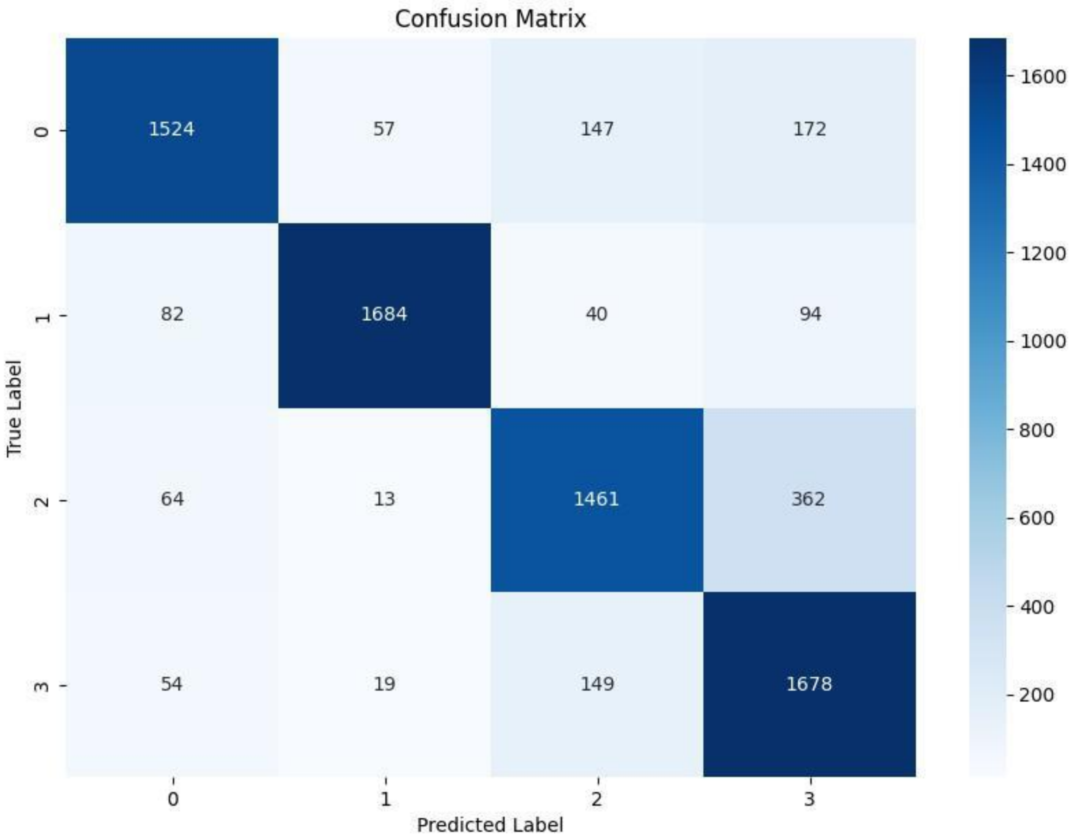
Loss: 0.5069, Accuracy: 0.8433

Accuracy: 0.8432894736842105

Precision: 0.8437262172000559

Recall: 0.8432894736842105

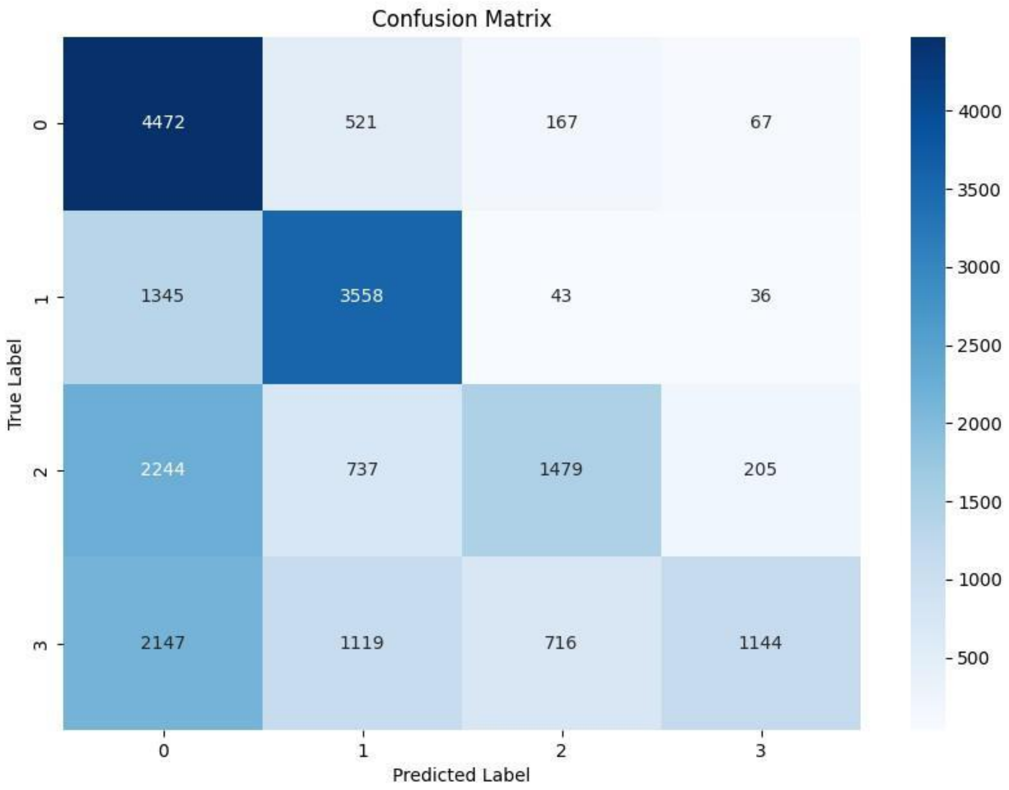
F1 Score: 0.8434486642651126



Learnable Function

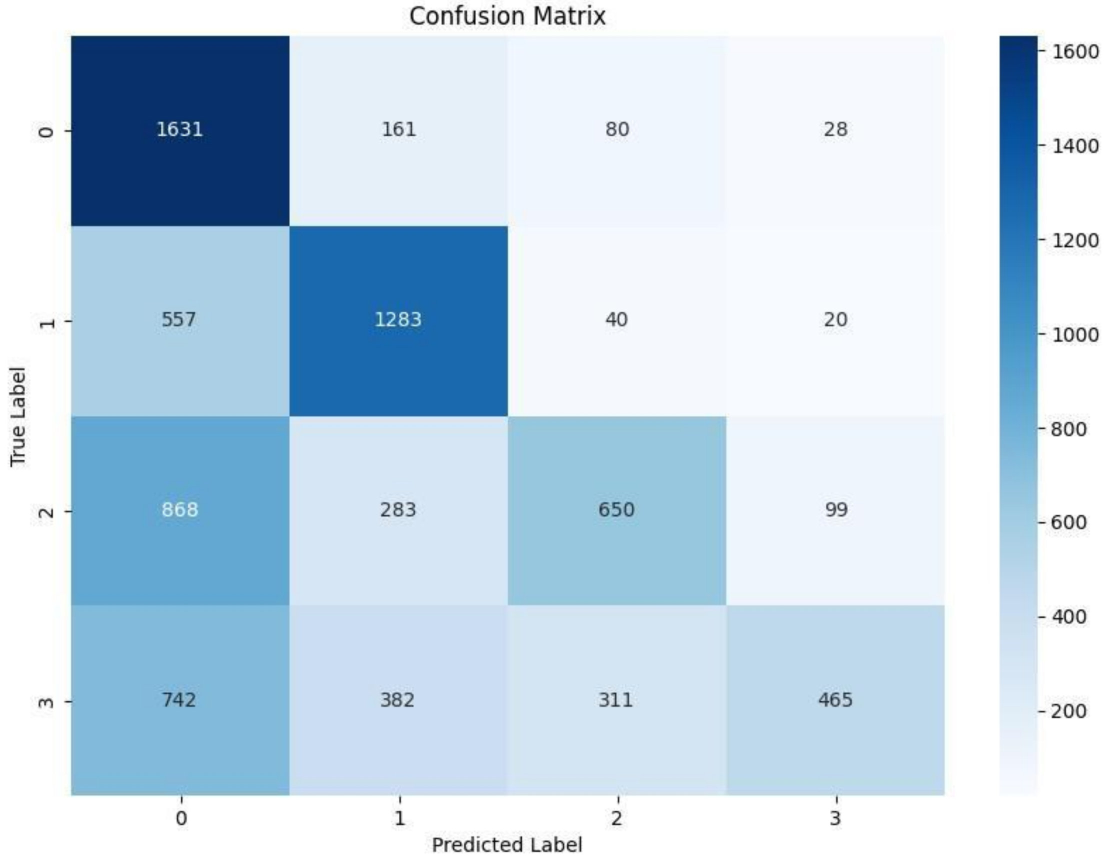
Train Metrics

Loss: 0.0835, Accuracy: 0.9792
Accuracy: 0.9792
Precision: 0.9795081860282755
Recall: 0.9792
F1 Score: 0.979240369576893



Test Metrics

Loss: 0.5824, Accuracy: 0.8250
Accuracy: 0.825
Precision: 0.8313844732618452
Recall: 0.825
F1 Score: 0.8268247235005447



SVD

Window Size=3

Train &Test Metrics

Train Accuracy= 78.329985178%

Test Accuracy:77.72368421052631

F1 Score: 0.7752065635158733

Precision: 0.7799399851780164

Recall: 0.7742416666666666

Confusion Matrix:

[[23678 1173 1706 2671]

[1678 24364 650 1938]
[2134 421 21169 5204]
[1623 611 2630 22564]]

SKIP-GRAM (word2vec)
Train & Test Metrics
Train Accuracy= 87.520656351587%
Test Accuracy: 83.424157897743%
F1 Score: 0.834572578433Precision: 0.82978054854164
Recall: 0.834689904332

Confusion Matrix:
[[26075 1034 1467 2310]
[1591 27311 654 1821]
[2189 305 23901 4765]
[1533 511 2407 25378]]

ELMO
Context Sensitivity: ELMo embeddings capture word meanings based on their context within sentences, allowing for a richer representation of word semantics.
Flexibility: ELMo embeddings can be fine-tuned for specific downstream tasks, making them adaptable to various natural language processing tasks.
Deep Learning Architecture: The use of deep Bidirectional LSTMs enables ELMo to capture complex language patterns and dependencies.

SVD (Singular Value Decomposition)
• Dimensionality Reduction: SVD effectively reduces the dimensionality of word embeddings while preserving semantic information, leading to compact representations suitable for downstream tasks.
• Global Co-occurrence Information: SVD considers the global co-occurrence statistics of words in a corpus, capturing broad semantic relationships between words.
• Computational Efficiency: SVD has relatively low computational complexity compared to deep learning-based techniques like ELMo, making it suitable for large-scale text Corpora.

Skip-gram (Word2Vec)

Context Window: Skip-gram captures contextual information by considering neighboring words within a fixed window, allowing it to capture local semantic relationships.

Scalability: Skip-gram is scalable to large text corpora and can efficiently train word embeddings on massive datasets.

Quality of Embeddings: Skip-gram embeddings often exhibit strong performance on various downstream tasks, owing to their ability to capture semantic similarities between words.